



ISSN 1720-7525

Vol. 1 [4]

[www.psychnology.org](http://www.psychnology.org)

## TABLE OF CONTENTS

Table of Contents.....	P. 327
------------------------	--------

### **TARGET ISSUE: Future Interfaces**

Editorial Preface.....	P. 328
------------------------	--------

User adaptive BCIs: SSVEP and P300 based interfaces.....	P. 331
--	--------

F. Beverina, G. Palmas, S. Silvoni, F. Piccione and S. Giove

Helping Viewers Press the Right Buttons: Generating Intuitive Labels for Digital Terrestrial TV remote controls.....	P. 355
--	--------

J. Lessiter, J. Freeman, R. Davis and A. Dumbreck

Software Design and Evaluation by Ergonomics Knowledge and Intelligent Design System (EKIDES).....	P. 378
--	--------

I. Jastrzebska-Fraczek and H. Bubb

Evaluating Collaboration and Core Competence in a Virtual Enterprise.....	P. 391
---	--------

R. Breite and H. Vanharanta

A Case Study for Reaching Web Accessibility Guidelines for the Hearing-Impaired .....	P. 400
---	--------

M. Namatame, M. Kobayashi, A. Harada

### **OTHER CONTENTS**

Virtual Realisation: Supporting Creative Outcomes in Medicine and Music.....	P. 410
--	--------

J. Waterworth

## Editorial Preface

Psychology is a very young journal, but is developing rapidly. Each successive number attracts broader interest, both in terms of the span of approaches contributed and the geographical origin of contributors and, presumably, readers. This is exactly how it should be. Psychology is an electronic-only journal, which permits more rapid and flexible editorial approaches than are possible with traditional print-based publications. But Psychology is also a serious journal, with peer reviews of articles, professional presentation, and full archival status. So, young and yet serious: much like the field the journal covers, the evolving relationship between people and ICT (information and communication technology). Both sides of this relationship impact on the other. Technologies must be designed, or evolve, in ways that match the capabilities and limitations of the people who use them. On the other hand, technology changes people; in fact it could be argued that without ICT, understood in its widest sense, people wouldn't be people. While this has always been the case, the rapid pace of technological innovation makes the changes to people brought by technological advances vastly more salient than in earlier times.

The target topic for this issue of Psychology, Future Interfaces, puts the focus firmly on the potential changes to everyday life that new ICT will bring to us. As the Call for Papers indicated, a clear trend is to design and develop technologies for specific purposes, needs and situations, such as mobility, group collaboration, disability, age, and so on. And of course another trend, as exemplified by this journal, is towards virtualisation; not only of information, but also of processes, organizations and activity patterns. The call produced a rich

response in the form of a large set of submitted papers from around the world, on a diverse range of topics highly relevant to the theme of the issue. So much so that accepted papers had to be distributed over two issues; in issues three and four you will find a total of 10 papers on the special theme of Future Interfaces, and 3 papers of more general interest, from Finland, France, Germany, Italy, Japan, Spain, Sweden and the UK.

Issue number four includes the second group of papers in response to the call along with a paper of general interest. A common thread in discussions of future interactions is the role of the body, and its relationship to information and the mind. Two contributions appearing in issue number three addressed these aspects (Waterworth and al.; Fukuda and Bubb). Directly interfacing the brain to the computer takes us a step further. *Beverina et al.* in this issue describe how Brain Computer Interfaces (BCIs) provide a new communication option for those suffering from neuromuscular impairment. By using data from electrodes placed on the scalp, it is possible for users mentally to control events at the interface. In principle, this is similar to work that has enabled monkeys to control mechanical limbs. The great advantage is that, unlike that work, this does not involve implanting electrodes in the brain. The drawback is the need for a high level of conscious attention from users. The authors suggest that more vivid and emotionally engaging interactions, using virtual reality and competition between users to elicit involvement and surprise, can enhance overall performance. This is in principle easy to do, but will take imagination and flair on the part of designers. *Lessiter et al.* describe work on a relatively down-to-earth topic, the design of remote controls for digital terrestrial television. They provide a

very useful and detailed account of usability methods to assess different designs of control, but the approach can be applied to a wide range of interactive media products and services. *Gamberini et al.* present a review paper on affective interfaces showing the attempts so far at interfacing computers with human emotional states.

*Namatame et al.* present a case study of webpage design for the hearing impaired. Using data on operation flow, length of stay, and eye and mouse movements, they identified two types of hearing-impaired users, those who are primarily text-oriented and those who are more picture-oriented. This is important work that has the potential to expand current guidelines for the design of webpages for the hearing impaired. The paper by *Jastrzebska-Fraczek and Bubb* describes EKIDES, a general design tool, based on a database of ergonomic guidelines, for reference and evaluation of work places and products. *Breite and Vanharnta* studied students attempting to set up and manage a virtual company by means of collaboration.

These twin issues are closed by an invited contribution of general interest by Waterworth. The paper suggests that the trend towards virtual realization of information provides the potential for designing computer-supported creative spaces, based on examples from music tuition and surgery planning.

It has been a great pleasure for me to serve as Guest Editor of these two themed issues of Psychology journal. The collection of papers in this and the preceding issue provide clear and fascinating pointers towards Future Interfaces.

John Waterworth  
Umeå, Sweden



## User adaptive BCIs: SSVEP and P300 based interfaces

Fabrizio Beverina<sup>^</sup> ♦, Giorgio Palmas ♦, Stefano Silvoni ♦, Francesco Piccione ♣,  
Silvio Giove ♥

♦ .STMicroelectronics -  
Advanced System  
Technology – Agrate  
Br.(ITALY)

♣ Dept. of Neuro-  
rehabilitation, S.Camillo  
Hospital, Venice (ITALY)

♥ Dept. of Applied  
Mathematics, University  
of Venice (ITALY)

---

### ABSTRACT

Brain Computer Interfaces (BCI) represent a new communication option for those suffering from neuromuscular impairment that prevents them from using conventional augmented communication methods. This new approach has been developing quickly during the last few years, thanks to the increasing of computational power and the new algorithms for signal processing (Independent Component Analysis, Wavelets decomposition, Support Vector Machine etc.) that can be applied to the studies made on brain waves. Here follows two methodologies of approach based on making the computer adapt to the human brain activity and not vice-versa. The P300 and the SSVEP based BCIs, here presented, have the characteristics of not demanding specific training to the user.

Keywords: *Brain Computer Interfaces, signal processing, human computer interactions.*

Received 27 October 2003; received in revised form 12 December 2003; accepted 19 December 2003.

### 1. Introduction

A good hundred years after the first discoveries were made on the brain's electrical activity, Jacques Vidal published an innovative work explaining how to use the brain's electrical potentials for building a mental prosthesis [Vidal, 1973,1977]. This was the starting point for BCI research. Nowadays, about thirty research groups are following this approach for interfacing the computer [Kronegg, 2003].

In the first international meeting on BCI technology, which took place in 1999, at the Rensselaerville Institute of Albany (New York), Jonathan R. Wolpaw formalized the definition of the BCI system :

*"A brain-computer interface (BCI) is a communication or control system in which the user's messages or commands do not depend on the brain's normal output channels. That is, the message is not carried by nerves and muscles, and, furthermore,*

---

<sup>^</sup> Corresponding Author:  
Fabrizio Beverina,  
E-mail: fabrizio.beverina@st.com

*neuromuscular activity is not needed to produce the activity that does carry the message" [Wolpaw et al., 2000a].*

Through this definition, BCI systems appear as a possible and sometimes unique mode of communication for people with severe neuromuscular disorders like spinal cord injury or cerebral palsy. As a matter of fact, such neural diseases can break the slim and fragile line between thoughts and actions. In these cases, neither medicine nor surgery can be of any use to give back to the person the control of his/her body. However, utilizing the residuals functions of the brain, it seems possible to give back a hope of communication to these people.

The human brain has an intensive chemical and electrical activity, partially characterized by particular electrical patterns, which occur at specific times and at well-localized brain sites. All of that is observable with a certain level of repeatability under well-defined environmental conditions. These simple physiological issues can lead to the development of new systems to communicate.

Here follows a list of the most utilized electrical activities of the brain for BCI:

- $\beta$  and  $\mu$  Rhythms

These electrical activities are observable inside a frequency range from 8 Hz to 12 Hz ( $\mu$ ) and 12 Hz to 30 Hz ( $\beta$ ). These signals are associated with those cortical areas most directly connected to the brain's motor output and can be willingly modulated with an imaginary mental movement for example. The increases/decreases of this rhythm have been used several times as a support for a BCI. [Pfurtscheller, 1989], [Pfurtscheller and Lopes Da Silva, 1999], [Mc Farland et al., 2000].

- P300 Evoked Potential

This wave is a late appearing component of an Event Related Potential (ERP) which can be auditory, visual or somatosensory. It has a latency of about 300 ms and is elicited by rare or significant stimuli, when these are interspersed with frequent or routine stimuli. Its amplitude is strongly related to the unpredictability of the stimulus, the more unforeseeable the stimulus is, the higher is the amplitude. This

particular wave has been used to make the subject chose between different stimuli [Farwell and Donchin , 1988] , [Donchin et al., 2000].

- Visual Evoked Potential

They are ERPs with short latency that represent the exogenous response of the brain to a rapid visual stimulus. They are characterized by a negative peak around 100ms (N1) by a following positive peak around 200ms (P2). Since Vidal's innovative works in the early 70's, these potentials are being used as clues indicating the direction of the user's gaze [Vidal, 1973, 1977], [Sutter, 1992].

- Steady-State Visual Evoked Potentials (SSVEP)

These signals are natural responses for visual stimulations at specific frequencies. When the retina is excited by a visual stimulus ranging from 3.5 Hz to 75 Hz, the brain generates an electrical activity at the same (or multiples of the) frequency of the visual stimulus. They are used for understanding which stimulus the subject is looking at in case of stimuli with different flashing frequency. [Morgan et al., 1996b],[Muller et al., 1997]

- Slow cortical potentials (SCP)

These electrical activities are slow potential variations generated in the cortex after 0.5 – 10.0s. Negative SCPs are generally produced by movement, instead positive SCPs are associated with reduced cortical activation. Bimbaumer and his colleagues [Bimbaumer et al., 1990] demonstrated that people, adequately trained, can control these potentials and use them to control the movement of a cursor on the screen.

BCIs have been investigated from different perspectives. For the sake of simplicity we will split them into two methods, direct (invasive) and indirect (non-invasive).

In the first approach, recording devices are required inside the brain. This makes it possible to capture the electrical patterns near their sources. This solution requires

challenging technological, scientific and psychological competences. For example, highly complex implanted micro-sensors, biologically compatible, are pioneering researches. Algorithms which process huge amounts of data generated by neurons at high rates (>20 kHz), and which filter and classify in real-time the brain's electrical activity, are another critical point for the BCI direct approach. Furthermore, from a psychological point of view, it is not yet sure that anyone in the future will approve to receive an implanted device in the brain.

Probably the work of neurologist Philip Kennedy and his colleagues [Kennedy et al., 2000] is the most impressive example for the direct approach. Johnny Ray, a patient of Kennedy, lived and talked to the world using cortical implanted electrodes. Other researchers, like Chapin from the Medical College of Pennsylvania and Nicolelis at the Duke University [Chapin et al., 1999], have used implanted electrodes inside monkeys' brains to control a robot-arm from distance.

The non-invasive way for BCI is less technologically and psychologically demanding. It requires sensors (electrodes) placed on the scalp to record electrical patterns. The necessary experimental set-up (i.e., electrodes, amplifiers, medical competence...) to carry out the indirect approach is more available to laboratories today than the require set-up for the invasive one. The disadvantage of listening to the brain activity from the outside of the scalp lies in the very low quality of the signals, due to the damping of the electrical activity signals on their way to the electrodes.

This non-invasive and easy to set-up approach has been used to study several sets of electrical patterns. These sets can be grouped into two main classes: the electrical patterns evoked by external stimuli (i.e., a blinking arrow) and the electrical patterns generated by means of wilful execution of particular cognitive tasks (i.e., imaging a spinning cube). It is the research goal or the physical boundaries imposed by the patients that decide the use of one or the other kind of electrical pattern.

The usage of this approach has given good results, even if there are no real and useful applications yet. To cite the most published works inspired by it, it's worthwhile to highlight the pioneering researches of Prof. Birbaumer. He was one of the first to use brain waves, influenced by the human will, to drive a speller [Birbaumer et al., 2003]. Another well-known researcher is Prof. Wolpaw. He developed a BCI based on  $\mu$  and  $\beta$  EEG rhythms [Wolpaw et al., 1991]. The patient, after specific training, is able to move a cursor up and down just modifying those rhythms. A different strategy was proposed by the psychologist Donchin [Farwell and Donchin, 1988] using the P300 wave to control a speller. Whereas the previously cited works (Birbaumer and

Wolpaw) are based on the user's ability to control brain behavior, in Donchin's method a quasi-uncontrollable brain signal, the P300, is used.

The general idea of Donchin's solution is that the patient is able to generate this signal without any training. This is due to the fact that the P300 is the brain's response to an unexpected or surprising event and is generated naturally. Donchin has developed a BCI system able to detect an elicited P300 by signal averaging techniques (to reduce the noise) and used a specific method to speed up the overall performance.

The SSVEP (Steady State Visual Evoked Potential) activity is another successfully investigated brain signal. As presented previously, SSVEP [Morgan et al., 1996b], [Muller M.M. et al., 1997] is the natural brain response when the retina is excited by flickering visual stimuli. The SSVEP signals are strongly modulated by a selective spatial attention process: these signals are well defined within the extent, delimited by the user's visual attention. Outside this area, flickering visual stimuli don't generate the same meaningful activity.

Starting from the P300 and SSVEP works and trying to overcome their boundaries, two BCI systems have been developed in our laboratory.

## **2. Materials and methods**

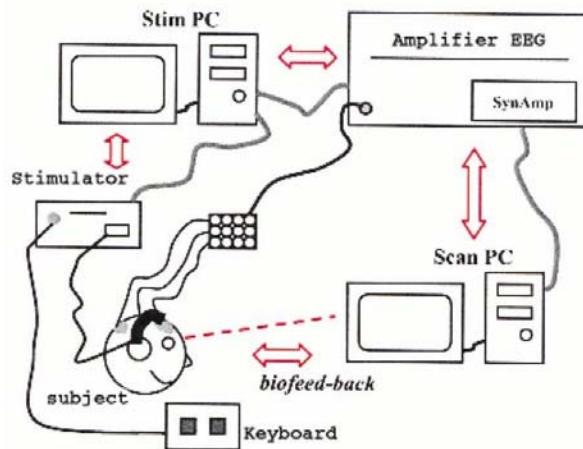
A fundamental aspect for a human machine interaction system, such as BCI, is the need for a proper development environment that allows a real time interaction between the subject and the machine. (This necessity is directed by the fact that the users need a short response time to keep up their attention level). For this purpose we created a flexible modular environment useful to develop and to experiment various BCIs

There are essentially two parts that every BCI system needs:

1. A dedicated hardware system that manages the stimulation, the EEG acquisition (electrodes and amplifier samplers etc) and the feedback (visual, acoustic, haptic ...)
2. A system that deals with advanced signal processing and interpretation.

We chose to use for the first part a commercial system for EEG analysis and for the signal analysis tool we opted for Matlab, which gives a great flexibility and easiness to the developing of the algorithms. In this system we studied two kinds of BCI: a first one based on an ERP (P300), and a second one based on SSVEP.

Both the BCIs studied share the video interface and the acquisition in the first elaboration part.



**Figure 1:** System for BCI development

## 2.1 The system: Hardware

The core of the first part is the NeuroScan System. This is a commercial system composed of :

- Neurostim: PC that manages the visual and acoustic stimulation for the subjects.
- SynAmps: which amplifies and filters high frequency components of the EEG signals.
- ScanPc: PC that manages the functioning of the whole system for acquisition, processing and visualization as well as the exchange of EEG data with external programs ( ex.; Matlab). The possibility to exchange data with other programs is one of the main reasons to use Neuroscan for a research and development project.
- A set of electrodes (that can be fixed on a "quick cap") to be fitted on the subject's scalp.
- A 4-button keyboard whose stimulations are recorded synchronously with EEG by the SynAmp.

## 2.2 The system: Software

Thanks to NeuroScan's characteristics, which allow data exchange between different processes, we developed the artificial intelligence algorithms needed by our BCIs in Matlab and C++ code. This choice was determined by the notable flexibility of the software written in Matlab whose characteristics permit easy trials of new algorithms. The proposed system (Hardware and Software) carries out some fundamental operations in order to test the detection of the electrical potentials in real time. We use for this reason, the same system for both the BCIs we are studying.

### **2.3 P300 based BCI**

For a P300-based BCI, we need to elicit an ERP. Therefore, to start with, the subject must be stimulated with an appropriate interface. Then the generated electrical potentials should be recorded and eventually processed in real time. Here follows the list of the main characteristics necessary to achieve these three steps (stimulation, recording and processing):

- The definition of an elicitation paradigm to evoke this ERP, using different kind of stimulations (acoustic or visuals modality according to the user's capabilities).
- Neuroscan that performs on-line acquisition of EEG data, synchronized with the stimuli (collected as ERPs epochs).
- A processing procedure that allows to reduce noise and enforce P300-related information.
- A pattern recognition algorithm that permits to check the absence or the presence of the P300 wave in the recorded ERP epochs.
- The procedures that carries out a epoch labelling, according to the previous stimulus type (target or non-target), in order to use them during off-line operations.
- A feedback mechanism to the subject, which sends him/her a visible signal on the monitor correlated to the recorded epoch.
- A pattern recognition algorithms that, using the labelled epochs previously collected, adapts its rules according to the user characteristics.

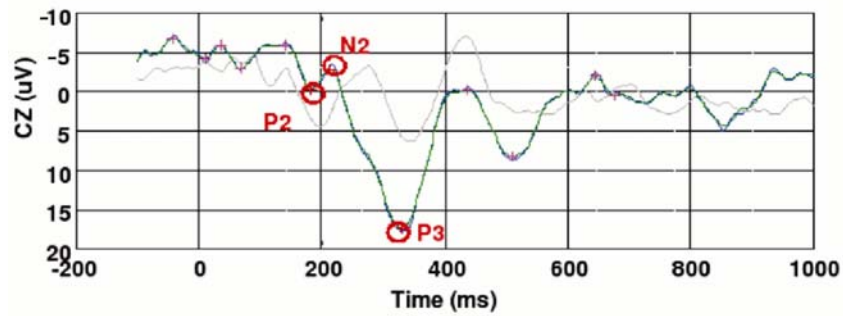


Figure 2: P300 wave

### 2.3.1 Algorithms for P300 studies: ICA and Support Vector Machine

#### Classification Algorithm

In order to understand whether a P300 pattern has been generated by the visual stimulus, a Support Vector Machine [Vapnik, 1995] was developed. Generally speaking, the Support Vector Machine implements the following idea: it maps the input vector  $x$  into a high-dimensional feature space  $Z$  through some non-linear mapping  $K$ , chosen a priori. In this space, a hyper plane is constructed. This hyper plane, in our case, separates the P300 patterns from the non-P300 patterns.

The core of a SVM classifier is the kernel function

$$K(x) \Rightarrow Z \quad (1)$$

One of the most used kernel functions, as in our experimental sessions, is the radial basis kernel

$$K(x) = e^{-\gamma x^2} \quad (2)$$

Using the SVM classifier, the following issues have been observed:

- A fast learning rate: typically a few seconds are sufficient to learn the training set.
- Quite coherent results between the off-line training, the testing phase and the real-time phase.
- Good numerical stability.

### ICA algorithm

The raw signal, acquired by NeuroScan, follows four processing steps: first, all the signals recorded by scalp electrodes are processed in order to obtain a set of independent components. Since the locations of the brain that generate ERP cannot be determined easily by the scalp recordings (resolution problem), many algorithms have been studied in order to separate each signal in a set of independent sources (i.e., originating from different areas). One of the most promising algorithms is the so-called Independent Component Analysis (ICA) [Comon, 1994]. ICA determines what spatially fixed and temporally independent component activations compose an observed time-varying response, without attempting to directly specify where in the brain these activations arise. Practically the problem that ICA solves, is to recover sources from their instantaneous mixture without any previous knowledge of the sources and the mixing channel. Differently from Principal Component Analysis PCA that finds components that are uncorrelated, ICA is a much stronger criterion because it is based on statistical moments of a higher order, so ICA requires more than the uncorrelatedness of the components. The most general case can be so characterized: we consider  $n$  unknown sources signals  $\mathbf{s}_i(\mathbf{t})$ ,  $i=1, \dots, n$ , which are mutually independent, and we model the sensor's output as

$$\underline{s(t)} = Ax(t) \quad (3)$$

where  $\mathbf{A}$  is an unknown non-singular mixing matrix,  $\mathbf{x}(\mathbf{t}) = [\mathbf{x}_1(\mathbf{t}), \dots, \mathbf{x}_n(\mathbf{t})]^T$ ,  $\mathbf{s}(\mathbf{t}) = [\mathbf{s}_1(\mathbf{t}), \dots, \mathbf{s}_n(\mathbf{t})]^T$ . With no knowledge of the source signals and the mixing matrix, we want to recover the original signals from the observed signals  $\mathbf{x}(\mathbf{t})$  by the following linear transformation:

$$\mathbf{y}(\mathbf{t}) = W\mathbf{x}(\mathbf{t}) \quad (4)$$

where  $\mathbf{y}(\mathbf{t}) = [\mathbf{y}_1(\mathbf{t}), \dots, \mathbf{y}_n(\mathbf{t})]^T$  and  $\mathbf{W}$  is the un-mixing matrix. Of course it is impossible to find the original sources without ambiguity, because they are not identifiable in a strictly statistical sense. However, up to some permutation, it is possible to obtain  $\mathbf{c}_i \mathbf{s}_i(\mathbf{t})$  where  $\mathbf{c}_i$  are unknown non-zero scalar factors. In order to separate the components, ICA works on a learning algorithm that minimizes the dependency

between the output components: such a dependency is measured by the Kullback-Leibler divergence (5) between the joint and the product of the marginal distributions of the output:

$$D(W) = \int p(y) \log \frac{p(y)}{\prod_{a=1}^n p_a(y_a)} dy \quad (5)$$

Where  $p_a(y_a)$  is the marginal probability density function (pdf). To perform this, some hypotheses are implicit and a training algorithm is needed to find the right un-mixing matrix  $\mathbf{W}$ . The hypotheses are the following:

1. The signals recorded from the electrodes are an instantaneous mix of  $\mathbf{n}$  statistically independent sources. This implies that the coefficients of the mixing matrix  $\mathbf{A}$  are linear and time-independent. From a physiological point of view this is equivalent to saying that the sum of the electrical potentials coming from different areas of the brain on the scalp electrodes, is linear. To be more precise, it is not the result of non-linear distortion or temporal convolution of the sources.
2. The number of sources  $\mathbf{n}$  does not exceed the number of electrodes. In physiology this means that the areas involved are stable and in a defined number.
3. The sources and the mixing process are stationary, they don't change their statistical properties in time.

The first hypothesis is well confirmed in literature [Makeig et al., 1997], [Jung et al., 2001]. The second hypothesis doesn't represent a problem because we can take as many sources as we want (in theory at least), in order to have the number of sources smaller than the number of electrodes. The third one is generally not verified but we can overcome this problem if we choose a time interval, small enough to consider with a good approximation, the signal stationary. According to these considerations we can apply the independent component analysis computation.

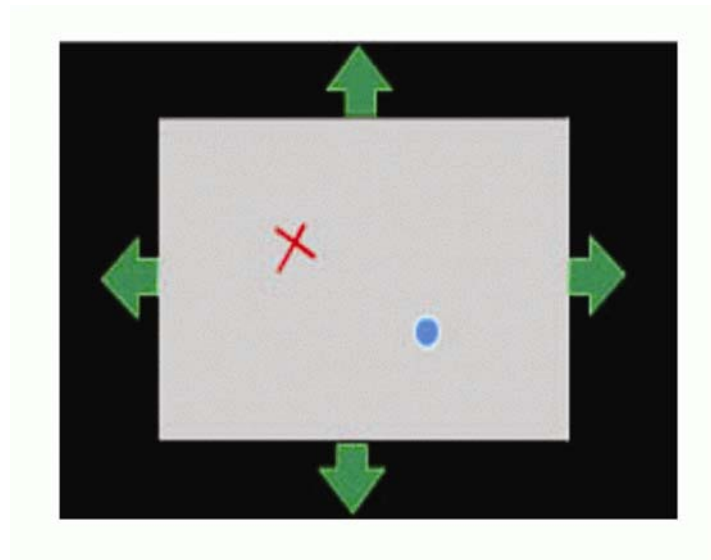
### 2.3.2 P300 based BCI: protocol and interface

The protocol used to test the proposed BCI device, consists of a P300 wave elicitation paradigm. It can be divided into two main phases, called learning and testing. Through this protocol we want to reach the following objectives:

1. Solving the "ad personam" system adaptation;
2. Quantifying the system performance;
3. Giving a visual bio-feedback to the subject.

