Construction of the Biocybernetic Loop: A Case Study

Stephen H Fairclough
School of Natural Sciences and Psychology
Liverpool John Moores University
Liverpool, UK
+ 44 151 9046311
s.fairclough@ljmu.ac.uk

Kiel Gilleade
School of Natural Sciences and Psychology
Liverpool John Moores University
Liverpool, UK
+44 151 9046314
k.m.gilleade@ljmu.ac.uk

ABSTRACT

The biocybernetic loop describes the data processing protocol at the heart of all physiological computing systems. The loop also encompasses the goals of the system design with respect to the anticipated impact of the adaptation on user behaviour. There are numerous challenges facing the designer of a biocybernetic loop in terms of measurement, data processing and adaptive design. These challenges are multidisciplinary in nature spanning psychology and computer science. This paper is concerned with the design process of the biocybernetic loop. A number of criteria for an effective loop are described followed by a six-stage design cycle. The challenges faced by the designer at each stage of the design process are exemplified with reference to a case study where EEG data were used to adapt a computer game.

Categories and Subject Descriptors

H5.2 [Information Systems]: User Interfaces; K.8.0 [General]: Games; J.3 [Life and Medical Sciences]

General Terms

Measurement, Performance, Design, Experimentation, Human Factors

Keywords

Psychophysiology, physiological computing, adaptation, games.

1. INTRODUCTION

Physiological computing systems receive data from the body, which is subsequently used to adapt the human-computer interface [1, 2]. This category of technology falls into a number of distinct groupings including:- brain-computer interfaces (BCI) [3] that use intentional changes in physiological activity as a form of input control system, and biocybernetic adaptation [4] where spontaneous changes in psychophysiology trigger adaptation at the human-computer interface. All physiological computing systems are constructed upon a biocybernetic control loop where raw physiological data is analysed in near-real time and transformed into input for software control.

This paper is concerned with the design cycle involved in the development of a working biocybernetic control loop. An early loop was developed at NASA where changes in electrocortical activity (EEG) quantified the level of task engagement exhibited by the operator [4, 5]. When the operator was engaged with a simulated aviation task, automation of joystick control was permitted, therefore reducing the mental workload of an alert person. If the level of task engagement fell, automation was terminated, effectively forcing the operator to control the joystick manually and re-engage with the task. This system was designed to manage the psychological state of the operator within acceptable limits. A similar type of biocybernetic control loop that responds to mental workload has been proposed by [6] and [7].

The biocybernetic loop is the elemental concept at the heart of all physiological computing systems. At a basic level, the loop describes the data processing protocol whereby live physiology is converted into control input for a technological system. However, the design of the loop also incorporates an explicit rationale with specific goals, e.g. to sustain a state of engagement, to prevent frustration, to select a desired command; this agenda defines the *modus operandi* of the system.

The loop is initiated by the collection of data from various psychophysiological sensors. Traditionally these sensors have been physically connected to the user, but recent developments in wearable computing [8] and remote sensing [9] demonstrates the availability of unobtrusive monitoring. Physiological data are streamed from sensors to be filtered or corrected for artifacts prior to detailed analysis. These data can be subjected to a range of analysis techniques from simple averaging to more complex analyses, such as Fast Fourier Transform. The output from analysis is used to trigger an adaptive response from the system, either by creating specific criteria or by categorizing data using pattern classification approaches, e.g. [10]; see [11] for review. The response triggered at the interface may fall into a number of categories; it could enable cursor control (move forward, left, right) or trigger specific commands ('select') in the case of BCI applications. Adaptation may trigger a discrete event, e.g. provision of help information, reduction of music volume, or make a covert adjustment to the interface that may not be noticed by the user, e.g. lighting changes to a virtual environment; see [2] for discussion of overt and covert categories of system adaptation.

The biocybernetic loop is designed to promote specific psychological states in the human operator. These states often represent a positive "approach" goal, which is seen to be desirable, such as the promotion of relaxation or productive engagement with activity. Other loops, such as the original version designed at NASA, are negative control loops designed to avoid undesirable operator states, such as disengagement from the task or an extremely high level of mental workload that would jeopardize performance quality. It is important to recognize that the biocybernetic loop must be imbued with a degree of autonomy in order to influence the psychological state of the user in a prescribed fashion.