To elicit the ERP (P300 wave) we submit the subject to a random sequence of visual stimuli on a computer screen, using a complex odd-ball paradigm. Inside our graphic interface, each stimulus has the shape of an arrow to give a directional meaning: up, right, down and left, for the total amount of four different stimuli (see Fig. 3). A stimulus consists of the single flash of an arrow lasting 150ms. The inter-stimulus interval is of 2.5 s, and the upper bound of a sequence of stimuli is of 90. Since the subject's task is to reach the goal (the red cross) with the movements of an object (the blue ball) he/she has to decide which direction (i.e.: the blinking arrow) is interested in and keep his/her attention on it. The stimulus desired by the user is called *target* (that is the one that would allow the object's movement in the direction chosen by the user), otherwise it is called *non-target*.

The main hypothesis, that we will use later, is that every target stimulus elicits the P300 wave.



**Figure 3:** P300 interface: testing phase

### 2.3.2.1 Learning phase

During the learning phase, the subject learns to reach the red cross using the arrows.

For this phase two strategies can be followed, depending on the capacity of the subject to use the keyboard or not. For those who can't, a predefined path appears on the screen that they are taught to follow looking at the correct arrows. The object's movements are controlled by the software, thus every time that the subject receives a target stimulus (the next direction along the predefined path) the object makes a single step toward the cross, otherwise it does not make any step.

If the subject is able to move (one hand at least), we can apply another strategy for the learning phase: the person informs the system when he receives a target stimulus by pressing a key, so that every time he presses the key, the object on the screen makes a single step according to received stimulus.

Both protocols make the subject believe that every time he/she wants to move the ball, the machine is able to *read* the subject's will and moving consequently the object. This is done to recreate the closest situation to the testing phase. At the end of the learning phase, the system has recorded a set of signals (ERP epochs) that should be similar to the ones it will deal with during the testing session. In such a way the adaptive algorithm can learn the specific P300 wave of each subject. It can then discriminate the ERPs data epochs between target stimuli (the one which elicits P300s) and non-target stimuli. To do this, some off-line operations must be performed:

- trace filtering (low-pass filtering);
- ICA decomposition, to extract the un-mixing matrix  $\mathbf{W}$ ;
- feature extraction and normalization;
- Support Vector Machine training.

Once the system completes the training, the testing phase can start. The above processing and training procedure can be performed after every testing session, in order to improve the system performance.

### 2.3.2.2 Test phase

During the testing phase, the subject actually performs the task, like in the learning session, but with a substantial difference: the object's movements are controlled by the output of the adaptive algorithm, trained on the subject's previous tracks. So, in this

phase, the user can see directly on the screen the result of the classification algorithm (the movement of the ball).

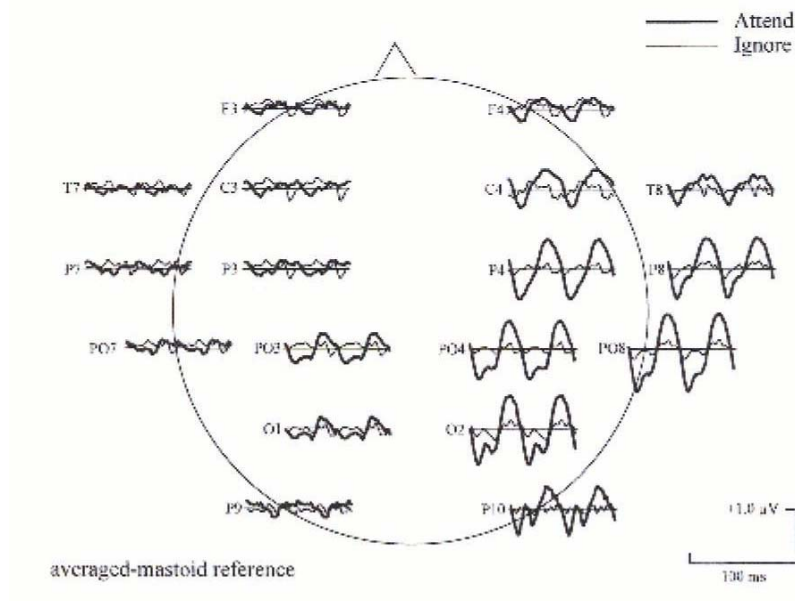
The recognition algorithm assesses the presence of the P300 wave into the single-sweep tracks related to every stimulus. If a P300 is detected, then the system moves the object on the graphical interface according to the stimulus just received by the subject, otherwise the object stays still. If the subject is interested in a particular direction, left for example, he normally will elicit a P300 wave only when he sees the left arrow flashing. If the classification algorithm works correctly, the subject should see the ball moving one step left. Otherwise, he/she will see the ball moving in a direction he/she is not interested in or no movement at all. The subject will consider as a reward the movement of the ball in the desired direction (positive biofeedback). Otherwise, a non-desired movement (negative biofeedback) will push the subject to concentrate more on the stimulus, trying to control it better. As in the learning session, the subject informs the system about target stimulus by pressing a key if he/she can. This allows track labelling and, successively, re-training the system using information retrieved during the testing session.

## **2.4 SSVEP-based BCI: protocol and interface**

The second BCI we are studying is based on the steady-state visual evoked potential (SSVEP). The SSVEP is a continuous and periodic signal, elicited by visual stimulus flickering in the frequency range between 3.5 Hz-75 Hz. SSVEP is described as a near sinusoidal signal oscillating at the same, or multiple, stimulus frequency and it's particularly detectable in the occipital-parietal region of the skull. These signals are readily quantifiable in the frequency domain and can be easily extracted from background electroencephalogram noise.

### **2.4.1 Assessment studies**

The first step of our research consisted in assessing the feasibility of creating a screen based visual interface able to elicit SSVEP signals. The visual stimulus is given by some items of the graphic interface that flicker at different frequencies. An important part of the work concerned with the definition of the item features, such as light intensity, colour, shape, dimension and flickering frequency. We performed a set of experiments in order to understand how to increase the amplitude of the recorded brain responses to the flickering stimuli.



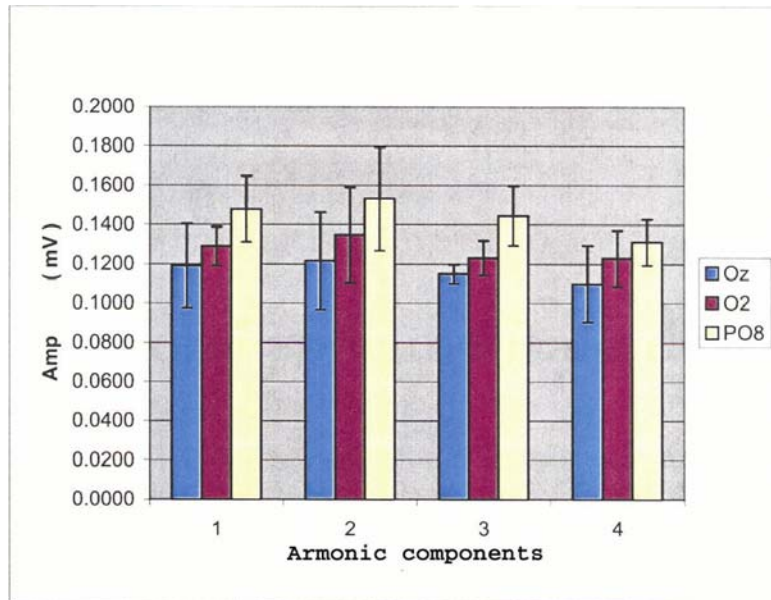
**Figure 4:** SSVEP distribution over the scalp

We reached these conclusion:

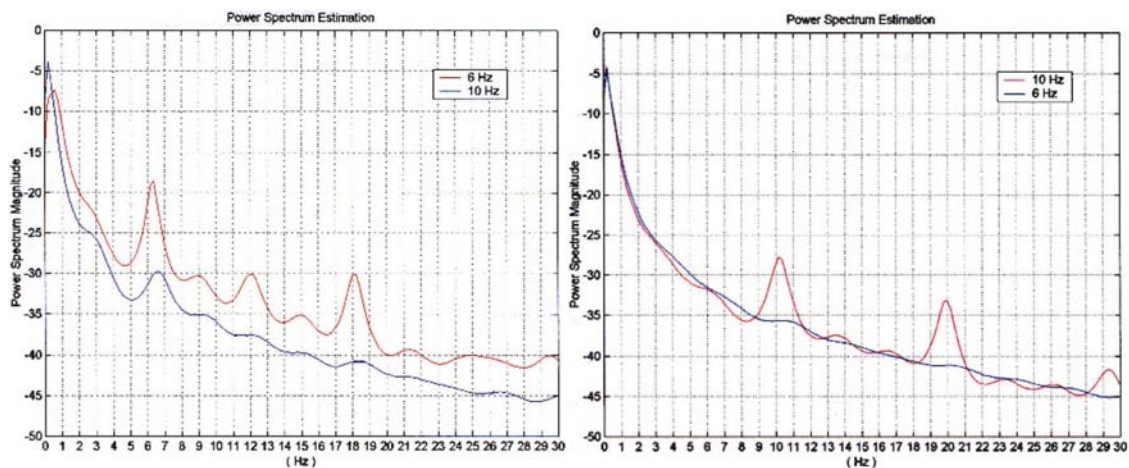
- Confirming what is said in literature [Regan, 1989], the occipital electrodes, specially *PO8* (figure 4), are the one where the response to the stimuli is higher. This was verified through a set of experiments, made on 5 subjects (figure 5). From the figure, It can be seen also that the second harmonic is the one which has the greatest amplitude [Ventura, 2002 ].
- We confirm the fact that the increase of the SSVEP amplitude is directly correlated with the spatial attention processes [Muller and Hillyard, 1997], [Morgan et al., 1996], [Silberstein, 1990]. To test this hypothesis we associated to the action of concentrating on the stimulus, a cognitive task like counting a blue spot that appears randomly on the regions where the stimuli are flashing. This task increases the amplitude on the SSVEP recorded.
- It has been shown [Blanchard and Epstein, 2000] that it is possible through the feedback to make some subjects increment some brain wave activities. For such a purpose we put a bar next to every flashing symbol, showing in real time the brain activity correlated to that flashing frequency.

Thanks to this feedback we noticed (with a high statistical significance ) an increase of the wave amplitude of 10-20% (Table 3) [Ventura, 2002 ].

- We found also that the couple of frequencies easier to discriminate was the couple 6 Hz and 10 Hz. This couple was the one in which each frequency interferes less with the other (as can be see on the figure 6).



**Figure 5:** Harmonic Components of 6 Hz (left) and 10 Hz (right) stimulation for the different electrodes: as it can be see the PO8 is the one with higher amplitudes



**Figure 6:** Power spectrum of SSVEP for the couple 6-10 Hz when the subject is looking at the flashing stimulus at 6 Hz on (figure on the left ) and at 10 Hz (figure on the right)

#### **2.4.2 Algorithms for SSVEP studies: MA, ALE, Subspace Averaging**

The EEG signal is processed with a method named *Subspace Averaging* [Davila and Srebro,2000],[Davila et al.,1998].

This technique is a combined application of two signal elaboration methods called *Signal Averaging* and *Signal Space Projection*.

The latter one was introduced in 1987 by Ilmoniemi and is used to separate the EEG signal from a generic noise signal (cardiac or oculographic artifacts). See [Ilmoniemi and Williamson, 1987], [Ilmoniemi et al., 1987] and [Ilmoniemi and Uusitalo, 1997]. This technique has been widely applied in the telecommunication field and, recently, has been applied in the analysis of physiological signals such as EEG or MEG. The computational simplicity and the short processing time, make this method preferable to others and allow an easy and efficient implementation, especially for on-line application.

The classification has been implemented by a linear threshold algorithm, that recognizes the SSVEP signal if its amplitude exceeds the threshold value for, at least, 2 seconds.

#### **2.4.3 Protocol**

The experimental protocol has been divided in two phases: training and testing. The training phase has the purpose of establishing the parameters of the typical SSVEP activity of every subject (maximal amplitude) in order to fix the threshold for the classification part. The subject sits at 70 cm distance from the screen and the electrodes (Oz,O2,PO8) are applied. In this first phase, during which the feedback signal is not given to the subject, the subject has to focus his/her attention for 20 s on the left flashing symbol, after 2s at the centre of the screen and finally for 20 s at the right flashing symbol. To check the exact moment in which the subject changes direction, he/she has to press one of the three buttons according to which flashing light the subject is looking at.

In this second phase, the subject has been told to focus his/her attention on the flashing arrows, following the sequence of the directions to look at on top screen (D for Right, S for Left as it can see on the figure 7). The current letter (corresponding to a direction) will change colour if the subject manages to keep the amplitude (according to the magnitude bar) of the brain wave corresponding to the frequency of the stimulus he is looking at, for at least 2 seconds. We measure, for each trial, the correct selections, the errors, the null events (defined as the event when both the amplitude

bars reach the top for two seconds) and the time to complete the trial. After each trial the subject's amplitude thresholds are recalculated adaptively in order to track the changes on the individual behaviour. This simple game is an easy trick to keep the subject always engaged.

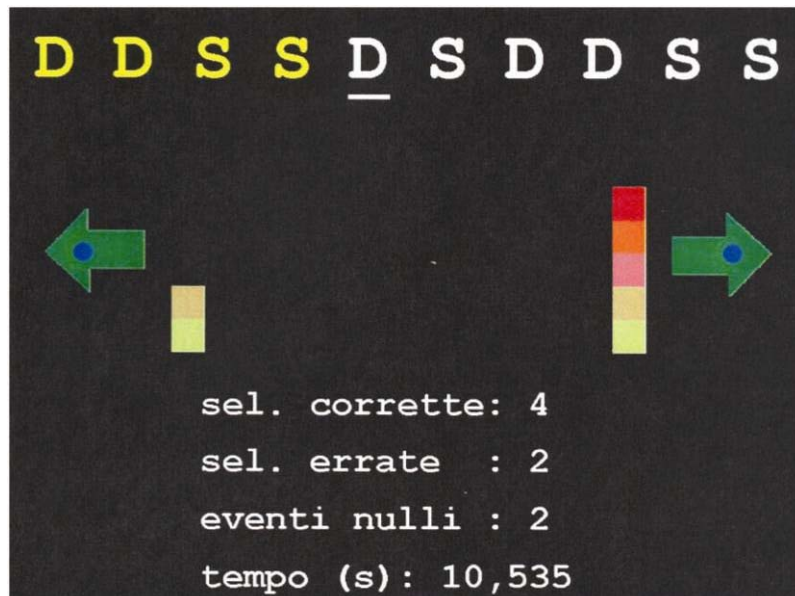


Figure 7: SSVEP visual interface during the testing phase

### 3. Results

#### 3.1 P300

P300 based BCI has now achieved a good developing level, in terms of both the system architectural choices and the early results. Since the beginning of our project, 3 years ago, the whole system has changed many times. We started with an acoustic protocol, but we had to drop this approach because of the difficulty to handle a stimulus with a high semantic component (P300 had latency and shape that depended too much on the particular stimulus). The evolution of the system passed through hundreds of tests on healthy and pathological subjects and through many modifications of the classification algorithm (we started from a simple Bayesian classifier, we shifted to neural networks and finally we chose a Support Vector Machine).

The following tables sum up briefly the results we had. The first one, related to the neural network approach, (Table 2) is divided into healthy and pathological subject,

and the second one represents the new promising SVM approach (Table 3). In the tables we define:

- *Performance*: the percentage of the exact classification;
- *Instr./min.*: the number of correct instructions per minute;
- *Err P300*: error in classification of P300 waves;
- *Err. Not P300*: error in classification of non-P300 waves;
- *Err. tot.*: total error.

	Healthy		Pathological	
Subject N	7		5	
Performance	66.8		56.7	
Instr./min.	3.39		3.59	
Age: Mean- Var	33	22-43	40	30-53
Err.P300: Mean –StD	0.4966	0.16	0.5864	0.1544
Err. Not P300: Mean –StD	0.2707	0.0794	0.3665	0.106
Err. tot.: Mean –StD	0.3318	0.0777	0.4327	0.986

**Table 2:** P300 results: NN approach healthy Vs Pathological

Subject N	5	
Performance	71.1	
Instr./min.	5.34	
Age: Mean -Var	32	23-45
Err. P300: Mean -StD	0.59	0.109
Err. Not P300: Mean -StD	0.123	0.0291
Err. tot.: Mean -StD	0.2888	0.0397

**Table 3:** P300 results: SVM approach.

It is necessary to keep in mind the way our interface works in order to interpret these data tables. A random stimuli sequence is presented to the user, who has to concentrate only on the one is interested in. It appears evident that the most critical error is due to the misinterpretation of the not-P300 wave. This causes a wrong action on the interface while a P300 wrong classification doesn't create any output. That is the reason why we tend to minimize the error on not-P300 classification without caring of the corresponding error on P300. Using such a strategy, the user will only experience a great difficulty to move the object in the desired direction and this sensation will push him/her to concentrate more.

A simply way to increase the bit-rate without changing the classifier, consists in reducing the interstimulus time. Presently, the interstimulus period is of 2.5 seconds whereas the computation time is of 400ms. This implies that we can reduce the interstimulus at most to 400 and this will bring up the bit-rate magnitude of 5 times. This solution will be investigated accurately because it hides many interrogatives about what the user reaction will be to such high frequency stimulation.

### **3.2 SSVEP**

The experimental sessions were run at the San Camillo Hospital in Venice. 5 healthy subjects, aged between 24 and 32 years old were tested. The experimental sessions validated the feedback efficacy in improving the man-machine communication process. The importance of the feedback signal is particularly evident in its capability of adapting the SSVEP response of the subject to the classification algorithm request. This adaptation process has been observed in all the subjects. As it can be seen from the table 4 there is a significant increase of the amplitude of SSVEP when using the biofeedback bar. Another interesting result is that all the subjects show a decrease of the signal amplitude after 7-8 trials. When asked, the subjects replied that this interface is very tiring and that they find it hard to concentrate. The last observation, according to the results obtained in the testing phases, is the accuracy of this kind of interface (the average accuracy is of 95.71% with a standard deviation of 2.73%). This high accuracy was obtained because we kept the threshold for selecting the stimulus high in order to decrease the false negative classification that is very harmful in BCI systems; in other words, only 4.3% of the instructions are misinterpreted. Practically this results in increased difficulty for the

subject that has to keep concentrated to reach the selection. In spite of this increment of difficulty, the subject achieved a satisfactory communication rate (10 instructions in 97.16 seconds).

Freq. Stim.	Subject	Percent. Incr.	P
10 Hz	1	10.93	0.085
	2	28.22	0.045
	3	10.01	0.013
	4	21.35	0.004
	5	10.09	0.034
6 Hz	1	9.08	0.025
	2	34.71	0.027
	3	14.51	0.046
	4	10.67	0.01
	5	19.58	0.029

**Table 4:** Tables comparing the percentual increase of amplitude for every subject and the statistical significativity *P*.

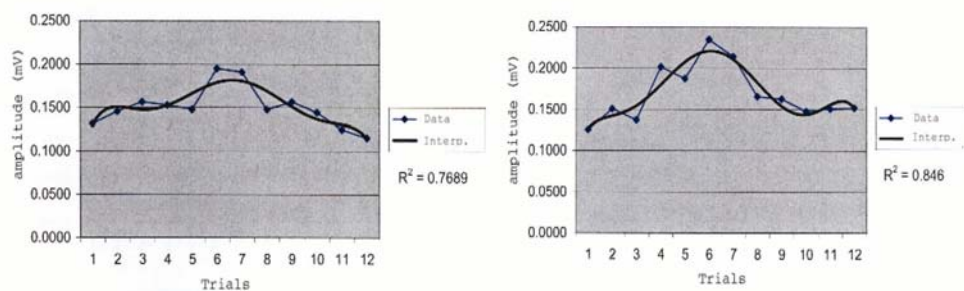
#### 4. Discussion

Both BCI systems studied have the advantage that they do not need a particular training (Birbaumer, Wolpaw) because they exploit a natural brain behaviour. Thanks to the feedback, subjects can learn how to improve the communication rate. We see also that the performance can be greatly improved by increasing the engagement of the subject in the task. It turns out that the subject participation and willingness is as important as the classification tools of the system. It is evident that the more the subject is concentrated on the task, the higher will be the bit-rate of whole systems. Therefore the study of physical interfaces and of the stimulating strategies become critical.

Many ways can be explored to improve the whole system performance:

- virtual reality involvement;
- competition between subject in a video-game, one after the other or one against the other;
- reward/penalty distribution strategy improvement.

The drawback of such an attention-dependent system is that it is very tiring and consequently the performances decrease in time (as it can be seen on the figure 8 referring to one SSVEP session).



**Figure 8:** SSVEP amplitude ( 6 Hz, 10 Hz) decreasing during the testing session for 2 subjects

## 5. Towards the future

P300 interface and SSVEP interface are still at this stage an open research topic. The results we show are in their early stages but some interesting points can be observed.

In general, it's quite clear that the two interfaces require a good attention level by their users. So, a particular effort has been dedicated to all the aspects regarding the user's side of the interfaces. Two simple examples are, the colours used for P300 interface arrows and the feedback channel studied for the SSVEP interface.

The classifier performance or the ICA filtering is not as important as the user's feeling toward the interfaces. The more the user is aware of the task proposed by the BCI, the more the brain activity will be recognizable by the machine.

To improve the involvement of the subject, which has been seen as a critical point for the performance of both the interfaces, many solutions can be explored. A massive use of Virtual Reality is probably the first step toward a real and usable BCI, especially for all the aspects regarding the multiple sensorial stimulations provided by this

technology. Another really important issue is the user's emotional involvement when he/she uses the interface.

A sort of game-competition, maybe mediated in a Virtual Reality environment, could amplify and focus the brain activity toward a specific target. As an example, the P300 wave is strongly related to the level of surprise: it's quite easy to imagine an engaging game, with colours and sounds, able to generate unexpected events.

## 6. Bibliography

- Birbaumer N., Elbert T. Canavan A.G.M. Roch B., (1990). *Slow potentials of the cerebral cortex and behaviour*, *Physiol. Rev.*,70,1-4.
- Birbaumer N., Hinterberger T., Kubler A., Neumann N., (2003). *The thought-translation device (ttc): neurobehavioral mechanisms and clinical outcome.*, *IEEE Trans Neural Syst Rehabil Eng.*,11,120-3.
- Blanchard E.B, Epstein L.H., (2000). *Biofeedback, Manuale Introductivo*, Bollati Boringhieri Editore.
- Davila C.E., Srebro R., (2000). *Subspace averaging of steady-state visual evoked potentials*, *IEEE Transaction on Biomedical Engineering*, 47, 6, 720-728.
- Davila C.E., Srebro R., Ghaleb I.A., (1998). *Optimal detection of visual evoked potentials*, *IEEE Transaction on Biomedical Engineering*, 45, 6, 800-80.
- Chapin J.K, Moxon K.A., Markowitz R.S.,Nicoletis M.A.L., (1999). *Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex*, *Nature Neurosci.*, 2,664-670.
- Comon P, (1994). *Independent component analysis: a new concept?* , *Signal Proc.* 36, 287-314.
- Donchin E., Spencer K.M., Wijesinghe R., (2000). *The mental prothesis: assessing the speed of a P300-based brain-computer interface*, *IEEE Trans. Rehabil. Eng.*,8,174-179.
- Kennedy P.R., Bakay R.A., Moore M.M., Adams K., Goldwaithe J., (2000). *Direct control of a computer from the human central nervous system*, *IEEE Trans. Rehab. Eng.*, 8(2), 198-202.
- Kronegg J. (2003). [http://vision.unige.ch/MMI/bci/papers/Kronegg2003b\\_\\_BCI\\_research\\_worldwide.pdf](http://vision.unige.ch/MMI/bci/papers/Kronegg2003b__BCI_research_worldwide.pdf)

- Farwell L.A., Donchin E., (1988). *Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials*, *Electroenceph clin Neurophysiol*,70, 510-523.
- Ilmoniemi R.J., Williamson S.J. (1987), *Analysis of the magnetic alpha rhythm in signal space*, *Soc. Neurosci. Abstr.*, 13, 46-50.
- Ilmoniemi R.J., Williamson S.J., Hostetler W.E., (1987). *New method for the study of spontaneous brain activity*, *Biomagnetism*, Tokyo Denki University Press, 182-185.
- Ilmoniemi R.J., Uusitalo M.A.,(1997). *Signal-space projection method for separating meg or EEG into components*, *Med. Biol. Eng. Comput.*, 35, 135-140.
- Jung T.P., Makeig S., weserfield M.and Townsend J., Courchesne E., Sejnowski T.J., (2001). *Analysis and visulization of single-trial event-related potentials*, *Human Brain Mapping*, 14, 166-185.
- Makeig S., Jung T.P., Bell A.J., Ghahremani D., Sejnowski T.J. (1997). *Blind separation of auditory event-related brain responses into independent components*, *Proc. Natl. Acad. Sci. USA*, 94, 10979-10984.
- Mc Farland D.J., Miner L.A., Vaughan T.M., Wolpaw J.R., (2000). *Mu and beta rhythm topographies during motor imagery and actual movement*, *Brain Topography*,3,177-186.
- Morgan S.T., Hansen J.C., Hillyard S.A., (1996). *Selective attention to stimulus location modulates the steady-state visual evoked potential*, *Neurobiology*,93,4770-4774.
- Muller M.M., Hillyard S.A., (1997), *Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 hz range*, *Cognitive Brain Research*,6,249-261.
- Pfurtscheller G., (1989). *Functional topography during sensomotor activation studied with event-related desynchronization mapping*, *J. Clin. Neurophysiol*, 6,75-84.
- Pfurtscheller G.,Lopes Da Silva F.H. editors (1999). *Event-related desynchronization*, *Handbook of electroencephalography and clinical neurophysiology*, Amsterdam Elevier,6.
- Regan d., (1989). *Human Brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*, Elsevier Publisher, New York.
- Silberstein R.B., (1990). *Steady-state visually evoked potential topography associated with a visual vigilance task*, *Brain Topography*, 3,.337-347.
- Sutter E.E., (1992). *The brain response interface communication through visually-induced electrical brain response*, *J. Microcomp. Appl.*, 15,31-45.

- Ventura G., (2002). Basi Metodologiche per lo sviluppo di sistemi BCI: applicazioni on-line di segnali SSVEP [*Methodological frameworks to develop BCIs: real-time application for SSVEP signals*], Md Thesis, Politecnico degli Studi di Milano.
- Vapnik N.V, (1995). *The Nature of Statistical Learning Theory*. Springer.
- Vidal, J.J., (1973). *Direct brain-computer communication*, Ann. Rev. Biophys Bioeng, 2, 157-158.
- Vidal, J.J., (1977). *Real-time detection of brain events in EEG*, IEEE Proc., 65,633-664 [special issue on Biological Signal Processing and Analysis].
- Wolpaw J.R., Birbaumer N., Heetderks W.J, McFarland D.J., Peckham P.H., Schalk G., Donchin E., Quatrano L.A., Robinson J.C. and Vaughan T.M. (2000a). *Brain-Computer Interface Technology: A Review of the First International Meeting*, IEEE Trans. Rehabil. Eng.,8,161-163.
- Wolpaw J.R., McFarland D.J., Neat G.W., Forneris C.A., (1991). *An EEG-based brain-computer interface for cursos control*, Electroenceph. Clin. Neurophysiol., 78, 252-259.

# Helping Viewers Press the Right Buttons: Generating Intuitive Labels for Digital Terrestrial TV remote controls

Jane Lessiter\*, Jonathan Freeman, Robert Davis and Andrew Dumbreck\*

i2 media research limited  
Department of Psychology  
Goldsmiths College  
University of London

\*External R&D Contract Manager  
Technology, R&D  
Ofcom  
London

---

## ABSTRACT

This paper describes a four phase research study into the ease of use of DTT (digital terrestrial television) remote controls, focusing on button labelling. The aim was to identify intuitive remote control button labels, thereby enhancing the ease of use of DTT remote controls. In phase 1, a series of brainstorming (discussion) groups was run involving members of the British public to generate intuitive button labels for a variety of DTT functions. The second phase involved expert screening of these labels. The revised lists of labels for the functions tested were then compiled into a questionnaire and distributed to members of the general public in phase 3. Respondents were asked to rank order their favourite labels for each function. The final phase involved exploring the extent to which the subjective preferences derived from the results of phase 3 mapped onto viewers' behavioural responses. Key measures explored included the speed and accuracy of correct button label identification. Generally, the results revealed close correspondence between the subjective preference data and behavioural responses. Whilst these results suggest that subjective preferences for labels are reasonable indicators of behavioural proficiency, we recommend that behavioural measures also be collected where possible. The approach reported here with reference to remote control button labels can be applied to a wide range of interactive media products and services

---

Keywords: *Digital Terrestrial Television, button labeling, remote control.*

Received 24 October 2003, received in revised form 9 December 2003; accepted 20 December 2003.

## 1. Introduction

Evidence from previous research and anecdotal reports suggest that people are often confused by the button labels used to convey particular functionality on remote

---

\* Corresponding Author:  
Jane Lessiter,  
i2 media research limited,  
Department of Psychology,  
Goldsmiths College,  
University of London,  
New Cross,  
London SE14 6NW, UK  
E-mail: J.Freeman@gold.ac.uk/J.Lessiter@gold.ac.uk

controls for a range of interactive media products and services. The issue is particularly important with reference to digital TV, given the ubiquity of television and the benefits that easy to use digital TV can provide for large sections of the population, both in terms of increased choice of programming, and access to interactive entertainment and information services. Whilst digital TV penetration in the UK is high relative to other countries, for people who have not yet adopted digital TV to readily adopt it, it needs to be perceived as having benefits to the user and as being easy to use. DTV affords increased functionality over the offerings of conventional analogue TV. Potentially it could provide added benefits to users for whom TV is already a valued product in the home, with increased entertainment in the form of, for instance, more choice, more specialist offerings, and more time shifting ability to suit the individual's lifestyle. DTV can also provide a useful information resource, giving access to government and local service information via a familiar and trusted medium. However, the added functionality of enhanced interactive programme features can also increase the complexity of the system. In this context usability is of paramount importance, especially so for a product like television which has almost universal penetration and is generally perceived as being very easy to use. Earlier work funded by the UK's Independent Television Commission (Freeman & Lessiter, 2001) explored the perceived ease of use of dTV and interactive TV (iTV) relative to 17 other products pre-rated (n = 51) for ease of use. The results revealed that the ease of use of dTV and iTV were perceived to be on a par with personal computers – and fell among the bottom third *least* easy to use of the rated products. In contrast, analogue TV was rated amongst the top third most *easy* to use products.

The ease of use of digital TV, and perceptions relating to its ease of use, are of particular importance when it is considered that the sensory, cognitive and dextrous abilities of the general population - and their confidence with technology - vary substantially. Further, because of the typical uses of TV - entertainment, relaxation and information - and the context in which it is typically used (the home environment), people do not expect to expend effort in learning how to use it. If dTV can satisfy the expectations viewers have of television, based on their experiences with analogue TV, and it is perceived as being easy to use, viewers' positive experiences will generate positive word of mouth for digital TV generally and improve non-users' perceptions of digital TV. Improved ease of use will also enable viewers to fully exploit the

information and entertainment capabilities of digital and interactive TV, making its use a satisfying, enjoyable and fun experience.

In 2002, a large scale survey was conducted which aimed to identify what characterised people with the biggest ease of use problems (Freeman, Lessiter, Williams & Harrison, 2003). The survey, based on over 1300 people, asked respondents about their attitudes to technology, media consumption and dTV. Following factor and cluster analyses of the resultant attitudinal data, seven clusters of people were identified representing different types of media consumer. The clusters differed in their perceptions of the ease of use of technology generally and dTV in particular, the hurdles they faced in dTV adoption and their motivations to adopt digital TV. Whilst the segmentation was based purely on attitudinal measures, the clusters identified were strongly differentiated in terms of dTV ownership. Across the clusters, dTV penetration ranged from 5.9% ('Strong Resisters') to 65% ('Technology Enthusiasts'). The seven clusters were further categorised into three higher order groups differentiated primarily by their perceptions of the ease of use of digital TV. The three groups were labelled 'It's too slow and clunky for me', 'I can deal with it if there's content I want', and 'It's too complicated for me'.

A series of usability trials were then run (in collaboration with the Consumers' Association) followed by focus group discussions about participants' experiences with the trial dTV services (Freeman, Lessiter, Williams & Harrison, 2003). These trials were conducted to explore the relationship between the 'type' of user and the extent to which they were affected by dTV usability issues. Very generally, while 'It's too complicated for me' users were more likely to make internal attributions about the usability issues, 'It's too slow and clunky for me' users were more likely to blame the inadequacies of the system. However, many specific usability issues were identified as problematic, either for themselves and/or for others, by all users irrespective of grouping. Thus, priority areas on which to concentrate further research resource were identified. Digital TV remote controls, which have been referred to as a viewer's handshake with the product (Daly-Jones, personal communication), were identified as a key priority. For instance, some DTT users did not understand the button labels on many remote controls in that the labels were sometimes deemed unintuitive or even misleading as shown by the following quote:

*“Also the exit button, I mean usually when you’ve finished something and you want it to stay on something you’re used to pushing ‘enter’ or ‘save’ or whatever, to push ‘exit’ just seems very bizarre in my mind...”*

This, combined with industry requests to assist in their work to generate recommended labelling for a variety of key dTV functions, prompted our research into identifying what labels (which could be words or symbols) are most simply and easily understood by the majority of the public to convey different functions that digital TV offers (i.e., what labels are *intuitive*?). The use of intuitive labels on remote controls promises to reduce manufacturers’ and retailers’ product support costs. For example, telephone help-line use and product returns should fall if products are designed to be more intuitive and easier to use.

## **2. Method**

To meet our research goal of identifying intuitive remote control button labels a four phase research programme was conducted. This involved both qualitative and quantitative methods, and subjective and objective measures.

### **2.1 Phase 1: Brainstorming Sessions**

First, a series of five brainstorming sessions - each involving (five to six) members of the British public - was run to generate ideas for labels for nine different dTT functions (‘Subtitles’, ‘Languages’, ‘Text’, ‘Exit Text’, ‘TV guide’, ‘Return to Sound and Vision [from any interactive application]’, ‘Back to Previous Screen’, ‘System Set-up Menu’, and ‘Power On’).

Fourteen males and thirteen females (mean age = 51.7 years, SD = 23.4) participated and were recruited from (a) our database of ‘willing participants’, (b) students and staff from Goldsmiths College, and (c) several elderly social groups in the south-east London area. Seven participants reported having dTV (3 Freeview, 4 Sky Digital). The groups used people with a range of attitudes towards technology – from the very confident (‘It’s too slow and clunky for me’) to the technophobic (‘It’s too complicated for me’). There was a slight deliberate bias towards those with average (‘I can deal with it if there’s content I want’) and low confidence (‘It’s too complicated for me’). It was reasoned that people with greater confidence in using technology would be more likely to explore and eventually learn the button labels regardless of any

ambiguities. Group members were simply asked to think of any labels (symbols or words) relating to the given function description.

The number of suggestions for each of the dTV functions ranged from 6 ('Text') to 33 ('Return to Sound and Vision [from any interactive application]')

## **2.2 Phase 2: Expert Screening of Generated Labels**

In phase 2 an expert screening panel comprising six members of the UK Government's Digital Television Project Technology and Equipment Group – Converter subgroup (TEG-C) inspected the list and deleted unsuitable suggestions. Suggestions that were too long for a button, ambiguous (or bizarre!), and that might be confused with labels for other functions were removed. The panel proposed minor modifications to some of the retained labels – such as presenting word labels in lower case, removing hyphens between words, and converting symbols to international standard versions. They also requested a few additions to the list to be tested in phase 3.

## **2.3 Phase 3: Questionnaire Survey – Ranking Button Labels**

Phase 3 involved a questionnaire survey where the screened list of labels (from Phase 2) were presented for each function and respondents were required to rank order their favourites (the number of choices ranged from 4 [for 'Text'] to 12 [for 'Return to Sound and Vision']). The 'Subtitles-Languages' and 'Power On' functions were removed from this survey (low priority), but the 'TV Guide' function was divided into two functions – 'Now and Next' and 'Electronic Programme Guide' (EPG). The same list of labels that had been generated and screened for 'TV Guide' were presented for each. There were eight functions in total that were addressed in the questionnaire – 'Subtitles', 'Text', 'Exit Text', 'TV Guide' (i.e., Electronic Programme Guide), 'Now and Next', 'Return to Sound and Vision', 'Back to Previous Screen', and 'System Menu'. Two versions of the questionnaire were constructed – the labels were presented in a different order in each to reduce the possibility of the presentation order of labels affecting respondents' ranking decisions.

Five hundred and fifty questionnaires were distributed (a) across England, Wales and Scotland via post to people who had previously taken part in our other research (but not Phase 1 of this research project) and who were willing to take part in further

research, and (b) to miscellaneous people around Goldsmiths College. As an incentive, participants were offered entry into a prize draw to win £100. Of the 550 distributed questionnaires, 237 were returned and were suitable for data analysis (43% response rate). Responses were received from people residing in 23 of the 39 counties in England. In addition, four people from Scotland and one from Wales responded. The age of the sample ranged between 17 and 92 years (mean age = 59 years, SD = 19). Forty-four percent of the sample were male. Thirty-six percent of the sample reported having dTV at home, a figure slightly lower than that for the UK population as a whole.

Each nominated label was given a score corresponding to its ranking (e.g., the label ranked first favourite was given a score of 1). Unranked labels were treated as missing data. As not all respondents gave all five ranks (or for the 'Text' function, 4) subjective preference was calculated by a simple tally of respondents' first choices for each label. Thus, only the results for first favourite were analysed.

Overall the results suggested that text labels were preferred to symbolic labels. For no functions did a symbol receive the highest number of first favourite nominations. Clearly it is an empirical question as to whether this subjective preference translates into behaviour. Interestingly, using a small sample of visually impaired users, the Royal National Institute for the Blind (Gill & Perera, 2003) found that visually impaired people responded significantly faster to symbolic rather than text labels.

For three functions ('Text', 'Back to Previous Screen' and 'Subtitles') there were clear 'winners' where the label directly corresponded to its function (labels 'text', 'back' and 'subtitles'). Each of these labels received over 50% of first favourite nominations and indicated that people tend to prefer labels that do exactly what they say, rather than abbreviations. It may simply be a matter of finding the shortest best descriptor word for that function.

For three of the remaining functions, the differences between first and second choices were marginal with around a 5% difference in frequency of first favourite nominations. For the 'Return to Sound and Vision' function, the label 'return' was closely followed by 'TV' (30.1% vs. 26%), for 'TV Guide-Electronic Programme Guide' function the label 'guide' preceded 'progs' (35.4% vs. 29.2%), and for 'System Setup

Menu' function the label 'setup' marginally received more nominations than 'settings' (29.9% vs. 27.6%). For these functions, which may be less well conceptualised by users, it may be more difficult to identify unique descriptor terms.

## **2.4 Phase 4: Experimental Evaluation of Button Labels**

### *Aims*

The aim of the final phase was to explore the extent to which the subjective preferences revealed in phase 3 translated into behavioural performance – i.e., are the most favoured labels more quickly and accurately identified than those less frequently nominated? It aimed to explore intuitiveness (on first exposure), memorability (learnability) and common mistakes. In addition, this phase measures the degree of consistency between the results found in phase 3 and the behavioural paradigm being used here. Careful design of this phase was crucial, and in order to reliably achieve its aims with a finite resource it was necessary to minimise the number of functions and corresponding labels to be tested.

### *Selection of Functions and Labels to be Tested*

Three functions were selected for testing in this phase – 'Subtitles', 'Return to Sound and Vision' and 'Back to Previous Screen'. These functions were chosen for various reasons. 'Subtitles' was selected because a dedicated button for this function is to be recommended by the TEG-C group. While the label in full can fit on some buttons, an alternative shorter label might be preferred by some manufacturers, and further, the range of label possibilities for this function also included a symbol. 'Back to Previous Screen' is a function for which there has been much debate as to the most suitable label. Finally, the 'Return to Sound and Vision' function was selected because there has been discussion in the industry regarding the introduction of this new 'comfort' button, and its inclusion in TEG-C's baseline receiver specification for Digital Terrestrial Television converter boxes has been discussed. As a new and potentially frustration-reducing function it seemed important to test labels for this unfamiliar function to explore whether participants would understand the label's meaning.

Three labels for each of the three selected functions were chosen based on the results from phase 3. Labels were tested that were both close contenders for favourite (i.e., where there were two very 'popular' labels), and also subjectively disparately favoured labels (i.e., the most and least frequently ranked favourite labels). For example, for the 'Subtitles' function, the clear winner – 'subtitles' (ranked favourite by

61% of the sample) - was compared with 'subT' (10%) and the internationally recognised 'subtitles' symbol (9%). For the 'Return to Sound and Vision' function, the labels 'return' (31%), 'TV' (26%) and 'out' (0.5%) were tested. And for the 'Back to Previous Screen' function, the labels 'back' (52%), 'back up' (2%) and 'cancel' (1%) were compared.

### *Design*

A 3 by 3 mixed measures design was used with one within group factor (function) and one between group factor (label). Thus, all participants were exposed to each of the three functions, but each participant was exposed to just one of the three labels for each function. This reduced the design complexity associated with employing a fully repeated measures design (in which button location would need to be an additional variable). Participants were exposed to each label a total of six times (trials) enabling learnability to be explored. Within each trial participants were given five attempts to identify the correct button (enabling common mistakes to be identified). Thus there were a total of 18 trials for all functions (i.e., 3 functions x 6 trials) and 30 attempts over all trials for each function (i.e., 6 trials x 5 attempts). The presentation of the functions was fixed in the following order: 'Subtitles', 'Return to Sound and Vision' and 'Back to Previous Screen'. Fixed order was preferential to counterbalancing as the analyses aimed to explore labels within functions rather than cross-function comparisons. Further, fixed order kept fatigue and practice effects constant on performance across function.

The independent variable was button label (*A, B, C*), so each participant within a sample of 27 people would receive a different combination of the three labels (3 x 3 x 3). The dependent variables were accuracy (correct/incorrect) and reaction time (RT: in milliseconds). The RT data were cumulative – i.e., the RT to correct response (across attempts) was logged. If there was no correct response for a trial (across the five attempts), no RT was registered for that trial. Every button press (coded for button function) was recorded in the experimental trials so that common errors could be explored.

### *Virtual Remote Control and Procedure*

A program was written to present the experiment (using Borland C++ Builder for Windows, Version 5.0). The graphic of the virtual remote control (see Figure 1) was constructed based on TEG-C recommendations for the required buttons on a DTT

remote control. The size of the buttons and the inter-button spaces conformed as closely as possible to guidelines produced by the Royal National Institute for the Blind (2003). Each function for which labels were evaluated in this Phase was placed on the same button throughout the experiment. In effect, only the labels of the three experimental buttons were varied within the experiment. Aside from a brief welcome and introductory note, the entire task was presented on an Iiyama INT3819Ts 15" LCD touchscreen.

Participants began with a series of 10 (unrecorded) practice trials which were unrelated in content to the experimental trials. The practice block aimed to (a) build participants' confidence with using a touchscreen in general, given that some would have never used one before, (b) familiarise them with the button size and halo catchment area for accurate button pressing, and (c) give them practice at positioning their finger over a red square presented at the bottom of the touch screen to initiate the onset of the experimental stimulus (i.e., the remote control). This was employed in an attempt to keep respondents' starting position for each trial as constant as possible so that measured reaction times could be meaningfully interpreted. In each of the practice trials the task was simply to press one of four buttons labelled A, B, C or D. Each time a button was pressed a tone was heard. As with the experimental trial, if the button press was correct, RT was recorded, the stimulus disappeared and the participant was moved on automatically to the next trial. If the button press was incorrect the stimulus remained on the screen until the correct button was chosen (up to five attempts).

Having completed the practice trial (which could be repeated if requested), participants were moved on to a familiarity task. The remote control replete with the specific configuration of button labels unique to that individual was presented on-screen for one minute. This aimed to reduce potentially large differences in exposure time to the remote control given likely individual difference in RTs on the experimental trials. For example, participant A might have taken longer to find the correct button on Trial 1 than Participant B and might have had the opportunity to more fully explore the remote control in this time. This could have potentially affected their RT for identifying the correct button for the next function. Before the remote control presentation, participants were simply instructed to familiarise themselves with the graphic rather than memorise it.

Participants were next moved on to the experimental trials. Each trial began with a function description. The function descriptions were written carefully to avoid using the critical labels in the descriptive text. The exception to this was 'Subtitles'. To describe this function without using the word 'subtitles' might have potentially confused participants – most people are now familiar with this function. Once they had read the instructions, participants pressed the red square to initiate the onset of presentation of the virtual remote control. On its presentation they were required to press the button they believed would activate the function described in the instruction they had just read. There were five attempts per trial and 18 trials in total. At the end of the experimental session all participants were thanked for their participation and fully debriefed. On average the full task took between 20 and 30 minutes to complete.

### *Participants*

There were 27 participants aged 44 years or under. Fourteen were male and 13 were female with a mean age of 31.3 years (SD = 7.6, range = 18 – 44 years).

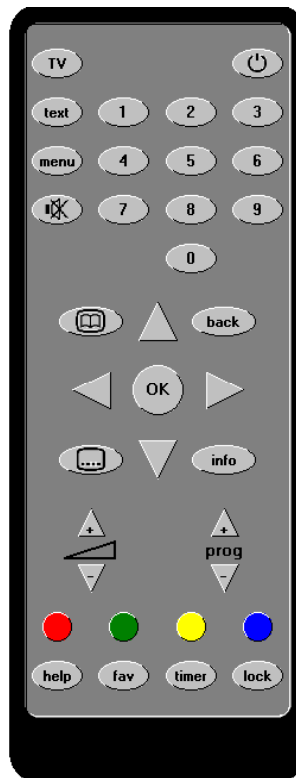
Another 27 participants, aged 45-84 years, have subsequently been run and data from these tests are currently being analysed. Data for this older sample will be available in a document prepared for the ITC to be released shortly, and a comparison of old and young responses will be published elsewhere.

### *Results*

The resultant data were explored in a number of ways: (1) the number of people to correctly identify the label, (2) the number of attempts to make the correct response (or time out) on each trial, (3) types of errors, and (4) average time to make a correct response for each label (a) on each trial and (b) over all trials.

With 27 different combinations of the three labels for each function, nine people were exposed to the same label for any particular function.

Results for each dependent variable are reported for each function below. Most of the data is descriptive but analyses of variance are reported for the RT data over all trials.



**Figure 1:** Graphic of the Virtual Remote Control

*Function 1: Subtitles*

*Number of people to correctly identify the label*

For the 'Subtitles' function, almost everyone identified the correct label on each of the six trials (at some point across the five attempts). For both the 'subtitles' and 'subT' labels, all nine people correctly identified their respective buttons on each of the six trials. Only the subtitles symbol fared marginally less well – on the first trial, one person was unable to locate this button, but for every other trial, all participants were successful.

*Number of attempts either to correct or time out*

So how many attempts were needed to find the correct button? Across all trials, participants, on average, tended to find the correct button with one attempt – thus, all labels were relatively easy to identify quickly. However, there was some negligible variation between the labels. Participants needed fewer attempts to find the correct button on the first trial for the label 'subT' (1.22 attempts) than they did for 'subtitles' (1.33) or the symbol, which on average took nearly two attempts to identify (1.78). This pattern did not remain consistent across the six trials; by the second trial, the 'subtitles' label took just one attempt while for the other two labels it took slightly more

(1.11 for 'subT' and 1.22 for the symbol). On trial 3, 'subT' took the lead once more, but by trials 5 and 6 all three labels took just one attempt to find. On average over all trials, 'subtitles' and 'subT' needed the fewest number of attempts to correct – 1.07 each – while the symbol needed 1.2 attempts.

In terms of the absolute number of attempts required to first time correct across trials (out of a potential maximum of 30 attempts), most people (8 of 9) found the 'subT' label with one attempt, the other participant found it within three attempts. Perhaps surprisingly, just six of nine people found the label 'subtitles' on the first attempt; three people found it in two attempts. For the symbol, five people found it on the first attempt, three people on the second attempt, and one person took eight attempts (i.e., third attempt, second trial).

#### *Types of errors*

So what buttons were being mistaken for the 'Subtitles' function button? Six of the eight people that failed to correctly identify the 'Subtitles' button on the first attempt, chose 'Text' instead. Presumably these participants are basing their decision on the analogue TV model where subtitles are accessed first by entering text. Indeed one person used up four attempts by pressing the sequence, 'text'-'8'-'8'-'8'. Two people made the error of pressing the 'TV Guide – Electronic Programme Guide' function button (which took the form of an open book symbol on the remote control). Perhaps the association of 'book' with 'words'/printed text misled these participants to press this button.