This paper is concerned with the biocybernetic loop as the essential element at the heart of all physiological computing systems. We wish to consider the challenge of constructing a biocybernetic loop that is both scientifically valid and effective from the perspective of user experience. In order to elaborate on this challenge, we present a case study from our own work.

2. CONSTRUCTION OF A WORKING BIOCYBERNETIC LOOP

The construction of a biocybernetic control loop must be transparent at every stage of development. The rationale and data processing protocol must be open to inspection in order to aid troubleshooting and build a knowledge base for future development. This transparency extends to the rationale and scientific validation of the loop. The goals of the loop, in terms of the intended effects of interaction on the human, should be clearly defined. Furthermore the fundamental process of psychophysiological measurement/inference must have a documented scientific basis.

The biocybernetic loop is a multidisciplinary construction that must satisfy several criteria if it is to work in a satisfactory fashion. If the loop fails to achieve its intended effects on user experience, there are many possible causes from inappropriate selection of measures or insensitive categorisation of data to poor HCI design at the interface. This is why transparency in the design cycle is so important. In the broader context, increased clarity of design will consolidate shared knowledge and understanding as biocybernetic loops are constructed across a wide range of subject domains.

The real challenge of designing a biocybernetic loop is to satisfy a range of criteria that span psychology, signal processing and computer science. In order to examine this issue in greater detail, we have proposed four criteria that must be satisfied in order to ensure that the loop is capable of impacting on user experience in a sensitive and reliable fashion.

1. Psychophysiological inference should be valid

   The complex nature of the psychophysiological inference (i.e. how psychological meaning is inferred from physiological data) has been described in detail [12]. This issue is highly relevant to the biocybernetic loop where physiological measures should correspond to specific psychological concepts across individuals and different environments. The reliability of psychophysiological measures over repeated sessions is a topic that has been relatively under-explored, however, see [13] and [14] for exceptions. It has been argued that biocybernetic loops must be based on valid psychophysiological inference [2], which may be described with reference to previous research. If the measure does not capture the psychological construct with sufficient sensitivity and reliability, the loop will not influence the user state in a predictable fashion.

2. The psychophysiological measures can deliver a sufficient representation of the user state.

   A valid psychophysiological inference delivers a degree of association between measure and construct. These measures must be subsequently translated into a representation of the user state within the loop. For example, we may have evidence that reduced alpha activity in the frontal cortex is related to high mental workload, but this association must be translated into a categorical representation of the user state (e.g. high/low workload) in order for the system to work. This second criteria contains an implication that the representation of user state is sufficiently detailed – but how can we assess the representation in this respect? We argue that the sufficiency of the user representation relates directly to the adaptive repertoire or functional vocabulary of the system, i.e. the number of responses supported by the system.

3. The classification of user representation is accurate

   The representation of the user must be categorised in real-time or near-real time if the loop is to operate dynamically. This criterion is defined by the data processing protocol that converts dynamic and variable streams of physiological data into a representation of the user. The first point to consider is the error rate of the data processing protocol, i.e. how many instances of target states, such as high workload or frustration, are missed? How many instances of non-target states are misclassified? The second point of consideration is related to the time window associated with the data sampling/processing protocol and its relationship to system response at the interface. With respect to latter, the output from a biocybernetic loop may take the form of an adaptation or a form of input control. These outputs are associated with a time window; for input control, this time window may be very short whilst response times may vary considerably in the case of adaptation. Obviously the time window for data analysis and categorisation sets the minimum response time at the interface.

4. The response at the interface has the desired effect on the user

   The output from the biocybernetic loop will largely determine the user experience and the influence of the loop. If interaction at the interface is poorly designed, then the loop will be ineffective, even if the previous three criteria have been satisfied. From the perspective of design, it is important that the adaptive response at the interface has the desired effect on user state and no unintended effects. Therefore, if the loop is designed to enable cursor control, it should do exactly that; if it is designed to minimize frustration, a reduction of negative emotion should be a tangible benefit. The user will expect the system to respond to changes in psychological state in a timely and intuitive manner, i.e. system response should be sensitive and the intention underlying the response should be logical from the users' perspective.