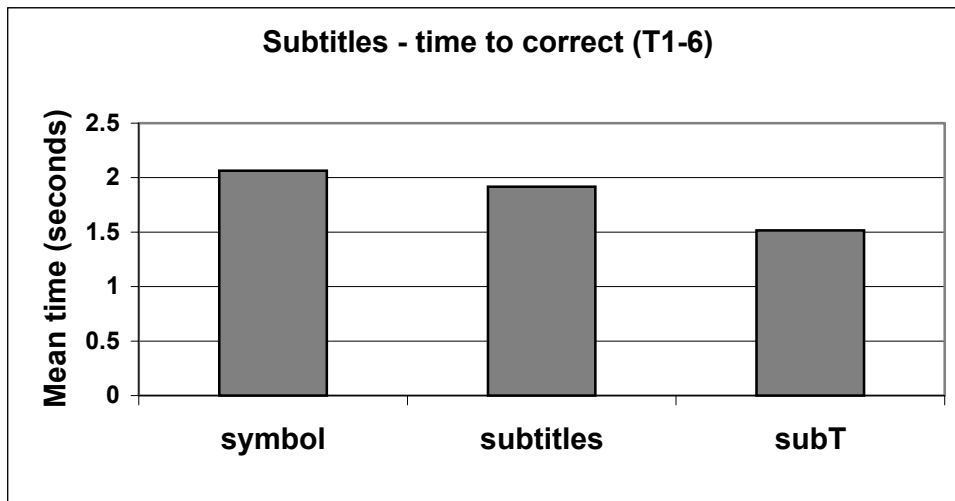
#### *Average time to make a correct response for each label on each trial*

In terms of the average time taken to identify the correct label (across attempts) within each trial, 'subT' was most quickly identified on the first trial. It took about half as long to find this label than for the other two labels, which both took around six seconds to correctly locate. By the second trial, the differences between the times for each label were getting smaller with the symbol taking the longest to correctly identify. By trial 3, the labels 'subtitles' and 'subT' were virtually equivalent in time, with the symbol still lagging behind. Although this pattern changed somewhat on trial 4 ('subT' took the longest to identify), the pattern was consistent with earlier trials for trials 5 and 6. 'subT' was the only label for this function that was correctly identified in under a second by the final trial.

*Average time to make a correct response for each label over all trials*

This pattern of results was maintained when the data were explored in terms of the average time to identify the correct label across all trials. Participants were able to learn each label with time (each trial), however, 'subT' was relatively easier to learn on average, than either 'subtitles' or the symbol. The label 'subT' took an average 1.5 seconds to find over all trials (n = 52) compared with 'subtitles' (1.9 seconds, n = 52) and the symbol (2.1 seconds; n = 51). A one-way ANOVA revealed no statistically significant differences between the labels in these mean times ( $F_{(2,154)} = 1.88$ ; ns) (see Figure 2). This may be due to low power because of the small sample size available here. Throughout the results section of this paper, this is a consideration of which the reader should be aware.

(The results presented in Figure 2 exclude 6 outliers i.e.,  $> 2.5 \times S.D. + \text{mean}$ , who responded in excess of 10 seconds. Before deletion of these cases, the mean times to correct across trials were 2.7 for the subtitles symbol (n = 53), 2.5 for 'subtitles' (n = 54) and 1.9 for 'subT' (n = 54).)



**Figure 2:** Mean time to identify the correct Subtitles label over all trials

*Function 2: Return to Sound and Vision (from any Interactive Application)*

*Number of people to correctly identify the label*

Although all participants correctly identified their respective labels by trials 5 and 6, there was some variation in the earlier trials. As predicted on the basis of the results

from the subjective preferences in Phase 3, 'out' fared least well – seven people (of nine) made a correct button press on the first trial, rising to eight by the second trial, but dropping again to seven people by the third trial. Nevertheless, by the fourth trial eight people were successful again and there was 100% success for the fifth and sixth trials. 'TV' and 'return' also showed similar variation, indicating some difficulty in remembering the correct label. 'TV' started with just one person unable to identify the correct button, but by trial 2, all participants made the correct choice. This dropped to eight people again for the third trial. By the fourth trial and for the remaining two trials, everyone correctly identified 'TV'. Finally, 'return' started well – for the first two trials, all participants found the correct button. But on trial 3, one person failed to correctly identify 'return'. Nevertheless, on the last three trials, 100% of participants found the correct button.

*Number of attempts either to correct or time out*

Overall people tended to have more difficulty with this function than with 'Subtitles'. On the first trial all participants, irrespective of label, took an average of around 2 attempts to find this button or time out. As anticipated, the label 'out' on average took more attempts (2.56), than 'TV' (2) or 'return' (1.78). Across the trials, 'out' also showed considerable variation and always took an average of around two attempts to find with each trial (except for trial 6 when it took on average 1.22 attempts). This indicates that 'out' was less memorable and more difficult to learn. Interestingly, 'return' also showed some variation – while it outperformed the other labels on the first trial and although the average number of attempts generally declined over trials, it increased on trials 3 and 6. It appears that 'TV' showed the most promising pattern of results; by trial 5 all participants were getting it right first time.

In terms of how many absolute attempts it took to get it right across trials, five (out of nine) people made the correct choice on the first attempt for both 'TV' and 'return', while only three people correctly identified 'out' on the first try. For 'TV' one person needed two attempts, two people needed three attempts, and one person needed six attempts to find this button. For 'return', three people found it in two tries, and one in six tries. And for 'out', two people needed two attempts, two people needed three attempts, one person needed six attempts, and another person needed 25 attempts to find the right button (i.e., they did not correctly identify the button until the last attempt of the fifth trial).

### *Types of errors*

So what common mistakes were made? Six out of the fourteen people who failed to locate the correct button on the first attempt made the mistake of pressing the button corresponding to the 'Back to Previous Screen' function. Two people first chose the 'Text' function, two selected the 'TV Guide – Electronic Programme Guide' function button (an open book symbol), and one person incorrectly selected each of the following: left pointing navigation button, 'menu', 'info' and the 'Subtitles' function.

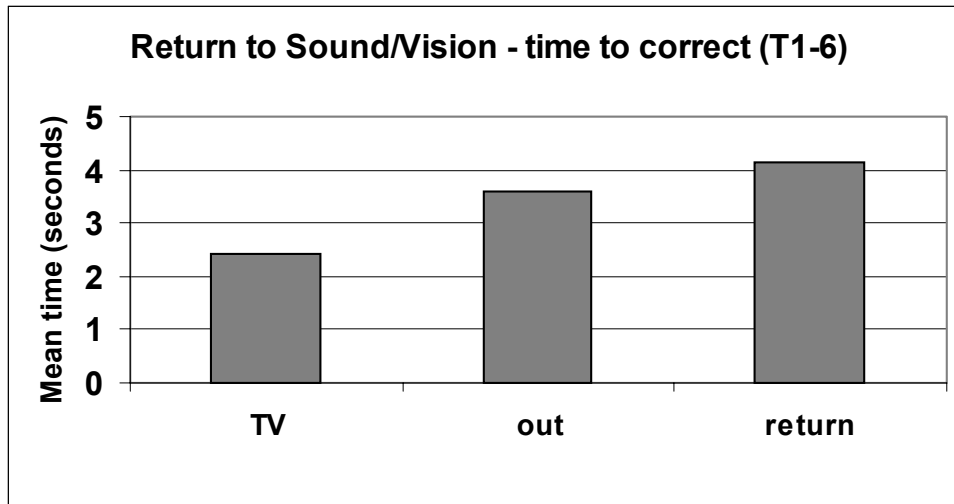
### *Average time to make a correct response for each label on each trial*

In terms of the average time taken to identify the correct label (across attempts) within each trial (learnability), on trial 1 'return' was the most quickly identified label, taking nearly five seconds less time to locate than for either 'out' or 'TV'. By trial 2, 'return' in fact took slightly longer than previously (8.2 seconds) while 'TV' and 'out' were much more speedily identified in around two seconds each. This was largely due to an outlier for the 'return' label – a participant who was unable to locate 'return' on the first trial, but who took 28.5 seconds on the second trial – about 13 seconds longer than the next slowest participant. When this outlier was removed, the mean time for 'return' on the second trial was 5.67 seconds. The pattern shifted over the trials. By trial 6, 'TV' was the most quickly identified label, taking about 1 second to find.

### *Average time to make a correct response for each label over all trials*

In terms of general learnability (time to correct across trials), 'TV' emerged as the winner. 'TV' took an average of 2.4 seconds to locate ( $n = 50$ ), compared with 3.6 for 'out' ( $n = 47$ ) and 4.2 for 'return' ( $n = 51$ ) (see Figure 3). One-way ANOVA suggested that these differences approached significance ( $F_{(2,147)} = 2.7$ ;  $p = 0.07$ ). Specifically, post-hoc tests (Games Howell correction – equal variances not assumed) revealed that the difference between 'TV' and 'return' approached significance ( $p = 0.052$ ).

(The results presented in Figure 3 exclude four outliers whose scores fell above 2.5 SDs from the mean – i.e. greater than 19.6 seconds. Before deletion of these cases, the mean times to correct (across trials) were 3.6 seconds for 'TV' ( $n = 52$ ; corrected = 2.4), 4.1 for 'out' ( $n = 48$ ; corrected = 3.6) and 4.6 for 'return' ( $n = 52$ ; corrected = 4.2).)



**Figure 3:** Mean time to identify the correct Return to Sound and Vision label over all trials (learnability).

### *Function 3: Back to Previous Screen*

#### *Number of people to correctly identify the label*

The results were most revealing for the 'Back to Previous Screen' function. The label 'back' was correctly identified by all nine participants on every trial. In comparison, 'back up' needed two trials before everyone made the correct choice (and continued to do so for the remaining trials). More people had problems with the label 'cancel'. On the first trial, just four people identified this button, rising to six people for the next two trials, then seven people for trials 4 and 5. It was only on the sixth trial that all nine people correctly located this button.

#### *Number of attempts either to correct or time out*

Of all the functions the biggest variation between labels in the number of attempts to correct was found for 'Back to Previous Screen'. The least popular label subjectively, 'cancel', on average took participants many more attempts to find than either of the other two labels. On the first trial, 'cancel' took an average of four attempts to correctly locate and press. Furthermore, although there was a steady decrease in the number of attempts with each trial for 'cancel' (indicating learnability), even by trial 6 the average number of attempts was around two (1.78). The label 'back up' required an average of around three attempts to find on the first trial (2.78). Again, this label showed a steady decrease in the number of attempts needed with each trial. But in contrast with 'cancel', for 'back up' all participants identified this label correctly on the first attempt by trial 4. Clearly, this label was easy to learn. However, the label 'back' was the clear winner here. It consistently needed around just one attempt to correctly identify and press from the first trial right through to the (final) sixth trial. Over all trials,

on average 'back' needed 1.09 attempts, compared with 1.46 attempts for 'back up' and 2.56 attempts for 'cancel'.

In terms of the number of attempts to first time correct, nobody was able to find the 'cancel' button on the first attempt. Three people were able to find it in two attempts, and one on the fifth attempt. Five people took seven attempts or more to find the label 'cancel'. In fact, one person found the correct button only on the last attempt of the sixth trial (i.e., 30 attempts to correct). For 'back-up', there was also some difficulty. While four people managed to correctly locate this button on the first attempt, one person did it in three attempts, two people in four attempts, one in nine attempts, and another in 13 attempts. The label 'back' was most easy to identify in the fewest number of attempts. Seven out of the nine participants found 'back' on the first attempt. One person needed two attempts, and another needed three attempts.

#### *Types of errors*

In terms of common mistakes, 10 of the 16 people who failed to choose correctly on the first attempt instead chose the left pointing navigation button. Two people pressed the 'TV Guide – Electronic Programme Guide' function button (an open book symbol) and one person each mis-selected one of the following: right pointing navigation button, 'Text', the 'Return to Sound and Vision' function button, and the downward pointing navigation button. That the open book symbol ('TV Guide – Electronic Programme Guide' function) was mistakenly pressed for this as well as the other two functions suggests that the meaning of this symbol is not entirely clear to participants. However, their readiness to try this button might indicate that it is perceived as a 'friendly', 'no harm done' function. It is interesting that the navigation buttons were a common error. Clearly, they were an intuitive option for many participants, and actually require no label whatsoever.

#### *Average time to make a correct response for each label on each trial*

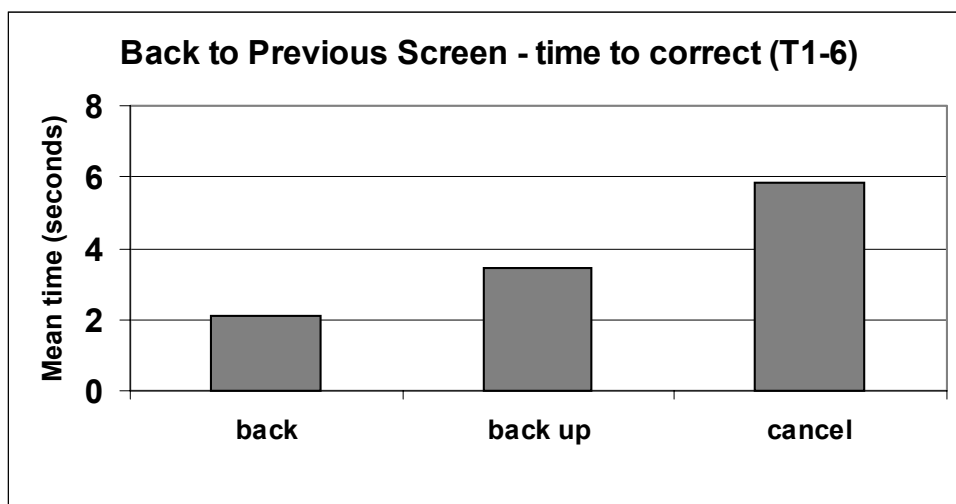
First, in terms of the average time taken to identify the correct label (across attempts) within each trial, the label 'back' was the most quickly located label on the first trial, taking about 5 seconds to find – three times as quick than for 'back up' (15.2 seconds) and almost five times faster than for 'cancel' (23.8 seconds). This pattern was similar for the next two trials, but with the differences in times between the labels getting smaller. By the fourth trial, 'back up' was the most quickly identified label, and this continued for the remaining trials. The differences between 'back' and 'back up' were

negligible, however. 'Cancel' was also more speedily identified with each trial. On trial 6, the mean time for 'cancel' was much larger than for the three previous trials for this label (19.3 seconds). This, however, was due to an outlier who took 135 seconds to find this label on the final trial. When this outlier was removed, the mean for trial 6 for 'cancel' dropped to 4.76 seconds.

*Average time to make a correct response for each label over all trials*

All labels were faster to identify over progressive trials, indicating learnability. On average over all trials 'back' took 2.1 seconds (n = 54). This was faster than for 'back up' (3.4 seconds; n = 50) and almost three times faster than for 'cancel' (5.9 seconds; n = 36) (see Figure 4). A one-way ANOVA revealed that there was a significant difference in time to correct (across all trials) between the three labels ( $F_{(2,139)} = 3.46$ ;  $p < 0.05$ ). Post-hoc tests (Games-Howell correction – equal variances not assumed) revealed that the difference between 'back' and 'cancel' approached significance ( $p = 0.053$ ) but there were no other significant differences.

(The results presented in Figure 4 exclude four outliers (time scores in excess of 2.5 S.D.s away from the mean i.e., in excess of 40.7 seconds). Prior to deletion of these outliers, the values for time to correct (across all trials) were 4.2 seconds for 'back up' (n = 51) and 11.3 seconds for 'cancel' (n = 39). There were no outliers for the 'back' function.).



**Figure 4:** Mean time (in seconds) to identify the correct over all trials (learnability)

### **3. Discussion and Conclusion**

Generally there was good correspondence between the subjective preference data and behavioural performance data, although the relative differences between the labels across the subjective and behavioural results were not always of equivalent magnitude. These results are reassuring and demonstrate that subjective preferences can be a good indicator of behavioural performance. Further, subjective preference measures are quicker, cheaper and easier to collect than are behavioural measures. However, the behavioural data are important to consider. For example, how many times will a viewer be willing to press the wrong button and get an unexpected system response before giving up on trying to use a service, and telling all his/her friends that it's impossible to use?

In terms of the behavioural data, the results demonstrate that for the 'Subtitles' function, whilst the label 'subtitles' was the (subjective) favourite label 'subT' outperformed 'subtitles' behaviourally (although statistically there was no significant difference between the times to locate these labels). This finding might relate to the label size – perhaps it was more difficult for subjects to read this text when it almost fills the button. Further, the 'subT' label might be more eye-catching because of the unusual capital letter at the end. It is also possible that younger people are more familiar and confident with using ad-hoc abbreviations from mobile phone text messaging, the use of which is particularly popular among younger age groups. It will be interesting to compare this result with data from the older age sample. The internationally recognised subtitles symbol was responded to least well, as predicted on the basis of the results from phase 3. It was interesting to note, however, that a common error in locating the 'Subtitles' function was access via the analogue TV route – in the UK, analogue TV viewers access subtitles by pressing text, and then 888. This suggests that in addition to having a dedicated subtitles button, designers should allow users to access subtitles via an option in the 'text' route - as this is what they are used to. Clearly, as a general rule, maximising the consistency between operating a novel technology (digital TV) and what people are used to (with analogue TV) is good practice, where possible, and will result in digital TV being easier to use for everyone.

For the 'Return to Sound and Vision' function, phase 3 showed that 'return' and 'TV' were close contenders for favourite, while 'out' was the least popular. The behavioural results for these labels demonstrated that 'out' caused participants the most difficulty.

Compared to the other labels, not only did fewer people actually identify 'out' on each trial, but they also needed more attempts to locate it correctly. The behavioural results were particularly revealing in identifying the relative usefulness between 'return' and 'TV'. Over the trials there was just one occasion where 'return' was not identified (at any point over the 5 attempts per trial) compared with two occasions for 'TV'. Further, 'return' needed marginally fewer attempts on average to find on each trial. However, on average (across all trials) participants responded to 'TV' more quickly than they did to 'return' and showed good learning for 'TV' over the trials.

While 'return' outperformed 'TV' in some analyses, during debriefing a few participants commented on the slightly ambiguous meaning of the term 'return'. Further, it was observed that the correct button for the functions 'Back to Previous Screen' and 'Return to Sound and Vision' were often confused by participants (e.g., pressing 'back' instead of 'return'), possibly because of the repetitive nature of the task. This confusion was exacerbated when the two terms for the different functions were more similar semantically and was reduced when the label 'TV' was used. It appears that, all results considered, 'TV' was the winning label.

Finally, for 'Back to Previous Screen', the behavioural results lent support to the subjective data. The label 'cancel' consistently performed least well of the three labels, as predicted on the basis of the results from phase 3. While 'back up' performed slightly less well to start, once learned, it was easily remembered. The label 'back' consistently performed well, as predicted on the basis of the results from phase 3. Indeed, the time to correct results showed that 'back' was responded to significantly more quickly than was 'cancel'.

A few observations are worthy of note in interpreting the results from this work. First, participants were often observed to be circling their finger around the centre of the remote control where the 'OK' and navigation buttons were located. The most interesting shapes and button spacings were located in this region suggesting that button location is an area worthy of future research.

Second, the observation that the 'Back to Previous Screen' and 'Return to Sound and Vision' functions were sometimes confused with each other, highlights the importance of context in terms of the labels given to the other buttons on a remote control. That is,

where a function may be ambiguous, function meaning is derived by comparison with possible alternatives.

Further, while it was aimed to present the behavioural study in an ecologically valid way, in everyday situations there are usually multiple means of accessing the same function - a reality that was not accounted for in this study. For instance, in many circumstances, a series of 'back' presses should eventually remove the user from an interactive application (as well as in one button press using the 'Return to Sound and Vision' function). In addition, in everyday remote control use, buttons are usually identified based on the users' understanding of what function they would like to activate, either directly when faced with the remote control, or indirectly when they are faced with options within menus and submenus. This is in contrast to the behavioural paradigm employed here where a description of a function was provided for them.

Thus, context (the influence of surrounding buttons and what they have been labelled) is clearly an important influence on people's decision making in this task. While this was kept standard (other than for the critical buttons which varied from person to person), and conformed to TEG-C recommendations, the extent to which the non-target buttons and labels used here affected performance on the critical labels is unclear.

The sequential methodology employed here, the design of the behavioural paradigm and the time scale available for this project made it necessary to reduce the number of functions and labels being tested at each phase. However, for a complete understanding of the intuitiveness and learnability of labels (for both existing and novel functions) on a given remote control, it would be necessary to explore all suggested function labels and possible label configurations.

The study used both qualitative and quantitative research methodologies in developing and answering the research question, enabling exploratory inquiry to develop into structured methods. Both users and 'experts' (from industry) participated in the study enabling usability (subjective opinion and objective behaviour) and pragmatic concerns of manufacturers to be addressed. The correspondence between phase 3 (subjective preference) and phase 4 (behavioural performance) was explored indirectly by comparing the patterns of results using both methods i.e., using

comparative estimations of 'best' through 'worst' label. However, there was no concrete benchmark by which to judge 'good' compared with 'bad' performance. Further, the labels were compared within-function – some functions are generally easier to understand and use than others. The results suggest that in general, subjective opinion may be a good, albeit rough, indicator of behaviour, though the differences between performance on the labels were of a lesser magnitude than the differences in subjective opinion. There was no evidence of very poor correspondence between phases 3 and 4 although it is possible that there might have been had all functions and labels been tested. Examples of poor subjective-behavioural correspondence would be of interest to designers. Subjective preference is an important criterion in usability research – it can be important to give people what they think they want as users' overt evaluations can sometimes override the performance benefits. *Perceptions* of ease of use are critical. However, they are not a direct substitute for behavioural measures, particularly more ecologically valid paradigms such as that employed here. Thus, whilst this study has shown that subjective preferences can be good indicators of behavioural performance, for new products and services in development, we recommend using a combination of both subjective and objective measures.

A final key point is when evaluations of the type reported here should be conducted. In terms of high support costs, high return rates, and bad word of mouth for brands (and product categories) it is potentially very expensive to design products without sufficient regard to their ease of use for their target users. Measuring target users' subjective and behavioural responses to software based mock ups of new products and services in development (using ecologically valid paradigms) is most definitely a worthwhile investment in the future success of these products.

#### 4. References

- Freeman, J., & Lessiter, J. (December 2001). Easy TV 2001 Report: ITC-UsE Ease of Use and Knowledge of Digital and Interactive Television – Results. [online] <http://www.itc.org.uk>
- Freeman, J., Lessiter, J., Williams, A., & Harrison, D. (January 2003). Easy TV 2002 Report. [online] <http://www.itc.org.uk>

Gill, J., & Perera, S. (2003). Accessible Universal Design of Interactive Digital Television. [online] <http://www.tiresias.org/reports/brighton.htm>

Royal National Institute for the Blind (2003). Guidelines for the Design of Accessible Information and Communication Technology Systems: Technology – Keypads. [online] [www.tiresias.org/guidelines/keys.htm](http://www.tiresias.org/guidelines/keys.htm).

# Software Design and Evaluation by Ergonomics Knowledge and Intelligent Design System (EKIDES)

Dr.-Ing. Iwona Jastrzebska-Fraczek\* and  
Prof. Dr. rer. nat. Heiner Bubb

Institute of Ergonomics, Technische Universität München,  
München, Germany

---

## ABSTRACT

The database system **EKIDES** (**E**rgonomics **K**nowledge and **I**ntelligent **D**esign **S**ystem) assists designers of technical systems, equipment, products and workplaces to meet ergonomic requirements for all system components and their interactions during the planning, development and subsequent design and blueprint processes. EKIDES is a design tool in form of an electronic reference system and an evaluation tool for work places and products.

The ergonomic tests can be carried out by using the *Basic, Consulting or Product Modules of EKIDES*. Furthermore, if the required measurement equipment is not available, or if a qualitative task or product analysis is sufficient, the module *Checklist* may be used.

---

Keywords: *design tools, electronic reference system, database system.*

Received 16 October 2003; received in revised form 12 December 2003; accepted 20 December 2003.

## 1. Introduction

The database system **EKIDES** (**E**rgonomics **K**nowledge and **I**ntelligent **D**esign **S**ystem) is a design tool in form of an electronic reference system and an evaluation tool for work places and products. EKIDES (formerly EDS - Schmidtke, H., and Jastrzebska-Fraczek, I (2000)) is composed of several modules (see figure 1). "Generic Design Module" can be used for non-specific analysis that utilizes generic ergonomics design considerations. "Application on Production Systems" can be used for specific applications.

---

\* Corresponding Author:  
Iwona Jastrzebska-Fraczek  
Institute of Ergonomics,  
Technische Universität München,  
Boltzmannstr 15. 85747 Garching  
Germany.  
Email: fraczek@ife.mw.tum.de

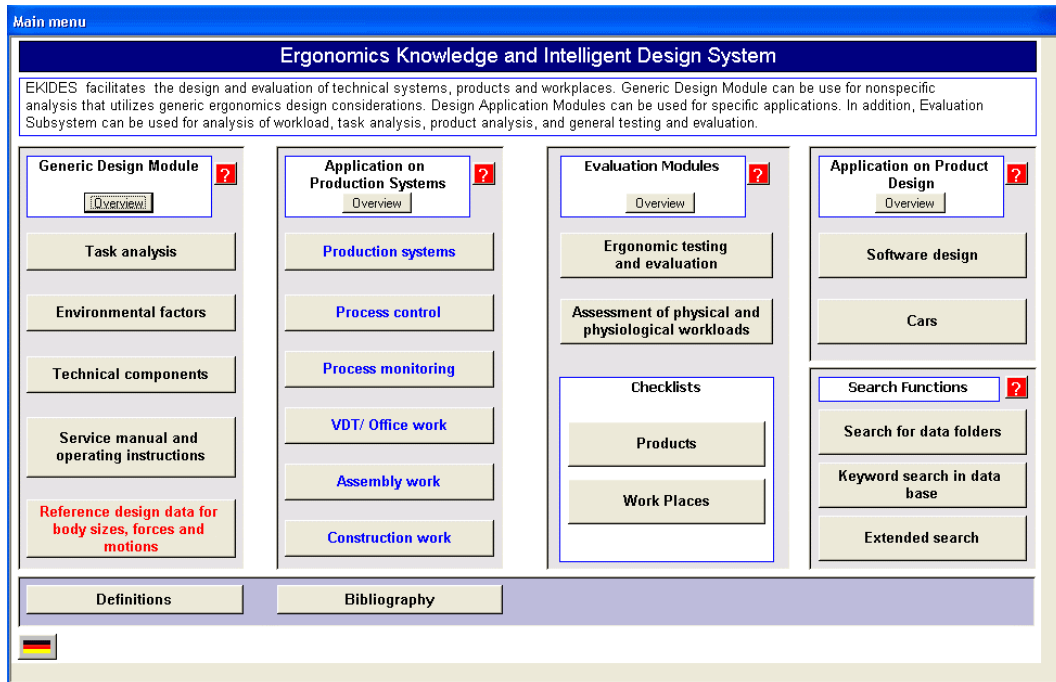


Figure 1: Main menu of EKIDES

In addition, “Evaluation Modules” can be used for analysis of workload, task analysis, product analysis, and general testing and evaluation. The user of EKIDES, sometimes will be confronted with the search of information about the specific facts relevant to his ergonomic analysis. In the part “Search Function” of EKIDES it is possible to search for data folders, keyword search and extended search for the user in such a situation.

### 1.1 Structure of Data

All ergonomic data in EKIDES are structured in the same way: For example design data about Software (in Product module – software design, see Fig. 2) contains 7 fields:

- Remarks to software ergonomics (with 6 data folders),
- User action and user guidance (with 9 data folders),
- Text editing and depicting of text (with 2 data folders),
- Processing and depiction of prescribed forms (with 2 data folders),
- Processing and depiction of tables (with 2 data folders),
- Processing and depiction of graphics (with 3 data folders).
- WEB - Design (with 6 data folders).

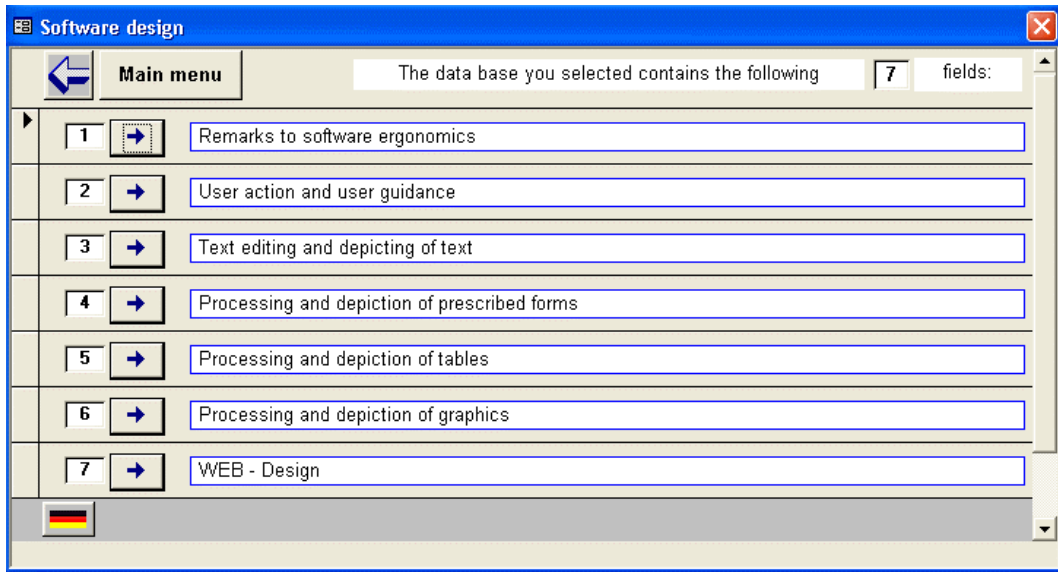


Figure 2: Software design, fields

The structure of data folders in all theme groups of database is similar (see Fig. 3). The first column on the left side (c), with a colour-coded sign, illustrates the meaning of the particular ergonomic requirement. Each requirement can be relevant overriding to health, safety, performance, reliability/dependability, or comfort. For this classification, in some cases no defined guidelines from law, regulations, or standards are available. In these cases, an expert judgment (ergonomists and company physicians) was used. In the second column, a given item (component) is verbally described.

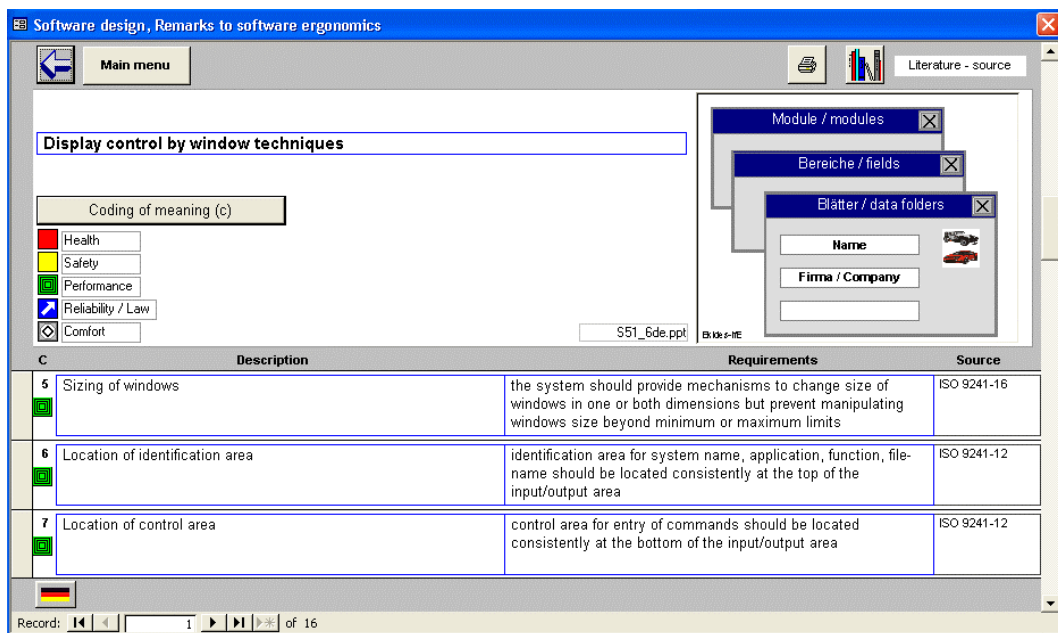


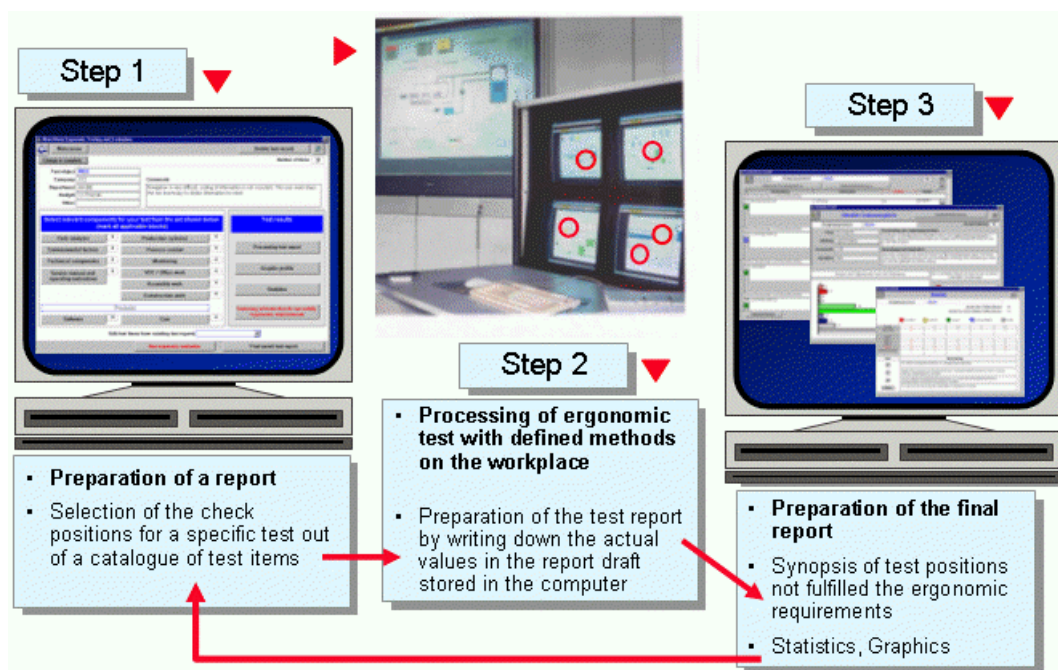
Figure 3: Data folder in EKIDES

The third column contains the ergonomic requirements. As much as possible, quantitative data are presented here. However, very often verbal assessments for the ergonomic requirements are indispensable. The fourth column contains information about the source of a reference. The general backgrounds for ergonomic requirements in EKIDES are international standards, books, and research reports.

## 1.2 Structure of Test and Evaluation Modules

### *Ergonomic Test and Evaluation*

This module offers the opportunity to prepare analysis with data sets from the Generic Design Module or the Application Modules in order to perform ergonomic tests and evaluations for a variety of workplaces, technical systems, and products (see Figure 4).



**Figure 4:** Three steps in ergonomic evaluation

A record form will be prepared automatically in which the data related to the test object can be prompted. The analyst has the opportunity to decide if the test results are in compliance with the ergonomic requirements. When the analysis is completed, the test report can be printed. A list of features which do not correspond to the ergonomic requirements is available. Additionally, a statistical evaluation of the results can be called up.

### *Application of Checklists*

Checklists offer the opportunity to examine workplaces or products with respect to the ergonomic requirements for design and layout. At present time, EKIDES offers checklists for six types of workplaces (VDT/ office work, process control, process monitoring, production systems, assembly work, and construction work) and five types of products (software in general, E-learning, WEB-design, cars, busses/coaches, trucks/heavy lorries and machinery for construction sites).

## **2. The Checklist and questionnaires for Software**

Normally used checklist for software comprises only questions the user must answer. The web-based user interface evaluation with questionnaires (Perlman 1998) has a very different structure and asks a different number of questions (10 heuristics questions from Nielsen, 1993 to 100 questions from Lin et al 1997).

By Chin et al 1988 is the following structure for an evaluation of user interface satisfaction with a 27 questions:

- Overall reaction to the software
- Screen
- Terminology and system information
- Learning
- System capabilities

By Lin et al 1997 are 100 questions presented in another structure:

- Compatibility
- Consistency
- Flexibility
- Learnability
- Minimal action
- Minimal memory load
- Perceptual limitation
- User guidance

The evaluation with questionnaires offers also a different sort of answer. There is a possibility of rating (0 points for bad software to 5, 7 or to 9 points for a good one), but also a composition of rating with an adjective for example (Chin et al 1988):

Overall reaction to the software:		
Terrible	0.....9	wonderful
Difficult	0.... 9	easy
Frustrating	0.... 9	satisfying
Inadequate power	0.... 9	adequate power
Dull	0.... 9	stimulating
Rigid	0.... 9	flexible

A further method of evaluation and help for enhance of functionality and universality of the WEB, 2002) is based of 12 main themes (<http://www.w3.org/TR/2002/REC-UAAG10-20021217/uaag10.html>). In the Table of Checkpoints for User Agent Accessibility Guidelines 1.0 (December 2002), the user can find three sections with priority. "Each checkpoint in this document is assigned a priority that indicates its importance for users with disabilities.

#### Priority 1 (P1)

If the user agent does not satisfy this checkpoint, one or more groups of users with disabilities will find it impossible to access the Web. Satisfying this checkpoint is a basic requirement for enabling some people to access the Web.

#### Priority 2 (P2)

If the user agent does not satisfy this checkpoint, one or more groups of users with disabilities will find it difficult to access the Web. Satisfying this checkpoint will remove significant barriers to Web access for some people.

#### Priority 3 (P3)

If the user agent satisfies this checkpoint, one or more groups of users with disabilities will find it easier to access the Web." The user of this checklist can read provisions about each checkpoint and can give comments with a notice of satisfaction. In a given case it could be very difficult to analyse the result of this checklist.

### 3. Evaluation of Software with the Checklist of EKIDES

The checklist evaluation of EKIDES is composed of three steps. At first the user must answer the questions referring problems he has with evaluated software (for example “usability” or “help functions”). At the second part of analysis, the system EKIDES offers the user a corresponding list of deficiencies. In this step of the evaluation the user makes a decision with “Yes”, “No” or “Not applicable”. In the third step the results of EKIDES checklists point out the deficiencies according to performance, reliability or comfort. The structure of the checklists for software design in general and WEB design in EKIDES is presented in table number one.

Fields of evaluation	Check positions for Software in general	Check positions for WEB design
Usability	12	4
Information display	20	5
User guidance	22	13
Error management	15	4
Help functions	10	4
Dialogue technique	11	6
Menu dialogues	10	10
Command dialogues	15	-
Documentation	8	-
Coding - relevance for:		
performance	105	40
comfort	11	4
reliability	7	2

**Table 1:** Structure of the checklists for software design generally and WEB design

All check positions are coded in their relevance for performance (105/40), comfort (11/4) and reliability (7/2). The Checklist for E-Learning is based on 60 check positions in 7 fields, for example processing didactical helps, motivation and so on. The evaluation of software with the checklist in EKIDES is very easy. Only three steps are to be done, however in order to check about 123 positions in the software under analysis the user needs about 60 minutes. In the next captions, a step-by-step description of the application of the checklists analysis allows to understand the facility of this method.

### 3.1 Start a New Product Analysis

The user starts a new analysis immediately from the main menu of software design (see Fig. 5).

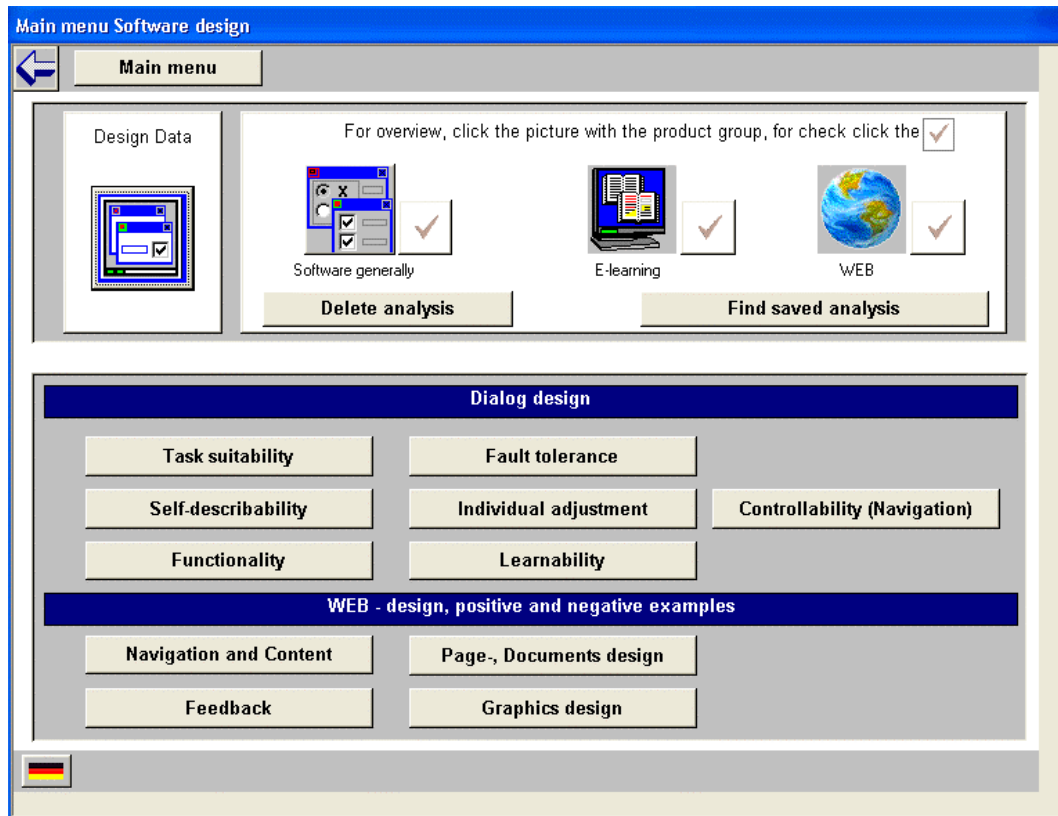


Figure 5: Main menu software design

In the first step the user answers the main questions (for example: “Are there any problems with usability?”). The problem groups as “user guidance” or “menu dialogs” will be explained, if the user clicks the command button with “?”. The corresponding definition of “usability” is shown in Figure 6 in the small window.

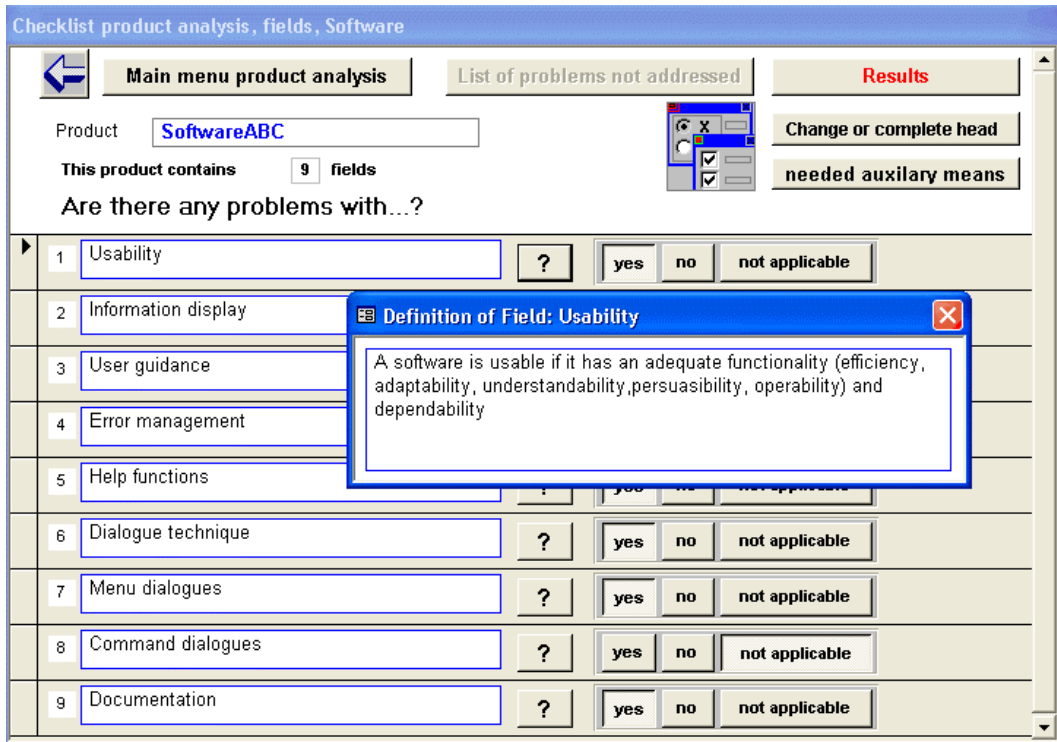


Figure 6: Start of checklist analysis

### 3.2 Selection of relevant Fields and Check Positions

After selection of each one problem (field) the window with a list of deficiencies (check positions) will be open (Fig. 7). The answers based of each check position are in this context focussed with “yes”, because the user decided he has problems with the usability of the software.

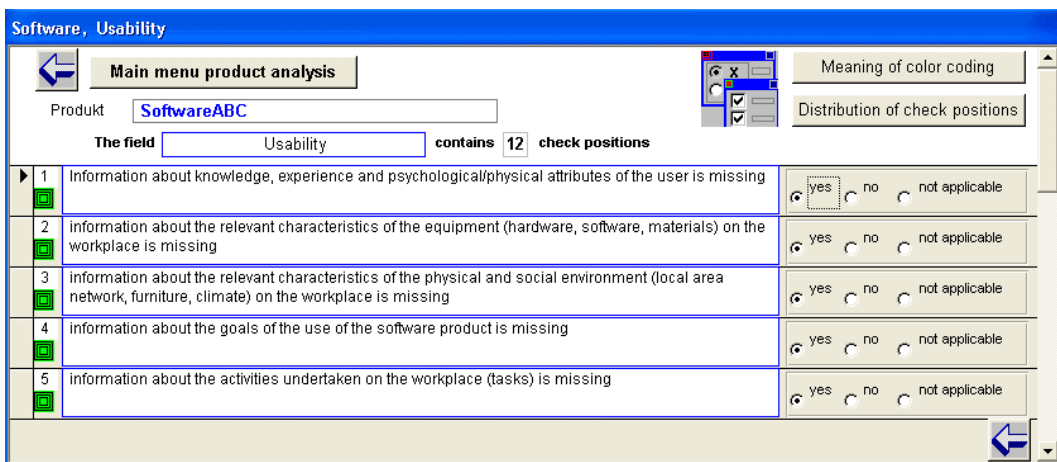


Figure 7: Check positions, deficiencies with usability

He can change his opinion and put another answer to each checkpoint. In the same procedure the user can decide, the field “command dialogues” is not applicable for evaluation. By this, in the window with checkpoints all answers will be set automatically to “not applicable”.

### 3.3 Results

After answering all the statements in the selected fields, the user can open a window with summary of results. He can see again the contents of analysis and change the answer he has given during the evaluation session if necessary or he can view or print out only deficiencies in evaluated software (see Fig. 8). In the left side in window “Summary of results” there are the command buttons with them the user can open the windows showing him content of analysis. He can see all positions sorted by fields or by criteria or he can obtain the view of problems critical to performance. The summary of deficiencies with evaluated software is available with the clicks of the buttons right in the windows.