The remainder of this paper will describe six stages in the construction of a working biocybernetic loop that was created in our own laboratory. This loop was designed to detect spontaneous changes in psychophysiology in order to adapt software in near-real time.

3. CASE STUDY

The case study for the construction of a biocybernetic loop is a physiological computing game [15-18]. This category of game is capable of responding to changes in the state of player and tailoring the gaming experience to the individual. The obvious purpose of the loop in this context is to create positive gaming experience, i.e. to minimize sustained periods of frustration and/or boredom and to promote engagement/effort.

In our case study, the loop was designed to manipulate the difficulty of the game in order to drive the experience of the player. We selected the Tetris game for the case study, our reasoning being that there was a relatively simple relationship between speed (of falling blocks) and game difficulty.

The development of the loop is described in six stages, from the conceptual origins to evaluation of a working system. Each stage presents specific challenges and an example is provided in our case study to present how we resolved those issues.

3.1 Stage One: Conceptual Model

The design of a biocybernetic loop generally rests on two kinds of conceptual model. The first is a psychological model of user behaviour, these models may describe the experience/expression of emotions or the relationship between frustration or fatigue with performance quality. These psychological models will often be derived from existing literature. The second model concerns the rationale behind the loop itself and how the designer anticipates the adaptive logic of the loop will influence the behaviour of the user. This first model is generic whilst the second represents a series of predictions for a specific usage scenario. With respect to the use of psychological models, it is recommended to adopt those models, which are (a) empirical, i.e. constructed on existing measures, and (b) predictive, i.e. associated with specific behavioural outcomes.

One challenge for the designer is how to reconcile the precise design of the biocybernetic loop as defined by the second model with general concepts from psychology. It is easy to describe those behavioural states that should be promoted or prevented by the loop in general terms. The rationale for the biocybernetic loop always seeks to enhance the experience of interaction through the promotion of positive, productive states and mitigation of negative states from boredom to frustration. Such labels serve adequately as a statement of general design goals, however, the biocybernetic loop requires a series of precise and objective formulations in order to achieve those goals.

The first step at this conceptual stage is to define specific categories of user behaviour to be detected; this process sets the context for the adaptive logic of the system. The designer may develop a scheme of desirable vs. undesirable user states; it is equally important to specify the adaptive repertoire of the system within the context of this scheme, i.e. how many states are to be detected and how many different responses must be supported by the loop. If possible, it is sensible to assess the response rate of adaptation to understand how quickly the system must react.

Example: Our basic model was to construct a version of Tetris that engaged the player by controlling the drop speed of pieces to sustain a positive gaming experience. The adaptive mechanic of the loop was straightforward but the definitions of desirable or undesirable states were vague and difficult to assess.

If a biocybernetic system is designed to promote an optimal gaming experience, how should we characterize that state in psychological terms? To be fully engaged with a task to the point that time seems to slip away is known as the *flow state* [19], which is sometimes referred to as being in the zone [20] or total immersion [21]. This state is a desirable experience as the level of challenge is matched with player skill, therefore attention is focused entirely on the game to the exclusion of other forms of mental activity. This type of experience not only provides an optimum level of challenge [22], it also avoids undesirable mental states (e.g. boredom) and promotes a positive emotional experience [23]. These kinds of psychological states represent a transaction between the skills of the player and the objective level of task demand. Representing these states using objective measures (game demand) is problematic because players' response to different Tetris speeds is an interaction between the level of skill and objective task characteristics.

From a psychological perspective, what are the characteristics of an optimal gaming experience? It is reasonable to assume that high levels of game demand would prompt increased mental effort into the game. A high level of effort is also necessary to stimulate skill development and there is evidence that the opportunity to acquire and demonstrate competence/mastery contributes to a positive gaming experience [24]. The relationship between effort investment and the task difficulty is described by motivational intensity model (MIM) [25]. MIM emphasizes a compensatory dynamic where effort is increased in response to rising demand; however, this relationship is nonmonotonic and includes a 'tipping point' where the investment of effort and inclination to continue may abruptly decline due to overload and an appraisal of unachievable demand. Both mental workload (the level of demand experienced by the player) and effort (motivation) are important drivers for optimal gaming experience. Therefore, we adapted the MIM to create four distinct player states: boredom (low effort due to low demand), engagement (increasing effort in line with increasing demand), zone (peak effort at highest level of demand attainable) and overload (low effort due to excessive demand). This scheme is illustrated in Figure 1.