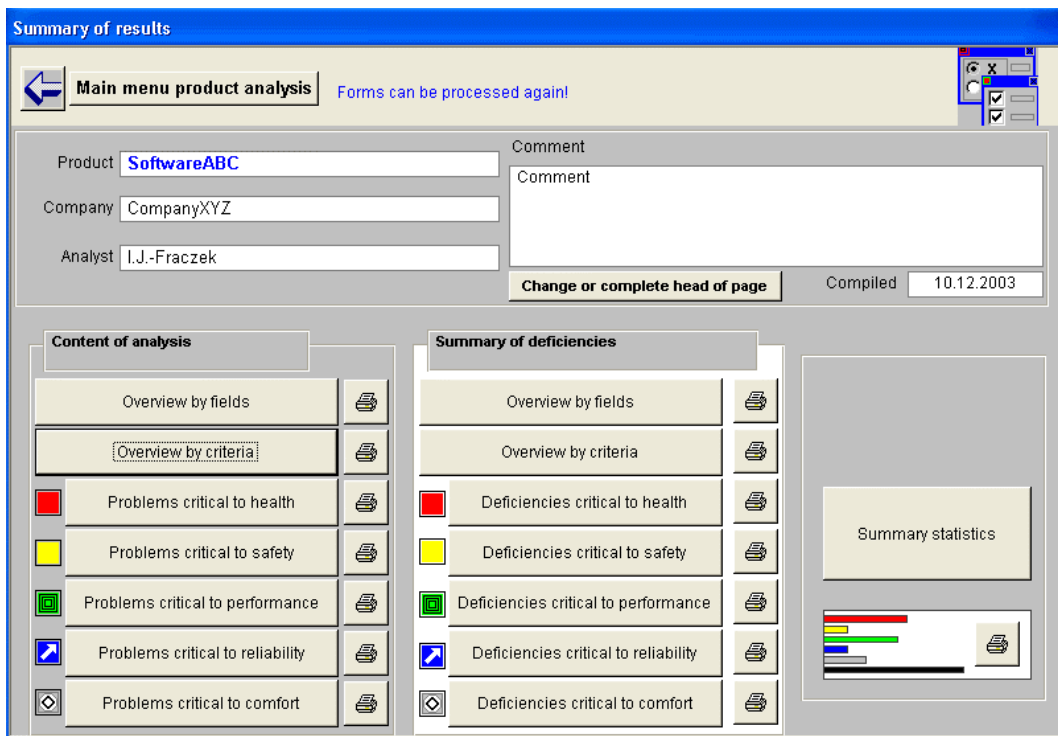
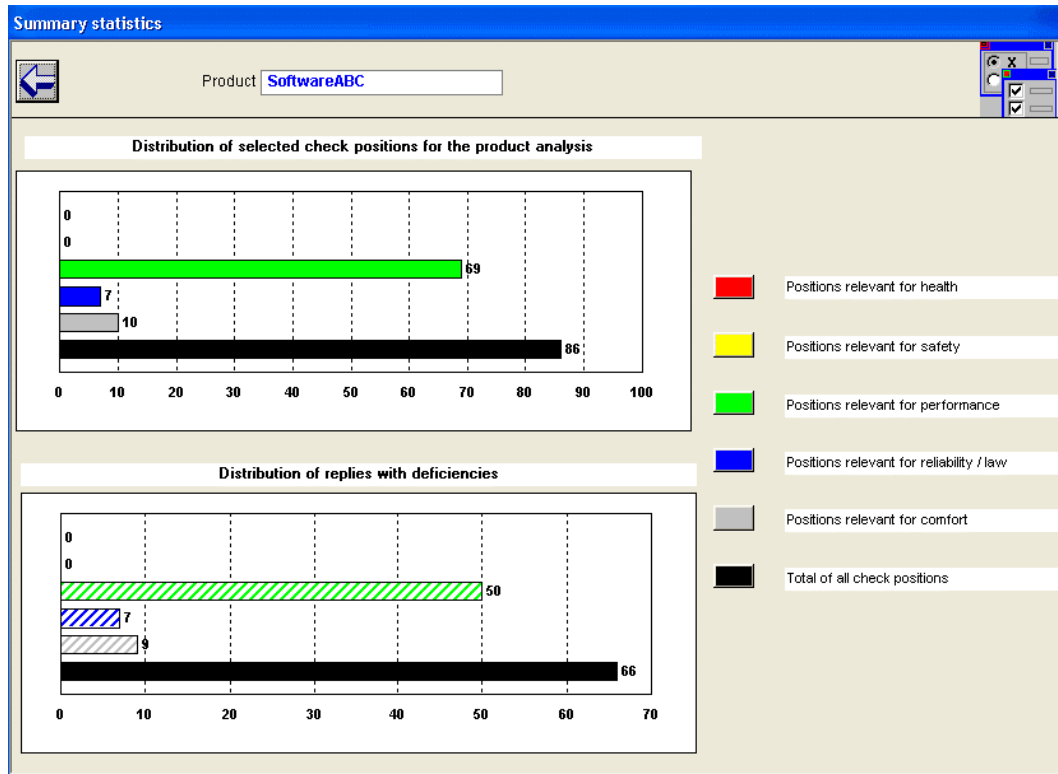


Figure 8: Result of the checklist

The button "Summary statistics" opens a graphic representation of the results (see Fig. 9).



**Figure 9:** Graphic representation of results.

These positions are classified according to the specific evaluation criteria. In the lower part of the graphical display one can see how many of the positions were answered with "yes", and, therefore, were classified as "deficient" (66). The positions that were answered with "not applicable" are not included in the summary statistics. Therefore, the total number of positions (123) and the total number of answers (86) will differ by the number of positions "not applicable". The complete test results can be printed as a test report containing graphic illustration of the full scale evaluation. To compare the software with the checklist of EKIDES, the graphic form of analysis gives a first impression. The study of the check positions in each group (e.g. performance, comfort) allows the designers to identify all deficiencies of the software.

#### 4. Use of EKIDES in the company and at the university

The database EKIDES is comparable with the existing ones only under consideration one aspects: obtain necessary data for design (or list of ergonomic requirements). But any others database offers the user an evaluation options. The most important attribute and advantage of EKIDES is the possibility to use the database as an evaluation tool.

The user can compose and include data in evaluation test and use it immediately as a tool for assessment.

The ergonomic database system EKIDES, including several computer aided evaluation methods, has been used in several German companies and offers the following advantages:

- ⇒ Current information on ergonomic requirements in the design process of products and workplaces always available to design-engineers,
- ⇒ Company specific collection of workplaces providing optimal conditions, documented in - house by means of EKIDES.
- ⇒ Documentation of omissions to be avoided in the future and improvements which should be taken into consideration for the design process

The application of EKIDES in the company and also at the university can certainly assume a teaching function. The analysis of work places and products by means of the system arise the interest in house ergonomic design engineering of staff and employees at the same time. In the next future it will be develop a new short method of subjective user evaluation of the software, based of the software assessment spider-diagram.

## **5. References**

- Bubb, H. Jastrzebska-Fraczek I.: Software Evaluation by the ergonomic assessment tool EKIDES. In Proceedings of 10 the International Conference HCI, Volume 3, Human-Centered Interaction: Cognitive, Social and Ergonomic Aspects. S. 1218-1222, LEA, Mahwah, New Jersey, London 2003
- Chin, J.P., Diehl, V.A., Norman, K.L. (1988) Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. ACM CHI'88 Proceedings, 213-218.
- Davis, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13:3, 319-340
- ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs); Beuth, Berlin (Status 12 Jan 2001) (Part 1, 8, 10-17)

Nielsen, J. (1993) Usability Engineering. Academic Press. Chapter 5, p. 115.

Perlman, Gary: Web-Based User Interface Evaluation with Questionnaires  
<http://www.acm.org/~perlman/question.html>

Purdue Usability Testing Questionnaire

<http://www.acm.org/~perlman/question.cgi?form=PUTQ> BASED ON: Lin, H.X.  
Choong, Y.-Y., and Salvendy, G. (1997) A Proposed Index of Usability: A Method for  
Comparing the Relative Usability of Different Software Systems. Behavior &  
Information Technology, 16:4/5, 267-278

Schmidtke, H., Jastrzebska-Fraczek, I.: The ergonomic database system (EDS) – an  
example of computer-aided production of ergonomic data for the design of technical  
systems. In: Landau, K. (Ed.) Ergonomic Software Tools in Product and Workplace  
Design. Verlag ERGON, Stuttgart 2000.

# Evaluating Collaboration and Core Competence in a Virtual Enterprise

Rainer Breite and Hannu Vanharanta \*

Tampere University of Technology, Pori, Finland

---

## ABSTRACT

During the Spring of 2000 and Autumn 2000 and 2001 students with a bachelor's degree in engineering participated in an estimation which was intended to investigate the possibilities to manage a virtual company by means of collaboration. The test subjects developed a virtual company operating on the Internet with suppliers available on the Internet. The test subjects' satisfaction and perception of the estimation were solicited. The task was complex, but taking into consideration the test subjects' background, their degree in engineering, as well as their several years of industrial experience, these students, if any, should have the capabilities to manage the task given and thus provide us with some valuable information about the possibilities of collaboration and virtuality in the Internet.

---

Keywords: *company management, collaboration, virtual enterprise.*

Received 3 October 2003; received in revised form 10 December 2003; accepted 20 December 2003.

## 1. Introduction

In our study the decision-makers are working through several www addresses in order to create a virtual company. Many of the found www addresses are discarded, while others are brought up for a more thorough inspection. Our main interest is to find out whether collaborative and virtuality is possible on the Internet. The main research question can be divided into the following sub questions: 1) Is it possible to find necessary suppliers on the Internet? 2) Is it possible to evaluate different suppliers on the Internet? 3) Is it possible to create a virtual enterprise on the Internet? 4) Does the Internet environment provide sufficient support for the decision-maker? We assume that the fourth question forms an essential part of the other three questions, as the decision maker needs support for his/her decision making process.

---

\* Corresponding Author:  
Rainer Breite,  
Tampere University of Technology, Pori  
Pohjoisranta 11  
P.O. Box Pori, Finland  
Email: rainer.breite@tut.fi

By using the hyperknowledge platform (Chang et al., 1989, 1993, 1994; Vanharanta et al., 1995), we assume that we can deal with the structuring and processing of decision problems, which involve creating the virtual enterprise. We start from the idea that our virtual enterprise is an object, which has to be formed on the knowledge base. Thus the decision maker needs support for his/her decision making process, through which s/he forms the virtual enterprise and affects the contracts between the decision maker / the virtual company and the supplier. The decision maker has to evaluate the ability to make the contracts solely by using the information which s/he gets from the computer screen. Thus we address the usability and the utility of the Internet in the context of collaboration by utilizing the hyperknowledge platform. Usability and utility are two equally important factors in the evaluation process. Newell et al. (1972) define utility as "the question of whether the functionality of the system in principle can do what is needed". Usability, in turn, is usually described in terms of criteria like learnability, efficiency of use, memorability, a small number of errors, and subjective satisfaction (Nielsen 1990).

The present study focuses on the two most critical components of the hyperknowledge environment: the user and the contents of the Internet applications. We try to find out whether the Internet applications used are a plausible system for evaluating the decision support activities of business contracts, and what are the advantages and disadvantages of Internet applications from the user's point of view. The validation methods are partially the same as those which have been used in expert systems (O'Keefe et al. 1987 and O'Leary 1988) and in the hyperknowledge system evaluations (Vanharanta et al. 1995), i.e. a form of performance validation combined with a questionnaire.

## **2. Theoretical Framework**

### **2.1 Hyperknowledge and Decision Support System**

The concept of hyperknowledge is wider than those of hypertext or hypermedia, although it follows the same general principles. Hyperknowledge is an ideal working and learning environment that holds knowledge and, at the same time, defines the nature of hypertext and hypermedia. The user can navigate freely in this environment (as on the Internet), and widen his/her own knowledge. (Chang et al. 1989, 1993, 1994). The basic goal of this framework is to serve active decision support that

enables the decision maker to participate actively throughout the decision process. The framework takes into account that the decision maker cognitively possesses and processes many diverse and interrelated pieces of knowledge e.g. procedural knowledge, descriptive knowledge, reasoning knowledge, etc. The user (i.e. the user's mind) is able to deal freely with and control these different pieces of knowledge. (Chang et al. 1989, 1993, 1994)

A decision support system (DSS) (See Dos Santos et al. 1989, p. 3), in turn, consists of three main components: *The Language and the Presentation System* mediate messages to and from the decision support system. The language System (LS) on the Internet is controlled by using a hypertext transfer protocol (http). The user activates a www page and a www address by using a mouse and keyboard. The presentation System (PS) on the Internet application is everything presented in www page/screen format and which can be printed in paper form. *The Problem Processing System (PPS)* handles all the user requests or responses to and from the various knowledge sources in the system. On the Internet this all means search functions and processes by the Internet application. *The Knowledge System (KS)* contains all the decision support system's knowledge and it stores, in groups, concepts that are related to each other by definition and/or by association. With the Internet applications this means that all available contract information, knowledge and procedures in www addresses, sites, and pages are placed in this system. Thus the decision support systems (DSS) should be a natural extension of the decision maker's internal activities.

## **2.2 Collaboration in the Internet**

In our case collaborative has been defined as follows: "The ability of two or more people or groups to transfer data and information with the capability of on-line interaction. The distinguishing feature is the ability for many-to-many interactions and information sharing, unlike e-mail where the interaction is one-to-one or one-to-many" (<http://www.collaborate.com>). Ideal collaboration in the virtual context means that it is possible to create a virtual enterprise by using the Internet. The virtual enterprise, based on collaborative contracts on the Internet, can then control and evaluate all the needed transactions. This, in turn, means that a company has the ability to form a network, which consists of suitable suppliers and distributors and the members' core competences support the company's core competence. In this context, the abilities of collaboration were estimated by taking the first step on the path of collaboration. In the

laboratory tests the companies' abilities and capabilities to collaborate were estimated with the help of the information on the www page.

### **3. Empirical Study**

We made three separated tests, where we conducted an exploratory study on M.Sc. (Eng.) students' capabilities to create a virtual company on the Internet. The first estimation made in the spring of 2000 (marked with S00), the second in the autumn of 2000 (marked with A00), and the third in the autumn of 2001 (marked with A01). Altogether (18 (S00); 30 (A00); 25 (A01) M.Sc. (eng.) students with a bachelor's degree in engineering participated in the estimation. All of the students have industrial engineering as their academic major. The mean age of the participants was (35 (S00), 28 (A00), 29 (A01)) years and the participants had on average (9.2 (S00), 3.6 (A00), 6 (A01)) years of industrial experience. The majority of the test subjects (i.e. 93 %) visit the Internet daily or several times per week. The test subjects are not experts in purchasing management but all are quite well grounded in the theories of purchasing management and they understand several concepts of contracts (cf. User knowledge in a hyperknowledge environment). In every test case the test subjects had already planned their own "case company" and the products which were needed via the Internet. The task given to the test subjects was twofold; (i) to create a virtual company on the Internet and (ii) to report on their experiences, and perceptions during the task. The test subjects represented the knowledge managers and all the contracts were to be made between the knowledge manager and the different suppliers. When choosing the suppliers a requirement was that the transportation costs should be reasonable when taking the geographical distances into account. The test subjects were asked to construct the virtual company organization, i.e. construct a flowchart of different suppliers with their www addresses. This flowchart was to be designed in a logical way and delivered as a report at the end of the estimation. No e-mail contacts were allowed during the estimation between the test subjects and the suppliers. The estimation was conducted in one of the PC-classes at Tampere University of Technology in Pori. The estimation started at 8 a.m. and the test subjects were given 8 hours to complete the given task.

### **4. Results**

Our utility analysis focuses on research constructs, which are based on the hyperknowledge framework, and the validity and utility model for the hyperknowledge environment.(c.f. Vanharanta 1995). In this model the knowledge of the system has been divided into five different types of knowledge: linguistic knowledge e.g. computer explanations, descriptive knowledge e.g. fact data etc., procedural knowledge, reasoning knowledge and presentation knowledge (c.f. Chang et al. 1994). Our research paper is only concerned with descriptive knowledge, thus we examine how the information which is transmitted via the www pages, supports the user's evaluation of business under his/her decision-making. We also refer to the statistical results calculated from the questionnaires. The test subjects' assessments of their own performance was measured on (i) the overall satisfaction rate, (ii) attitudes towards the task and (iii) how experienced they were with the Internet. The attitudes towards the Internet as a source for creating a virtual furniture company were measured with the following factors, (i) finding relevant suppliers, (ii) making comparisons between the suppliers, (iii) choosing the suppliers, and (iv) the usability of the suppliers' web sites. The responses were mainly measured using the Likert attitude scales (5-point scales ranging from "strongly agree" (5) to "strongly disagree" (1)).

In this paper the above mentioned research constructs used were divided into three different categories: The first category describes traits and feelings perceived by the test subjects themselves during and after the tests. In our study these constructs are "lost in space", and "cognitive overhead". The second category describes perceived outcomes for the test subjects interacting with the Internet application, e.g. what was the contribution of the virtual environment for the user. In our study these constructions are: "learning", the "creation of comprehensive understanding", "understanding the Internet and its applications", "usefulness in acquisition", and the "utility of the Internet". The third category (satisfaction) describes user satisfaction. (cf. Conklin 1987; Vanharanta et al. 1995)

The task was completed by all the test subjects, however with varying results. The number of suppliers needed for a virtual company was on average 12, the maximum number being 20 and the minimum being 6. The number of visited web sites also varied. On an average the test subjects worked through 9 out of 12 sites when creating the virtual company, the maximum being 13 and the minimum being 5 sites.

#### **4.1 Performance Validity and Utility Assessment**

Finding the suppliers on the Internet. According to Vanharanta et al. (1995) one of the problems in hypertext and hypermedia applications is the feeling of being “lost in space”. This means that the users lose their “coordinates”. In Internet applications, especially www sites, this means that the user is unable to find “the right page” and s/he is not sure, where s/he exactly lays and what the context of the www page is. This means, that the user does not have any familiar framework. (cf. Lost in space in a hyperknowledge environment). In the questionnaire we asked for the subject’s opinion by evaluating his/her ability to find “the best supplier” among all the suppliers. On an average the test subjects did not regard finding the suppliers on the Internet as too difficult, when the mean values were (2.9 (S00); 2.7 (A00); 2.8(A01)), the std. deviation values being (0.97 (A00); 0.91 (A01)). The lowest limits for the t-distribution were (2.35 (A00); 2.43 (A01)), the upper limits being (3.08 (A00); 3.18 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar.

Making comparisons between the suppliers. We tested learning by asking the test subjects to estimate their ability to make contracts with suppliers. So the test subjects had to evaluate suppliers by using the information which was available on the suppliers’ www sites. (cf. Learning and creation of comprehensive understanding in a hyperknowledge environment). Once test subjects find the suppliers they need to make comparisons between them. The comparisons were to be based solely on the information available on the Internet. Intuitively this phase would be more difficult than finding the suppliers on the Internet. The results provide some evidence for this but the difference is not significant. We did, however, find without exception that the average values for making comparisons are in some cases higher than those for locating finding suppliers on the Internet (means 3.3 (S00); 2.6 (A00); 2.9 (A01)), the std. deviation values being (0.76 (A00); 0,59 (A01)). The lowest limits for the t-distribution were (2,33 (A00); 2,67 (A01)), the upper limits being (2,90 (A00); 3,16 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar.

The usability and utility of the suppliers’ web sites. Our purpose was to indicate how well the user can handle the knowledge which the Internet contains, how aware s/he is of that knowledge, and how close the Internet can be to the ideal hyperknowledge

environment. We tested the utility of the Internet by asking about the usability of the suppliers' websites. On an average the test subjects did not find the usability of the suppliers' web sites too problematic when creating the virtual furniture company, mean values 3.1(S00); 2.8 (A00); 3.0 (A01), and the std. deviation values being (0.95 (A00); 0.46 (A01)). The lowest limits for the t-distribution were (2.46 (A00); 2.76 (A01)), the upper limits being (3.17 (A00); 3.14 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar. This means that the quality of the web sites, the relevant information not provided on the site, such as price and delivery terms, no interactive sites, missing company information which does not give a very reliable picture of the suppliers, just to mention a few.

Satisfaction and expectations. On an average the participants considered the estimation to be realistic in comparison to things that individuals and organisations do in a typical or common business situation, the mean values being (2.3 (S00); 3.3 (A00); 3.3 (A01)), and the std. deviation values being (1.17 (A00); 0.79 (A01)). The lowest limits for the t-distribution were (2.09 (A00); 2.95 (A01)), the upper limits being (4.47 (A00); 3.61 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar. On an average the students were not very happy with their performance, the mean values being (2.6 (S00); 3.0 (A00); 2.7 (A01)), and the std. deviation values being (0.96 (A00); 1.10 (A01)). The lowest limits for the t-distribution were (2.45 (A00); 2.14 (A01)), the upper limits being (3.55 (A00); 3.27 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar. When asked about it, the majority of the participants 61% (S00); 64% (A00); 56% (A01) thought that it is possible to manage a virtual furniture company, 28% (S00); 28% (A00); 32% (A01) being not sure of the idea. The majority of the students thought, based on their experience from creating the virtual company that it would be possible to agree on co-operation in general and on contracts on either a permanent or a temporary basis. Considering partnership the opinions were divided. In general the test subjects did not find the effort to be too demanding, mean values (2.4 (S00); 2.8 A00); 3.1 (A01)), and the std. deviation values being (0.71 (A00); 0.58 (A01)). The lowest limits for the t-distribution were (2.23 (A00); 2.84 (A01)), the upper limits being (3.37 (A00); 3.33 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar. The level of interest as

well as the challenge of the task were quite good, both scoring (3.05 (S00); 4 (A00); 3.5 (A01)) on average), and the std. deviation values being (0.91 (A00); 1.12 (A01)). The lowest limits for the t-distribution were (3.62 (A00); 3.06 (A01)), the upper limits being (4.38 (A00); 3.98 (A01)). The std. deviation and t-distribution values in the (S00) case were not calculated, but we suppose that these values are similar.

## **5. Conclusions**

We can conclude, that there were no problems in finding the necessary suppliers on the Internet. The problems start with the evaluation of the suppliers, i.e. comparing and choosing the suppliers. The fact that the participants were not allowed to send e-mails to the suppliers made the task difficult. If the necessary information was not available or not found on the suppliers' web sites then the participants had to live with that and make the comparisons as well as choosing the best suppliers without proper information. According to our findings we can see a slight trend of growing complexity from finding to comparing and to choosing the best suppliers. The differences are, however, not significant.

All the students, with the exception of six, managed in principle to create a virtual company. Whether these companies would in practice be manageable was not tested. The participants were rather positive to the idea of managing a virtual company. Their opinions were divided regarding partnerships but the majority thought that it is possible to agree on co-operation in general and on contracts on either a permanent or a temporary basis. According to the user satisfaction results we can conclude that users seem to need more relevant and supporting information for their decision- making.

The paper illustrates the theory connection between hyperknowledge and the Internet in the context of collaboration in the virtual environment. We can conclude that utilizing the constructs of "lost in space", "learning", "creation of comprehensive understanding", and "utility of the Internet" supported the estimation of the Internet as an implement of collaboration in the virtual working environment.

The results and experiences suggest that tests should be extended into real companies and business situations. Moreover, in the framework of the theories used, it will be fruitful to examine how different forms of information on the www page support decision-making in the context of collaboration.

## 6. References

- Chang, A-L. Holsapple, C.W. Whinston, A.B. (1989) A Decision Support System Theory. *Working paper University of Aritzona* Tucson, USA, 53p.
- Chang, A-L. Holsapple, C.W. Whinston, A.B. (1993) Model Management Issues and Directions Decision Support System 9(1) pp.19-37
- Chang, A-L. Holsapple, C.W. Whinston, A.B. (1994) A Hyperknowledge Framework of Decision Support Systems. *Information Processing & Management*, 30 (4) pp. 473-498
- Conklin, J. (1987) Hypertext: An Introduction and Survey. *IEEE Computer* 20(9) pp. 17-41
- Dos Santos, B.L. Holsapple, C.W. (1989) A Framework for Designing Adaptive DSS Interfaces *Decision Support Systems*, 5 pp. 1-11. <http://www.collaborate.com>
- Newel, A. Simon, H. (1972) Human Problem Solving Prentice Hall, Enlewood Cliffs, New Jersey
- Nielsen, J. (1990) Hypertext and Hypermedia Academic Press Inc., San Diego
- O'Keefe, R. Balci, O. Smith, E. (1987) Validating Expert System Performance. *IEEE Expert* 2(4) pp. 81-89
- O'Leary, D. (1988) Methods of Validating Expert Systems. *Interfaces* 18(6) pp. 72-79
- Vanharanta, H. Käkölä, T. Back, B. (1995) Validity and Utility of Hyperknowledge – Based Financial Bechmarking System. *Proceedings of the 28<sup>th</sup> Annual Hawaii International Conference on System Sciences*, pp. 221-230.

# A Case Study for Reaching Web Accessibility Guidelines for the Hearing-Impaired

\*Miki Namatame\*, Makoto Kobayashi, Akira Harada

Department of Design  
Tsukuba College of Technology, Ibaraki, JAPAN

---

## ABSTRACT

This research is a case study, but it has the possibility to make Web Accessibility guidelines for the hearing-impaired. Web Content Accessibility Guidelines by W3C suggests "Provide non-text equivalents of text". We suppose this guideline is applicable to the hearing-impaired people. Our research expanded W3C's guidelines to specify that the "non text equivalents" convey a significant meaning about the contents accurately. We performed experiments and considered how different web designs effected WEB accessibility of people with different disabilities. We recorded the operation flow and the length of stay on each page as well as eye tracking and mouse movement. We found some differences between non-impaired and hearing-impaired people in this experiment. We have considered that two user types exist Text-oriented for the non-impaired vs. picture-oriented for the hearing-impaired. We formed hypothetic Web Accessibility Guidelines for the hearing-impaired people based on this case study.

---

Keywords: *WEB accessibility, WEB Design, Hearing-impairment, Information Design, Interface Design.*

Received 19 October 2003; received in revised form 20 December 2003; accepted 20 December 2003.

## 1. Introduction

W3C's WAI recommendations have been establishing guidelines for the standardization of WEB technology used in building WEB-sites <sup>□□</sup>. These guidelines can be applied internationally, however, as yet, W3C suggests only to "Provide equivalent alternatives to auditory and visual content" <sup>2□</sup> for hearing-impaired people. Specifically, this guideline has been stated as the following: "Describe the sound of

---

\* Corresponding Author:  
Miki Namatame  
Tsukuba College of Technology,  
Department of Design  
4-3-15 Amakubo  
Tsukuba Ibaraki, Japan.  
Tel: +81 29 8589352  
Fax: +81 29 8589352  
E-mail: miki@a.tsukuba-tech.ac.jp

auditory content" <sup>3</sup> and "Provide non-text equivalents of text" <sup>4</sup>). The different guidelines (the US Government Section 508) related to Web accessibility standards are suggested to hearing-impaired people. The guidelines recommend attaching the synchronized caption for hearing-impaired people in the case of audio material, video material, and multimedia material <sup>5</sup>). The problems of accessibility for hearing-impaired people used to be mentioned only about sound. These are because the guidelines have been supposing that non-impaired people and hearing-impaired people have little difference in use (except for sounds) of WEB contents. We have observed how hearing-impaired use WEB at our college for several years. Do non-impaired & hearing-impaired people have the same WEB content needs? We hypothesize that the differences in use of WEB contents of non-impaired people and hearing-impaired people are not only sounds. Generally, hearing-impaired people have less vocabulary than non-impaired.

We can see such a phrase for urging an understanding about hearing-impaired people. "Rephrase rather than repeat a misunderstood sentence" <sup>6</sup>). We have to consider that this problem is related to the text information of web pages. We think it important to investigate the hearing-impaired person's computer operation characteristics.

## **2. Purpose**

We had three purposes. One was to examine the operation processes of the research samples on the sample homepages, and to extract the key items of information. The 2<sup>nd</sup> was to investigate the features of the operation and recognition of people with different disabilities. And the 3<sup>rd</sup> was to form Web Accessibility Guidelines for the Hearing-Impaired.

## **3. Method**

On the Internet, the conclusive factor of practical use is to find information efficiently. How do people find the required information? We observed the characteristics of information selection. We performed experiments and considered how different web designs effected WEB accessibility of people with different disabilities.

We choose some goals from existing websites with different structures. These goals were simple choices with no comparisons or judgments. We established some items and classified the homepages. We choose some standards such as Amount of information, Balance of texts and pictures, Picture Types, Motion, Navigation Type, Navigation Layout, and Page divisions, then choose good websites. The aim was to research how quickly a specified goal was attained using existing web designs (see Table 1). Existing homepages were Automaker, Official Approval, Book Store, and Cosmetics. We gathered 6 persons for experiment (3 non-impaired persons & 3 hearing-impaired persons).

We used equipment to track eye movement and the computer operation processes were recorded on video. This system has video based eye and target tracking equipment using infrared lighting <sup>7)</sup>. Personal computer with 17 inch monitor, was operated by the examinees using only the mouse. Since this equipment requires no head-mount or glasses, tests were performed under natural conditions. We observed the mouse location, eye movement and time taken by video records (see Figure 1).

**Table 1:** Structure of the Home pages.

	Amount of information	Balance of texts and pictures	Picture Types	Motion	Navigation Type	Navigation Layout	Page divisions
Automaker	Medium	Equal	Photograph (car only)	Animation	Text	Left	Right-left
Official Approval	Medium	Equal	Icon	None	Text	Left	Right-left
Book Store	High	High text information	Photograph & Icon	None	Index tab / Text	Top / Left	Top & right-center-left
Cosmetics	Low	High picture information	Photograph / Illustrations / Logos	Flash movie	Picture / Text	Center	None



**Fig.1:** Examinee (left) and researcher.

## 4. Results

About each website; Since the “Major Book Store” website had the function showing the genre perused last time, it was not applicable to the experiment. There were no differences in operation when accessing the “Official Approval Guidelines” website and “Major Cosmetics Maker” website. On these websites, the subjects were asked “Please search the schedule of examinations” or “Please choose your favourite one from new lipsticks”. When accessing the “Automaker” website, differences were seen in access time and operation characteristics. On this website the subjects were asked, “Please choose the color which you think is the most suitable for the indicated type of car. So we analyzed the “Automaker” website.

### 4.1 The operation flow and length of stay on each page

Figure 2 is the operation flow and length of stay on each page of the “Automaker” website. Only one route can reach the right position. It is indicated with the bold lines. The left box shows “Examinee A” ’s operation and the right box shows “Examinee B” ’s operation. “Examinee A” had a no loss operation flow and stayed on the homepage for a longer time with accurate operation procedures. “Examinee B” ’s confusion can be observed from an operation flow. This figure shows the results of “Examinee A” (shortest time) and “Examinee B” (longest time) on the “Automaker” website. The time required to complete the task was 100sec by “Examinee A”, and 350sec by “Examinee B”. The number of jumps between pages were 3 times by “Examinee A”, and 22 times by “Examinee B”. Except the top page, “Examinee B” ’s lengths of stay were shorter than “Examinee A” ’s lengths.

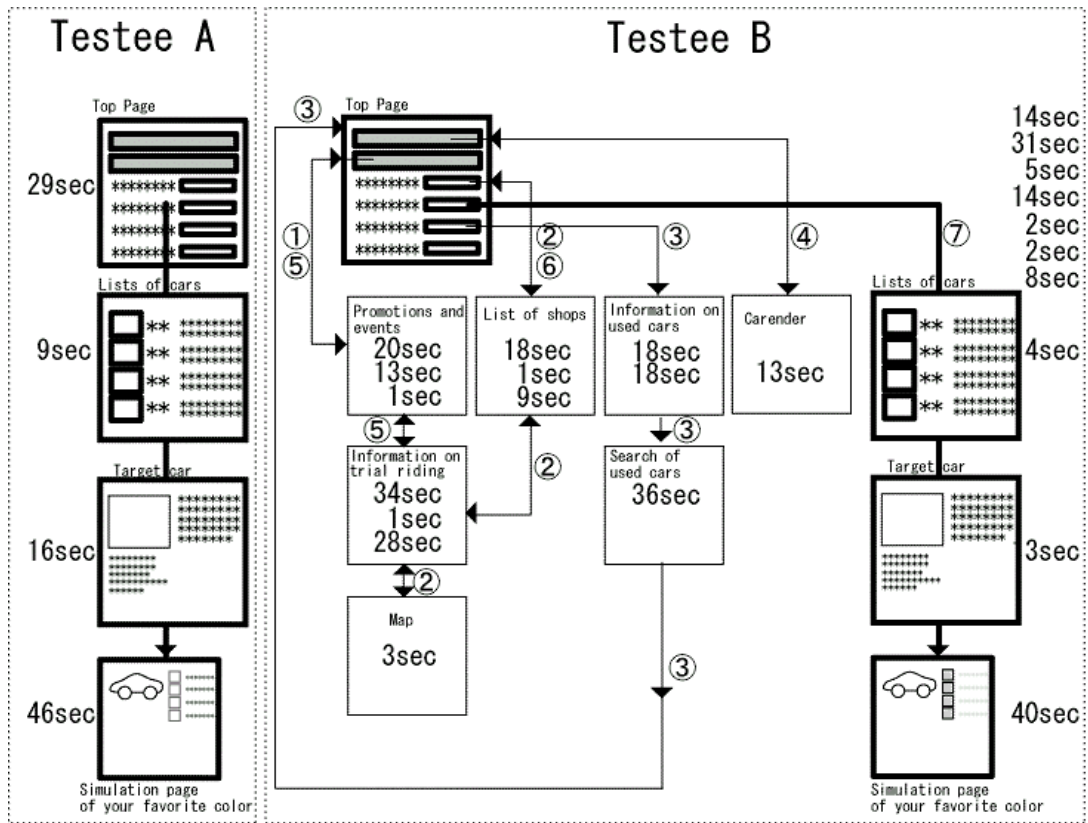


Fig.2: The operation flow and length of stay on each page of the “Automaker” website.

#### 4.2 The operation tendencies of each Examinee (Home Page)

We analyzed the operation tendencies of each examinee. On Figure 3 images the orange lines represent the eye movement while green lines show the mouse movements.

Figure 3 can be seen on "Examinee A" 's image, on the left, that text was read and the appropriate link was used. The eye movement clearly showed the operation of reading the text on each page. "Examinee A" used the text links. "Examinee B" 's image shows inefficient tracking of both the eyes and the mouse. "Examinee B" 's eye movement was broad and sweeping. "Examinee B" had a tendency to operate animation and large picture links with a top to bottom sequence. Small font text links were used much less than picture links.

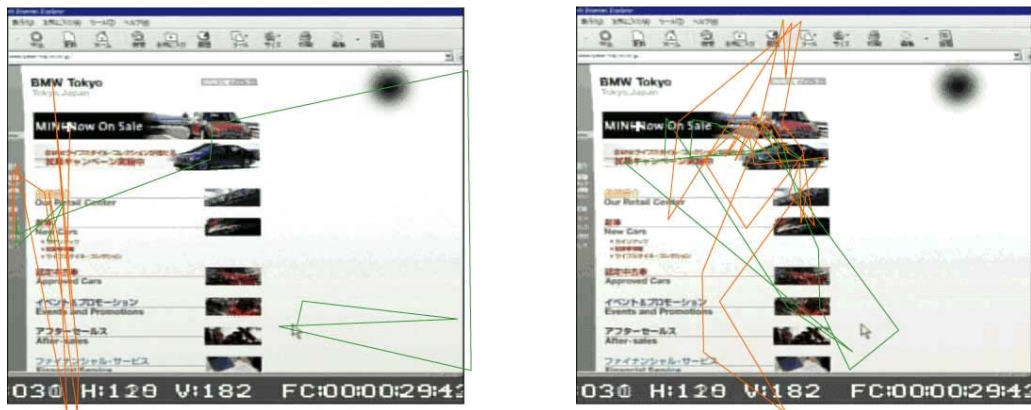


Fig.3: The Mouse Location & Eye Movement of the homepage.  
"Examinee A" (left) and "Examinee B".

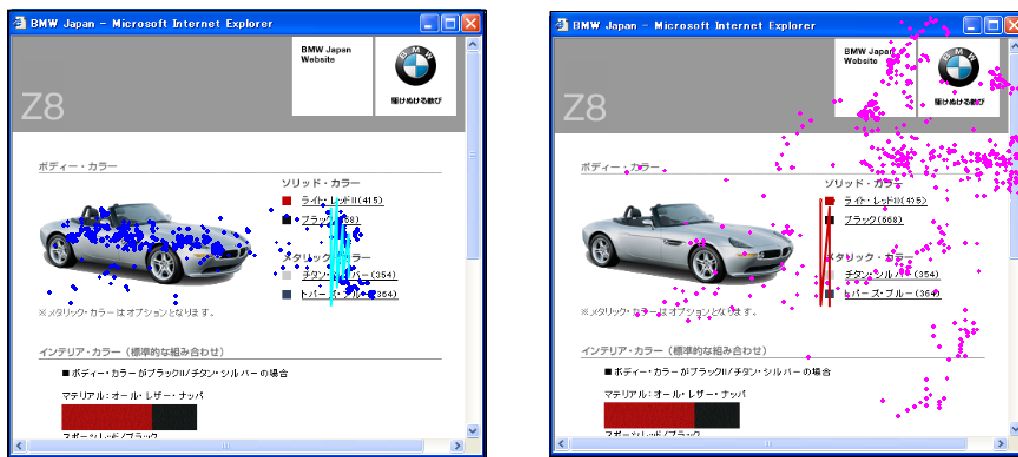


Fig.4: The Mouse Location & Eye Movements of the Color Determination page.  
"Examinee A" (left) and "Examinee B".

### 4.3 The operation tendencies (Color Determination page)

After the experiment, we questioned "Examinee A" about the activity. "Examinee A" did not understand the link labelled "Line Up" and so was unable to progress to the next step. As such we determined that this page was unsuitable to use for comparison in this experiment. The final times required for decision making were 46sec and 40sec respectively. The time between the first click and the final click was 19sec, while the color was changed 10times. "Examinee A" and "Examinee B" had the same results so we decided this page was suitable for examination and comparison.

The record lengths required for decision making (determination of the favorite color) were almost same. It took 19 seconds for each examinee from the first click to the last click to change colors. We judged that this page had an equal condition for "Examinee A" and "Examinee B". Then, we analyzed these 19 seconds in detail (see Figure 4).

On Figure 4 images The points show the eye tracking coordinates for each 0.02 seconds over a 19 second period. The lines show the mouse movements.

"Examinee A" 's eye movements concentrated on the texts (name of color) and the photographs of car alternately. "Examinee B" 's eye movements ran around the whole page broadly. "Examinee B" 's eye movements had no action of gazing at a click object. "Examinee B" 's eye movements were always quick and moved widely. The number of clicks for each examinee was 10 times. However, "Examinee A" clicked only the texts. And "Examinee B" clicked only the tiny pictures, oppositely. The link targets which "Examinee B" used were just 10\*10 pixel pictures but the eye tracking shows that "Examinee B" never looked at them either.

## 5. Discussions

"Examinee A" is a non-impaired male and "Examinee B" is a hearing-impaired woman. Age and the years of Internet experience (but total accessing time is unknown) are equal. At present, the factor which makes the difference in operation cannot be specified and we cannot yet prove that the disability is related. But the examinee, types differ clearly. "Examinee A" is the type who gets information from text, and "Examinee B" gets information from pictures. We studied to three hearing-impaired people, all of them had the tendency to take information from pictures. We can show and contrast the differences of operation tendencies concretely by video records.

It is known that the hearing-impaired do not use text information as effectively as non-impaired people. Why did the "Automaker" website have some differences? This home page has the same amount of both characters and pictures. However, as most of the pictures are image photographs of cars, the differences in the contents cannot be drawn from the photographs, and text information must be used. But, the Keyword "LINEUP" of this task was written in *katakana-English* and "Examinee B" could not understand the meaning (*Katakana-English is when English words are written using Japanese characters and sounds*). "Examinee B" was able to access no information from both pictures and texts. Therefore, it was a difficult interface for "Examinee B", who operates depending on picture information. We considered the difference between each examinee by studying the eye-movements. The characteristics of the person who operates depending on picture information (like "Examinee B") are to gather information from the whole page except text information and to judge something

in an instant. We suppose that as hearing impaired people gather more information visually than the non-impaired, non accurate visuals and graphics mislead the hearing impaired more. We think it is important and simple, to make internet visuals & graphics accurate so that everyone can access information efficiently.

## **6. Conclusions**

We found some differences and similarities, between non-impaired and hearing-impaired people in this case-study. The "Automaker" website had differences in reaction to text and picture information. We have considered that two user types (Text-oriented for non-impaired vs. picture-oriented for hearing-impaired) exist in vision information processing of WEB operation. The "Official Approval Guidelines" website and the "Major Cosmetics Maker" website were accessible for hearing-impaired people. The "Official Approval Guidelines" homepage and the "Automaker" homepage had same standards such as amount of information, balance of texts and pictures, navigation type, navigation layout, and page divisions. But the picture type of "Official Approval Guidelines" homepage was icon. On this homepage pictures and text were presented as a set, it cannot be classified as either a text based type or a picture based type. The "Major Cosmetics Maker" homepage was based on picture information, and photographs of goods.

The fact that there was no difference may indicate the same interface was provided to both types equally. Using suitable picture information will allow an easier interface for picture dependent operators such as "Examinee B".

Our research hopes to further W3C's guidelines to specify that the "non text equivalents" convey a significant meaning about the contents. We formed hypothetical Web Accessibility Guidelines for Hearing -Impaired People based on these case studies.

### **For Picture:**

- Pictures should express the contents accurately
- Use concrete objects
- Use photographs of goods and suitable icons

For Text:

- Text should be accompanied with pictures

For Japanese (also applied to elderly people):

- Avoid using English and katakana notation in menus

## 7. Acknowledgment

Some papers suggested that "new technology frontier" has also created enormous roadblocks and barriers for people with disabilities <sup>8)</sup>, but another researches try to develop new browser. The web accessibility is always discussed as one of technical problems. However, we believe that the web pages become more accessible for hearing-impaired people by visual design. This research is just preliminary and we plan to have more samples in our next experiment. In the future, we'd like to consider the correlation between user types and different factors such as sex and impairment. We will use a questionnaire to rate the examinee's interest in the test subject and ask how much they access to the internet every day (or a week). We assume that the operation characteristics are connected with the examinee's interesting and knowledge. Our next plan is to research the composition elements related to web accessibility, according to the result of an experiment in which we will investigate the features of operation and recognition in a typical web design model. We are going to continue such field-work.

## 8. References

- 1) <http://www.w3.org/TR/WCAG20/>
- 2) <http://www.w3.org/TR/WCAG10/#gl-provide-equivalents>
- 3) <http://www.w3.org/TR/WCAG10/#auditory-description>
- 4) <http://www.w3.org/TR/WCAG10/>
- 5) [http://usability.gov/web\\_508/tut-b.html](http://usability.gov/web_508/tut-b.html)
- 6) <http://www.nysashhh.org/index.htm>
- 7) ISCAN, Inc. : <http://iscaninc.com/HomePage.html>,  
<http://www.creact.co.jp/>
- 8) Lila F.Laux, Peter R. McNally, Michael G. Paciello, Gregg C. Vanderheiden, Designing the World Wide Web for People with

Disabilities: A User Centered Design Approach, ASSETS '96,  
Vancouver, British Columbia, Canada, pp.94-98

## **9. Appendix**

\*Daniel M. Berry, Requirements for Maintaining Web Access for Hearing-Impaired Individuals, IEEE 3rd International Workshop on Web Site Evolution(WSE'01), Italy, Nov.10-Dec 10, 2001, pp.33-41.

\*Namatame, M. Kobayashi, M. Harada, A. A Study of Web Accessibility for the Hearing Impaired. Universal Access in HCI: inclusive Design in the Information Society. Volume 4 of the Proceedings of HCI International 2003,p.1010-1014

\*This research's project number of Grant-In-Aid for science research is 13680262 and 15500646.

\*Printing of pictures had obtained permission of BMW Japan co. and BMW Tokyo co.

<http://www.bmw.co.jp/Contents/> ,

<http://www.bmw.co.jp/Product/Automobiles/>

(We accessed BMW-Site at 20th Feb 2002. BMW-Homepage was changed after our experiments).

# Virtual Realisation: Supporting Creative Outcomes in Medicine and Music

John A. Waterworth<sup>^</sup>

Interactive Institute  
Tools for Creativity Studio  
Umeå, Sweden

---

## ABSTRACT

In this article, I describe findings on the impact of virtual realisation on professional skills and creativity, based on observational studies and interviews with surgeons and musicians. I also present a vision of computer-supported creativity in terms of a modular set of virtual and augmented-reality environments based around an explicit model of the creative process. I suggest that by combining these two types of study, the potential of virtual realisation technology can be utilised in a way that transcends physical distinctions of time and place, while reinforcing the cognitive distinctions that are essential for the generation of creative outcomes.

---

Keywords: *virtual reality, augmented reality, virtual realisation, surgery, music* .

Received 27 October 2003; accepted 20 December 2003.

## 1. Introduction: VR and Professional Work

Investment advisors, surgeons, engineers, composers of music, industrial designers, meteorologists, visual artists, geologists, cinematographers, crystallographers, molecular biologists, and many others, are using computer graphics, and sometimes sounds and also tactile displays, to experience information in ways that would have been impossible, or required tremendous imaginative efforts, to realise previously. In this way, new information technologies such as virtual reality and multimedia radically change the nature of professional work and the skills required of professionals in carrying out that work. The adoption of such technology has an enormous impact on several aspects of each profession in which it is utilised; on education and training, on

---

<sup>^</sup> Corresponding Author:  
John A. Waterworth,  
E-mail: jwworth@informatik.umu.se

self-perception, on creativity and the generation of new ideas and approaches to problems. This brings fundamental changes to what it means to practise a skilled profession, with consequent changes in the public and self-perception of the professions, in what being a professional is thought to be.

At the Tools for Creativity studio, part of the Interactive Institute (of Sweden), we are currently investigating the impact of proliferating information technology (IT) on professionals such as surgeons, musicians and educators. Two examples will be described in some detail. The first is the use of a dextrous virtual reality environment to plan neurosurgery (see Serra et al., 1999; Waterworth, 2000), and the changes the introduction of this technology has brought to surgeons' ways of working, and what is possible through that work. A short account of the separation of Siamese twins with conjoined brains is used as an illustration. The second example is of a 'synaesthetic medium' (Waterworth, 1997a) used in piano performance education. This permits a performance to be experienced in a new, technologically-enhanced way, compared with other performances of the same piece, and so opens up a multi-modal channel of communication between a teacher and one or more students, or between students. Both of these technologies can be and are used between remote locations, including the home, via the Internet. Before describing the cases, it is worth clarifying what I mean by the often abused word: creativity.

## **2. What is Creativity and How Can It Be Supported with Information Technology?**

What do I mean by creativity? The term is often used very loosely. Does it differ from originality, innovation, improvisation, generating novelty, or problem solving? And if so, how? I will not try to define all these terms but rather suggest that we can see them all, and others, as referring to ways of achieving novel outcomes or novel ways of achieving desirable outcomes. In the application of professional skills (and probably generally) creativity is best thought of in terms of outcomes. As suggested by Janlert (2002), "the proof of creativity is in the result. Creativity is not identified as a special creative kind of feeling or mood of the person being creative. It is not identified as some spark of inspiration, god-sent or produced by some special creative conditions ..... Creativity is identified as a certain quality of a result produced, relative to the circumstances of the production and the perspective taken". Creativity, as used here, implies novelty in solving a real problem encountered by an individual in a certain situation.

Boden (1995) makes the point that human (and machine) creativity arises from the paradoxical blending of freedom and the application of constraints. Freedom to generate alternative possibilities, unhindered by practical considerations, must be combined with later reflection on how well the possibilities match practical constraints. The stages must be kept separate. Boden's view echoes evolutionary ideas in biology, such as Monod's (1971) characterisation of evolution as the result of chance alternating with necessity in a given environment. More recently, evolutionary ideas on creativity have been extended into the realm of culture and the creation of novel artefacts. Boden (2000) also differentiates between two different kinds of creativity, P-creativity (P for psychological) and H-creativity (H for historical). P-creativity is an idea that is novel to the mind of the individual only while H-creativity is when the idea is novel in the whole of human history (as far as anybody knows). This indicates that P-creativity is more fundamental than H-creativity, and that H-creativity is a special case of P-creativity.

West (2000) provides a useful summary of what is known about creativity and innovation at work. In line with the definition of creativity above, he sees the generation of a new idea as a cognitive process located within an individual. In general, people are at their most creative when alone, free from pressure, and feeling safe and positive. On the other hand, innovation often involves groups, and implementation is best done in teams. Innovation and its implementation often arise as a result of pressure or even threats from outside the group. We can see from this that very different conditions are needed to promote creativity versus the innovative implementation of new ideas.

Individual creativity is inhibited when others question a new idea too closely or too soon, when they make fun of an idea, or when they ignore a new idea. An individual raising a new idea is socially vulnerable and can be easily deterred from following through to the innovation stage. Implementation of innovative ideas, on the other hand, generally requires diversity of knowledge and professional background. But there should also be a high level of social integration within the group. An innovative group that does not inhibit individual creativity will be one where individuals feel safe, with little competition between members.

Dawkins (1976) first proposed the idea of memes, the cultural equivalent of genes, and others have developed the idea further (e.g. Blackmore, 1999; Dennett, 1995). Meme theory applies the genetic operations of copying, duplicating, resequencing, and mutating to ideas, tunes, objects, and other artefactual elements of differing complexity. Evolutionary creativity depends on variation, selection and inheritance (retention).

Variation maps onto non-judgmental openness, while selection corresponds to the application of constraints. Retention is what remains after unsuccessful candidates have been eliminated (in standard evolutionary terms, what is retained are genetic features of organisms that survive to have offspring with the same features).

In line with the idea that freedom from normal constraints is important to creativity, Arthur Koestler (1964, reprinted 1990) points to the importance of the suspension of rational thought, and the role of the unconscious in being creative:

"The moment of truth, the sudden emergence of a new insight, is an act of intuition. Such intuitions give the appearance of miraculous flashes, or short-circuits of reasoning. In fact they may be likened to an immersed chain, or which only the beginning and end are visible above the surface of consciousness. The diver vanishes at one end of the chain and comes up at the other end, guided by invisible links."