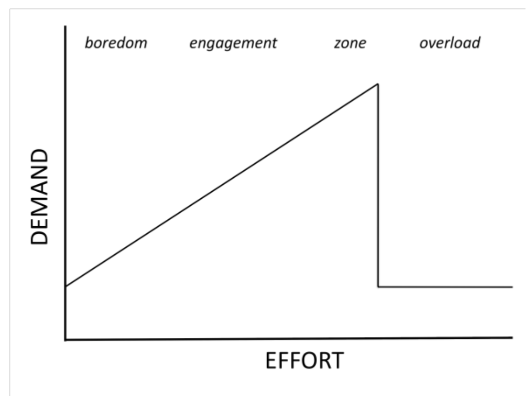

Figure 1. Motivational Intensity Model adapted from [25].

With this scheme in mind, the biocybernetic loop was designed to promote effortful engagement with the Tetris game via states of engagement/zone and to avoid the two states of boredom and overload. The adaptive logic of the loop would be to increase the speed of falling blocks when boredom was detected and to continue to increase speed until effort peaked at the zone. If a zone state was detected, the system was to make no further adjustment. If a combination of high demand and low effort was detected, the loop should reduce the speed of the falling block to take the player out of the overload state. This four-category model corresponds to the model of user representation that would be adopted by our biocybernetic loop.

3.2 Stage Two: Psycho-Physiological Inference

The conceptual model of user behaviour or representation provides an implicit direction for operationalisation. At this stage in the process, the designer must elect to choose a number of psychophysiological measures to represent the user state. The designer should carefully peruse the research literature because it is common for psychological concepts to be associated with more than one physiological measure. By definition, a good psychophysiological measure is sensitive and responds to changes in psychological states in an accurate and reliable fashion. It is also important for the psychophysiological measure to respond specifically to a target psychological state and not be subject to influences from other psychological states.

One important consideration when selecting a psychophysiological measure for the biocybernetic loop is the context used to assess the sensitivity of the measure. A measure may have a proven record of sensitivity in the laboratory but there is often a question about whether the same sensitivity will be apparent in the field. There is also the issue of the task context for sensitivity. Many studies in the psychophysiological literature used standardized stimuli and task protocols that may be very different from the task context for the biocybernetic loop. The designer must also consider several practical issues such as the degree of intrusiveness introduced by the sensor, i.e. a hand-worn sensor would be of little use for a task that required use of both hands.

The designer should at the very least examine the existing research literature from a critical perspective and make an intelligent estimate of whether this research will generalise to the usage scenario of the system. The best way to assess this issue is to run some systematic tests of the psychophysiological measures with the actual task scenario. This kind of testing serves to ensure that sensitivity of the measures is appropriate for the loop, provided that tests are conducted with a reasonable number of participants.

Example: The measurement of effort in the MIM is associated with beta-adrenergic activity in the cardiovascular system, using measures of systolic blood pressure reactivity and the pre-ejection period [26, 27]. These cardiovascular measures are sensitive to movement and were not practical for our purposes. We had already performed an EEG study in our laboratory where increased cognitive demand resulted in enhanced activity in theta band (4-7Hz) power over mid-frontal scalp [28-30]. In addition, the manipulation of cognitive demand leads to a suppression of absolute power in both lower (~7.5-10Hz) and upper (~10.5-13Hz) alpha bands [28]. However, this EEG data was derived from a standard experimental psychology protocol (a working memory task).

We performed a study where 20 participants played Tetris under three conditions: easy/boredom (drop speed of blocks was very low), hard/engagement/flow (drop speed of blocks was fast) and impossible/overload (drop speed of blocks was so fast that playing the game was impossible). The analysis of frontal theta activity at 6Hz revealed an significant effect for task demand [$F(2,18) = 21.89$, $p<.01$]. The observed trend illustrated that frontal theta increased from easy to hard demand and declined when task demand was extremely difficult (see Figure 2). The analysis of upper alpha activity at 11.5Hz revealed a linear trend with increased task demand [$F(1,19)=13.69$, $p<.01$], i.e. greater suppression of alpha activity as game demand increased. This linear effect was localized to the parietal region, e.g. P4.

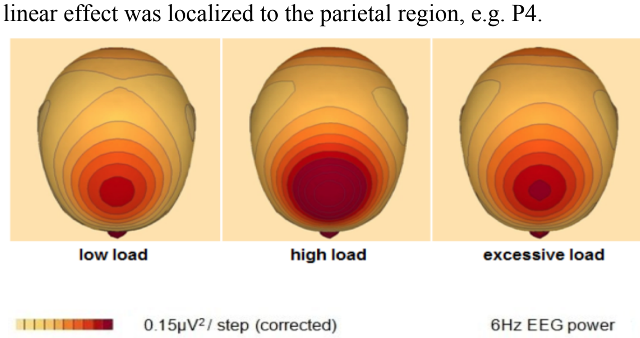

Figure 2. Frontal theta at low, high and excessive demand on Tetris (N=20).

At the end of the experiment, we had two candidate measures; frontal theta to represent effort invested in the game and parietal alpha to capture the level of mental workload/demand experienced by the player. Furthermore, both measures had been validated in the specific usage context of the biocybernetic loop.

3.3 Stage 3: A Quantified Model of User State

Once measures have been validated, the next stage of the process is to recreate the conceptual model of the user state illustrated in Figure 1 in a quantifiable form. The challenge at this stage is to operationalise the conceptual model of the user in an accurate and detailed form. Due to restrictions surrounding the act of measurement (e.g. availability of appropriate sensors, strength of the psycho-physiological inference), there may be a loss of fidelity during the transition from concept to measurement.

This quantifiable model is the way in which the user state or behaviour is represented within the biocybernetic loop. The minimum requirement of the quantified model is that the user is represented with sufficient fidelity to enable the process of adaptation within the loop. In other words, the target states that drive the adaptive logic of the loop should be clearly defined.

Example: The conceptual model of the user was provided by MIM at stage one. We adapted this model by measuring effort/motivation using frontal theta at Fz (Figure 2) whereas the demand dimension was represented by alpha suppression at parietal sites. We made this assumption based upon our experiment as frontal theta demonstrated the shark-fin curve associated with MIM (Figure 1) whilst parietal alpha had a linear relationship with game demand. The combination of both measures, combined with MIM, created the two-dimensional model of user state illustrated in Figure 4 (note: suppressed alpha power = increased demand). The desirable target states of ‘zone’ and ‘engagement’ are associated with high effort; undesirable states are defined by high demand in combination with low effort (overload) or low demand/low effort (boredom).

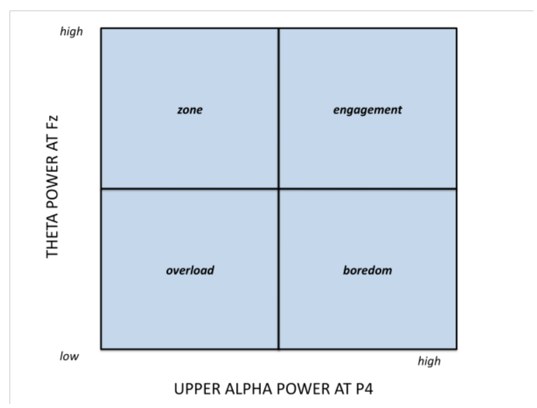

Figure 4. Two dimensional representation of the user state using EEG measures.

3.4 Stage 4: A Real-Time Model of User State

The quantified model of the user state is a generic formulation that defines the ‘data space’ of the monitoring activity of the loop. In order for the model to trigger adaptation in real-time, this generic model must be populated with classes and criteria that define the four regions of the user representation (Figure 4).