Much earlier, Poincaré (1905) had suggested that novel combinations in the unconscious have a role in the creative process. He distinguished four distinct phases, which are generally described as: Preparation, Incubation, Illumination, and Verification (to which a fifth Revision stage is often added). These are described in more detail in Section 5. Shneiderman (2000) describes work to provide "user interfaces for supporting innovation", based on a four part process consisting of Collecting (learning from published work in libraries and on the Web), Relating (consulting with peers and mentors), Creating (exploring, composing and evaluating possible solutions), and Donating (disseminating results). This is a very general framework for all kinds of "creative work", but the focus is on rather mundane activities such as searching for relevant information, composing documents and other intellectual products, consulting with others, and visualising information. There is also a useful review of a few "creativity-enhancing" software tools, such as the Axon Idea Processor for producing "Mind Maps". But existing software packages most often provide only very limited support for restricted parts of the creative process, or are concerned with general production tasks such as preparing a book manuscript for publication. Other recent authors have suggested that too much rational thought can stifle creativity and even intelligence (Caxton, 1997). De Bono (e.g. 1990) is famous for the idea of "lateral thinking" amongst other techniques that help suspend rational judgments. Part of the idea is to leave time and mental space for novel insights to develop. Another strand in

understanding how creative sparks arise is the catalogue of evidence suggesting that this depends more on sensory perception than on conceptual cognition. While conceptual analysis is a necessary part of developing new ideas, original insights may depend more on sensory exploration than on conceptual analysis.

Waterworth (1997a) pointed to the way in which new technologies, by allowing information to be experienced in a variety of vivid media and forms, expand the possibilities for creative inspiration. In other words, experiencing the same underlying information in different representations, and through a variety of sensory channels increases the range of concrete perceptions through which information is experienced. These richer perceptions may then lead to more original concepts. Some people have the natural ability to experience information through several senses at the same time, a phenomenon known as synaesthesia. Synaesthetic individuals appear to be particularly creative (Cytowik, 1989, 1995). Systems that provide experiences of the same material in different modalities have been dubbed “synaesthetic media” (Waterworth, 1997a), and an example is described in Section 4. The next section presents an unusually unequivocal example of how technology can enhance creativity, using virtual realisation to open up new possibilities in difficult surgical interventions.

### **3. Separating Conjoined Brains**

Surgical operations are rarely as difficult as the separation of two Siamese twins, joined at the head and with inter-mingled brain tissues. Conventionally, a surgeon must interpret large numbers of 2D ‘slices’ of the brain. These slices are radiographs and are of different types – coming from different scanners – according to the type of tissue they are sensitive to. A surgeon then had to form a detailed 3D image in mind – a mental realisation of the anatomical structure – combining large numbers of 2D scans from different scanners, some showing blood vessels, others showing bone structure, and so on.

In contrast, virtual realisation of medical data allows a radiologist or surgeon to interact with an integrated three-dimensional model that combines data from several different scanners. The model can be rotated, enlarged, viewed from any angle, and selected portions can be rendered completely or partially transparent. Possible surgical interventions can be explored using “virtual surgical instruments” such as a drill, a scalpel, a marker, and a ruler (which can be used to measure *inside* the dimensions inside an anatomical structure). Specialised software tools can be used to help the

surgeon identify and highlight particular structures. What difference does using virtual realisation make to planning and a complex surgical intervention? Benjamin Carson, MD, a neurosurgeon at Johns Hopkins Hospital, Baltimore, USA, puts it like this: 'It was fantastic. When I put those glasses on and went in to the virtual work station it was like I had them [the twins] right there in front of me.'

In his book *The Big Picture* (Carson, 1999) he writes: *'I can say it was the next best thing to brain surgery - at least in terms of my preparation and planning for the scheduled operation on the Banda twins. In a Johns Hopkins research lab in Baltimore, Maryland, I could don a special set of 3-D glasses and stare into a small, reflective screen which then projected an image into space so that I could virtually "see" inside the heads of two little Siamese twins who were actually lying in a hospital on another continent [Carson was in the USA, the twins were in Zambia].*

*Using simple hand controls I manipulated a series of virtual tools. A turning fork or spoke could actually move the image in space - rotating the interwoven brains of these two boys to observe them from any and all angles. I could magnify the image in order to examine the smallest details, erase outer segments of the brain to see what lay hidden underneath, and even slice through the brain to see what different cross-section would reveal about the inner structure of the brains. This allowed me to isolate even the smallest of blood vessels and follow them along their interior or exterior surface without difficulty or danger of damaging the surrounding tissue. All of which, of course, would be impossible in an actual operating room.'*

Carson continues: *'The chief benefit of all this was knowledge. I could observe and study the inner structure of the twins' brains before we opened them up and began the actual procedure of the operating table. I could note abnormalities ahead of time and spot potential danger areas - which promised to reduce the number of surprises we would encounter in the real operation.'*

The most difficult part of the actual operation was sorting out and separating the *'overlapping, interconnected, and shared blood vessels. It required a tedious and meticulous separation and closing off, cutting through and sometimes reconstructing a massive tangle of veins. Each vein had to be isolated and taken down in the right order - as carefully as if you were defusing a bomb. Make even a small mistake and the resulting blood loss could flood the surgical field and make it impossible to locate and control the problem in time to prevent serious brain damage or death. Being able to see, study, and memorize the vascular anatomy - particularly in the crucial structure around the brainstem, the ventricular system, and the skull base - proved an incredible*

*advantage. I've since tried to explain the benefit this way: a normal human brain is perhaps the single most marvelous and complex piece in the great three-dimensional master puzzle of creation. When you press two such complex organs together, the problems of orientation are compounded almost beyond belief. Finding our way around the abnormal venous structures of craniopagus [i.e. conjoined by fused skulls] twins is like being a cab driver dropped into the middle of a foreign city you have never seen to before, where you do not speak the language and cannot even read the road signs - and still being expected to do the job. This time I at least had a detailed road map to study before I got there. In fact, I felt almost as if I had successfully performed the operation.'*

Looking back on the operation, Dr Carson commented in an interview that *'There came a point when looking at their brains it appeared that there was no plane and that we wouldn't be able to get them apart, except that I knew from having done the virtual surgery before, that there was a plane and I remembered where the plane was so I began to tease in that direction and eventually the plane showed itself.'* Dr Carson was able to perform the separation successfully, and both the young twins went on to make a full recovery. This would almost certainly have been impossible without the use of virtual realisation.

The combination of detailed and integrated visualisation with interactive exploration of data lies behind this dramatic increase in the chances of successful outcomes, reflecting the increased creativity of surgeons using virtual realisation tools. Visual information has always been important to medical practitioners and in medical training, using visual aids such as charts, models and visual inspection of patients and cadavers. Of these, only cadavers allow explorative interaction but suffer from the limitations of any physical representation, as well as imperfectly reflecting the reality of a healthy individual. With this kind of tool ( see Figure 1) the surgeon has the possibility to choose between different views of the data on a spectrum from specific details to views that contain all data that has been collected (an holistic view).



**Figure 1** – The Dextroscope™ system used by Carson ©Volume Interactions 2003

In brain surgery, the scope of possible visualisations of accurate patient-specific information has, until the advent of VR techniques in the last decade, been limited to two-dimensional images from various types of scanner. Major decisions about a surgical intervention had to be made on the basis of this information, with no opportunity for direct visual inspection of the patient's brain. A surgeon would mentally realise the three-dimensional structure involved and, on the basis of this, drill down through the skull and into the brain to the target area. Virtual realisation means that a surgeon can inspect and interact with a three-dimensional rendering of the patient's brain, formed from a combination of data from the various scanners available. Each scanner reflects different types of material – typically, blood vessels, bone, and soft tissue – which can be visually represented as in nature, or differently to bring out a particular density of material, as required.

At this stage in the state of the art, there is a strong emphasis on visualisation. But this is coupled with manual interaction with the displayed data. A growing trend is to incorporate other modalities, such as touch and hearing, in the virtual realisation. Using force feedback devices allows the surgeon to feel the resistance of represented

materials, by pushing a probe into the virtual realisation. Different sounds are also used to reflect tissue densities, and while this is much less natural than the use of vision or force feedback, people seem to be readily able to interpret and use this switched-modality information. In fact, modality-shifting is one of the great strengths of virtual realisation, since it enhances the likelihood of novel insights and creative outcomes.

#### **4. Playing piano with feeling**

As with the practice of brain surgery, the important part of learning to play the piano is about performing physical actions in a highly-skilled way. Going beyond reading the notes to making quality music involves the acquisition of an intimate complex of physical and mental skills, and how a good teacher imparts those skills to a good student remains an art form. Virtual realisation can, however, facilitate communication between teacher and student during this learning process. It can also serve as a “synaesthetic medium” (Waterworth, 1997a) by which listening skills are enhanced through the use of dynamic graphical displays. Beyond establishing whether or not a performer gets the notes right, it is possible to use MIDI technology (Loy, 1989) to record several aspects of a performance, including dynamics, tempo, articulation, and synchronization. If these records are then displayed in a suitable representation, it will be easier for a piano teacher to communicate to a student the distinction between the art of playing piano and the technique of playing the correct notes.

What happens when an advanced piano student comes to his teacher? Usually, the student has been asked to prepare performances of one or more compositions. It is already taken for granted that the student is familiar with interpreting the notation of these compositions, so what happens next? Assuming that the student has been practicing these works, the lesson will most likely begin with his playing one of them, probably in its entirety, for the teacher. The teacher listens to this performance and takes note of specific aspects that need to be pointed out to the student. Some of these may involve mistakes in “decoding” the notation; but, as the student becomes more experienced in notation literacy, the teacher becomes more concerned with the subtleties of interpretation—the transition from just reading the notes and playing them according to the score to making good music. In other words, how can the student move on to become a creative musical performer? A related question is how can the teacher support the student's learning process in the best way, and in which form should this support be expressed?

This transition can rarely be expressed easily in words. Communication depends more on the ability of the teacher to demonstrate to the student how this subtle quality of interpretation involves more than playing the notes exactly as they are written. Such demonstration requires cultivation of the student's ability to perceive that such a difference does exist, and it also involves the teacher's ability to perceive what the student is or is not doing by way of appropriate interpretation. This is a very subtle process which has been examined considerably by Donald Schön (e.g. 1987) as a problem of educating talent in designing. Schön analysed several case studies of master classes in musical performance, which he interprets as dealing with "designing performance." One of the lessons to be drawn from the studies reported by Schön is that the role of the teacher is just as important (if not more so) as that of the "information being acquired" by the student. This is because one of the most critical difficulties is whether or not the student has really grasped what the teacher is trying to say (and in whether or not the teacher can establish whether or not the student has grasped it). This cannot be established through words. Just because a student knows how to say the right thing does not mean he will actually do it when he brings his fingers to the keyboard. This is not a problem of "knowledge acquisition" which would fit comfortably into the paradigm of expert systems (Clancey, 1989). Rather, it is a problem that depends critically on the ability of the teacher to exercise judgment; and, in music education, that judgment must be grounded in the skill of listening. The teacher can only assess the success of her communication with the student when she hears how the student is performing. However, if that student has not been successful, what and how should the teacher be communicating to him? As with exploring the structure of a brain, these are ultimately perceptual problems. The teacher hears certain features of the performance that indicate that the student is not playing as the teacher intended. The student must learn not only to play like his teacher but also to listen that way - to perceive the significant distinctions on which a skilled performance is based. Several piano teachers were interviewed about the observations they tend to make during a student performance, and the skills they are trying to impart. The assumption was that these observations would reflect what teachers wanted the student to actually hear. What we discovered was that the question of what students should be listening for could be restructured as a collection of related questions:

**i) How are the dynamics being interpreted?**

This is one of the best examples of how interpretation must be more than simply "decoding" the notation. In the work of many composers, such as Wolfgang Amadeus Mozart, dynamic markings are few, if they are present at all, while they are far more abundant in compositions by, for example, Franz Liszt. However, even when they figure heavily in the notation, a precise rendering of all dynamic markings is not necessarily a satisfying interpretation; or, put another way, dynamic markings at best indicate how a performance could sound, which need not also be a specification for how it should be executed.

**(ii) What are the onset times of the notes?**

Absolutely precise interpretation is rarely desirable in performance. We all know the experience of tapping the foot to a "beat" defined by these onset times; but we are probably not aware of subtle variations in the rate at which the foot is tapped. Indeed, variation from a uniform tempo is as important as variation from uniform dynamics, often for the same reason—as an indication of "inflection" which sets off phrasing (Desain and Honing, 1992).

**(iii) How are the notes articulated?**

Articulation is basically the percentage of the time until the next note onset during which the current note sounds. Notes which are notated as staccato tend to sound for only about 50% of the time between successive onsets, while the sound of a slurred note tends to endure right up to, or even beyond, the onset of its successor. Articulation tends to define the character of a phrase, just as dynamics and duration contribute to its shape; so observation of a student's articulation technique is very important. The detailed ways in which these questions are answered determines whether the performance is experienced as lifeless, mechanical, boring or moving, expressive, creative.

*pianoFORTE* (Smoliar et al., 1995) is a virtual realisation system which addresses these three questions by performing computations on MIDI representations of performances. Answers to each of the questions are then communicated through graphic displays. None of these answers implies correctness: the system produces a visual description, not an assessment (see Figure 2). A faithful reproduction of the score that the student is playing provides the infrastructure for all displays. A session

will typically begin by recording a student performance; but, because all performances may be easily stored as computer files, it may also begin with the teacher retrieving either a past student performance or a demonstration performance by one of possibly several teachers. The recorded performance allows the teacher to demonstrate specific elements of interpretative technique to the student, and the teacher interacts with the computerised score to select those particular portions of the recording which best illustrate the points to be made. Computations on performance data may be applied to any of these recordings, so the teacher communicates with the student by supplementing the act of listening to selected passages with that of viewing graphic annotations of the score describing how the music was performed, whilst also listening to the performance.



**Figure 2** – Using colour to convey performance dynamics in pianoFORTE

This new approach to communicating about piano performances demands a new kind of skill from the teachers, who are generally not used to the idea that viewing a performance can be a valuable supplement to listening to it. The very perception of music changes through the use of these displays; ultimately, their primary function is to

cultivate listening skills. Indeed, we have observed that even expert teachers have been able to hear things they had not previously noticed in a performance after examining the displays. Nevertheless, the displays should not be seen as a replacement for listening but as a tool for increasing awareness of what is being heard. With sufficient exposure and practice, the attentive student can eventually learn to hear those features which are first presented to him visually; so, even without the aid of computer displays, he will then be better equipped to discuss problems of performance with his teacher. User trials have revealed significant enthusiasm from the students and a ready acquisition of the necessary skills from the teachers. Students also acquire the skill of reading these displays on their own, as the graphics become a vehicle for communication with the teacher. There is thus the possibility that *pianoFORTE* can also serve as a tool for self-study. However, there is a danger of attaching too much importance to this role, since *pianoFORTE* has been designed to display the what of a performance, rather than the why. *pianoFORTE* does not detract from the role of the teacher but is designed to facilitate that role. While many treatises have been written about the art of playing the piano (Lhevine, 1972, being one of the leading classics in the literature), the nature of piano pedagogy remains highly intuitive. Ultimately the problem is one of communicating "knowing how" (as opposed to "knowing that") knowledge (Ryle, 1984), while much of our theoretical knowledge of education tends to concern the communication of "knowing that" knowledge. *pianoFORTE* uses graphic displays as a means of facilitating the communication of "knowing how," knowledge, taking advantage of the fact that a MIDI interface can be used to provide data concerning both what and when specific actions are taken. Because these actions are limited to the scope of moving keys and pedals, one may easily represent the "how" of piano performances (by both students and teachers) in terms of these data. It is the representation of performance in such terms which serves as a basis for teacher-student communication. Activities based on the recall of fleeting perceptions are realized as concrete data structures that may be analyzed and examined in detail. However, it is also important to observe that this new approach to communication will demand a new kind of skill from the teachers, who are not currently used to the idea that, for purposes of a critique, "viewing" a performance can be a valuable supplement to listening to it. It is not unusual for VR to lead to an increased emphasis on the visual. But what is more unusual, and perhaps important here, is the shift of modality (or, more accurately, the addition of an extra modality). The very perception of music may change through the use of these displays, and this seems likely to be an increasingly

frequent effect of virtual realisation. Any digital medium can be transformed into another – from vision to sound, from sound to vision, from vision to touch, from touch to sound, and so on. But, ultimately, the primary function of *pianoFORTE* is to cultivate listening skills. With sufficient exposure and practice, the attentive student should eventually learn to *hear* those features which are first presented *visually*; so, even without the aid of computer displays, that student will then be better equipped to discuss problems of performance with any teacher, even one who has not had experience with the *pianoFORTE* system. One of the biggest dangers of computer music as a technology is that it detracts from the behavioural aspects of music as an art form. Music is neither the notes on a printed page nor the motor skills required for the proper technical execution of those notes. Rather, it is a far more elaborate complex of behaviour in which the making of sounds is tightly coupled with their perception (Smoliar, 1994). Whether one is composing music, performing it, or just improvising, listening is still the paramount skill. The ultimate goal of this example of virtual realisation is to make us all better listeners. This skill is equally important whether we make our music at a piano keyboard or at a computer workstation.

## **5. Virtual Realisation: replacing cognitive space**

The two examples of virtual realisation presented above, clearly represent cases where information technology can enhance quality of outcomes, and perhaps creativity. But a major problem with many current uses of information technology is quite the opposite - that they reduce the 'cognitive space' in which people feel they can be creative. A large part of the reason for this is the progressive loss of the distinction between work and elsewhere – especially the home. When we are always contactable by e-mail, mobile phones, and SMS, it seems likely that our creative capacities will be compromised. This is because time spent out of the workplace – and, traditionally, away from the technology and communications found in the workplace – is actually vital to creativity. Home workers may organise things to have periods of inaccessibility at home, times when they do not look at e-mail or accept incoming phone calls. This is almost never the case in a professional work environment.

Technologies of virtual realisation, such as those mentioned above, also have the capacity to reinstate necessary distinctions as a by-product of the characteristic way in which they support the 'concretisation' of information. They can provide a range of types of virtual environment that support not only communication, conceptual work, and

rich sensory experiences, but also play, exploration, introspection, relaxation and even meditation. Our current approach is based on a three-dimensional model of experience, comprised of *locus*, *focus* and *sensus* (Waterworth and Waterworth, 2001a).

*Locus* captures the extent to which the observer is focused on the real world or a virtual model. *Focus* describes the nature of the user's attention, specifically whether they are attending to currently present stimuli (from the real or virtual world) – in which case they will experience a feeling of *presence* – or are attending to information which is not currently present in the real or virtual environment. The former can be characterised as perceptual (or concrete) processing, the latter as conceptual (or abstract) processing. I refer to this latter, reflective state of mind as *absence*. Finally, the *sensus* dimension refers to the level of attentional arousal of the observer, and ranges from awake and alert to totally unconscious. Refer to Waterworth and Waterworth (2000, 2001a) for more details of the model and its application. The three dimensions of the model describe a design space for what I call “Creative Spaces” – virtual realisation environments specifically developed to support creativity.

The *Creative Spaces* we are developing correspond to one or more of the stages of a hypothesised cyclical process, which carries the individual through changes along the three dimensions of experience outlined above. The suggested process consists of five stages: sensory stimulation, inactivity, conceptual exploration, recollection/reflection and selection. In more detail, the process is as follows:

1. Sensory stimulation (varying modal form for the same content). Here the idea of modality shifting in so-called *synaesthetic media* (Waterworth, 1997a) is applicable – emphasising perception rather than conception, thus inducing a strong sense of presence and emotional engagement. The two examples given earlier fall into this category

2. Inactivity (or HCI; but in this case, it is human-computer *inactivity*). Here the participant is encouraged, with technological assistance, *not* to focus consciously on the issues or problems under consideration. This can be done in various ways, but all involve the inducement of a calm, free-floating sense of low-level awareness. Seclusion in a personal creative space, safely open to emotions, will often support this phase (see Waterworth and Waterworth, 2001b).

3. Conceptual exploration (ideas in action, manipulating memetic elements). Here the strong emphasis on pure perception of 1., and the relative unconsciousness of 2., give way to more conceptual activities, such as the synthesis and evaluation of new combinations of memetic elements. The participant alternates between mild presence and mild absence as he or she explores information models in a detached way (see Waterworth, 1997b).

4. Recollection and Reflection (assessment of results against goals). Conceptual faculties are now fully engaged in a more analytic stage of the process. The participant is highly conscious but no longer feels present in the current (real and/or virtual) environment. Rather, he or she is fully engaged with concepts generated during the creative process. However, earlier phases may need to be re-experienced, through the replaying of earlier situations. This will tend temporarily to recreate the conditions of those particular phases.

5. Selection, and deletion of unwanted materials. In this stage the participant is engaged in critiquing and editing the results of earlier stages, and so feels strong absence, but is reconnected to the real world through reflecting on the feasibility of the ideas generated earlier in the process. The result of this phase may be to re-enter the process at an earlier phase.

The stages of the process match standard views of creativity (which are mostly derived from Poincaré, 1905) quite closely, but each component utilises computer-based informational techniques such as virtual reality (VR), augmented reality (AR), multimodal and cross-modal information presentation, and memetic manipulation (by computer and/or by the participant).

The transition from stage 1 to stage 5, as can be seen from the descriptions above, is a move from passive, conscious being, in a state of mental presence, to more active, conscious doing, in a state of relative mental absence (see Waterworth and Waterworth, 2000 for more details). Although the process is described in linear form, it is actually cyclical. The final stage will often lead to a realisation of the need to re-enter the process to develop insights further or just to start all over again. It will also often be

the case that participants feel the need to enter or leave the process at a particular stage. This flexibility is facilitated by a modular approach: people should use the components they feel they need in light of their creative needs.

The potential to use technology to help develop professional skills, and to carry out professional work in creative ways, has never been greater. Although some existing skills (such as mental realisation) will be diminished, greater possibilities are opened up. But technological trends towards mobility and networked information systems mean that this trend is no longer limited to the traditional workplace. More and more, professionals carry out parts of their work from home, and this is part of a larger tendency for the distinction between work and leisure, workplace and home, to diminish. However, what is known about how creative ideas arise indicates a strong need for distinctions of location and mindset of the kind that are now disappearing. This is the paradox of new technology: it provides tools that can, if used appropriately, enhance creativity in the exercise of professional skills; but it also often removes distinctions between environments and ways of working that are themselves necessary for creative work. To solve problems creatively people need, first, a problem they are intensely interested in solving. They need to attack the problem with all the technical resources at their disposal. But if the needed solution is truly novel, the approaches they have used before will be unlikely to succeed. This is when they need to step back, forget the constraints imposed by previous ways of working, and allow the cognitive space needed for creativity to open up.

## **6. Acknowledgements**

Thanks are due to Dr Luis Serra, of Volume Interactions, Singapore, and to Dr Stephen Smoliar of FX Palo Alto Laboratory, USA, for collaboration on the medical and musical applications, respectively; and to Dr Eva Lindh Waterworth and the editors for helpful suggestions for improving the text.

## **7. References**

Blackmore, S (1999). *The Meme Machine*. Oxford, UK: Oxford University Press.

- Boden, M A (1995). 'Creativity and Unpredictability'. In *Constructions of the Mind: Artificial Intelligence and the Humanities*. Stanford Electronic Humanities Review, 4
- Boden M A (2000). 'State of the art: Computer models of creativity', *The Psychologist*, 13, No 2, February 2000.
- Carson, B (1999). *The Big Picture*. Grand Rapids, Michigan, USA: Zondevan Publishing House.
- Caxton, G L (1997). *Hare brain, tortoise mind: Why intelligence increases when you think less*. London: Fourth Estate.
- Clancey, W J (1989). 'The knowledge level reinterpreted: Modeling socio-technical systems', *International Journal of Intelligent Systems*, vol. 8, no. 1, pp. 33–49, 1989.
- Cytowic R E (1989). *Synesthesia: A Union of the Senses*. New York: Springer-Verlag.
- Cytowic R E (1995). 'Synesthesia: Phenomenology And Neuropsychology - a review of current knowledge', *Psyche* (electronic journal) 2
- Dawkins, R (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- De Bono, E (1990). *Lateral Thinking: A Textbook of Creativity*. Harmondsworth, UK: Penguin.
- Dennett, D (1995). *Darwin's Dangerous Idea*. Harmondsworth, UK: Penguin.
- Desain, P and Honing, H (1992). 'Tempo curves considered harmful', in *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence* (P. Desain and H. Honing, eds.), pp. 25–40, Amsterdam, HOLLAND: Thesis Publishers, 1992.
- Janlert, L-E. (2002). 'Tools for Creativity: breakers and bracers'. *Internal Report*. Department of Computing Science, Umeå University, Sweden.
- Koestler, A (1990). *The Act of Creation*. Reprint edition. Arkana.
- Lhevinne, J (1972). *Basic Principles in Pianoforte Playing*. New York, NY: Dover, 1972. Foreword by R. Lhevinne.