There are many techniques available for the classification of psychophysiological data within the biocybernetic loop, see [11] for review. Many of these machine learning algorithms require training data, i.e. exemplars of psychophysiological data that describe target states. The designer may rely on stage two in order to generate this kind of example data, but often triggers or criteria require a degree of calibration to the individual user. The system may require each user to undergo a calibration process prior to usage. Classification algorithms often require bespoke categories of data to enable training, but there are several approaches to the generation of these training sets. If the biocybernetic loop is designed to detect naturalistic states, such as sleepiness, then it is relatively straightforward to generate example data. e.g. sleep deprivation. Similarly, the generation of data to describe extreme states of high mental workload may be achieved by simply exposing individuals to high cognitive demand. There are a number of emotion induction protocols available that may be used to generate example data for different emotional states. These protocols can involve active manipulation (e.g. asking participants to recall emotional events from the past or placing people in a situation designed to induce a specific emotional state) or passive approaches (e.g. exposure to pictures/movies and/or music designed to induce a particular emotion). Alternatively, psychophysiology may be continuously recorded and participants are required to provide a subjective self-assessment in order to categorise emotional states.

The major obstacle for the generation of example data for the biocybernetic loop is concerned with the context of data collection. Therefore, the real-time model of user state should be calibrated using example data that is fully representative of the context of system usage. The problem with formal protocols is that data obtained may not generalize when the system is being used in a different context. On the other hand, a reliance on subjective self-assessment is complicated by other reasons, e.g. limitations of introspective techniques, retrospective bias, individual differences. However, given that the biocybernetic loop must respond in accordance with subjective perceptions, it is clear that calibration in line with self-reported experiences is an important approach.

Once example data has been translated into a categorization algorithm, the designer must consider the real-time limitations of the model. The response window of the adaptive system has been specified during stage one and the capability of the real-time model to classify data within that response window should be considered.

Example: We developed a number of calibration protocols to deliver trigger values that were tailored to each individual user. Our experimental work had created a version of Tetris where drop speed could be manipulated. We initially created a version with ten different drop speeds from slow to impossibly fast. Prior to system use, we exposed participants to each of ten levels and asked them to indicate a preference for each, i.e. which speed do you enjoy playing the most. This was a standard procedure wherein each person played at each drop speed for two minutes and was asked to deliver a subjective preference. It was relatively straightforward to characterise extreme states of boredom and overload using this manipulation. However this method had two problems: the trigger level depended on the specific characteristics of the connection and two minutes of playing at a constant level was unrepresentative of the game. Baseline levels of frontal theta and parietal alpha were taken to derive a second method of trigger definition. Baseline levels of EEG were collected whilst the participants viewed a relaxing piece of video footage. This second method was based on criteria for triggering adaptation, which were generated relative to baseline levels of EEG activity. For example, if frontal theta activity increased or decreased by a 100% (twice baseline) in any 5 sec window, whilst parietal alpha increased or decreased by 100% then the system adaptation may have been triggered. In our case, if frontal theta increased by 100% whilst parietal alpha fell by 100%, this defined the player as being in the zone (see Figure 4).

We also had to define our criteria within the context of a time window. This represented another design decision. If the time window for the trigger was too short, it was easily influenced by sudden fluctuations in EEG power that were unrepresentative. However, opting for a long time window (>5sec) ran the risk of reducing the sensitivity of the adaptive loop.

3.5 Stage 5: Design of the Adaptive Interface

The biocybernetic loop 'communicates' the results of monitoring/classification to the user via the interface. This may be an overt response that is obvious to the user or the system may adapt in a covert way that may or may not be noticed [2]. The critical concern for the designer at this stage is to ensure that adaptation is perceived as accurate, timely, intuitive and does not have any unintended consequences. User perceptions of interface adaptations are hugely influenced by the visibility of the adaptation. Overt adaptations tend to get noticed and accuracy is assessed in terms of self-appraisal (to what extent does the assessment of my internal state match my actual experience?), timeliness (is this adaptation what I need right now?) and intuition (does this adaptation make sense?). Overt adaptations can have an immediate impact on the state of the user and offer an effective response. However, they do run a risk of drawing attention to system error, which may reduce the trust of the user in the system. Covert adaptations are much less risky in this respect, being inherently ambiguous and involving subtle changes at the interface. The drawback of covert adaptation is that it tends to have a cumulative impact and is less effective than an overt strategy over a short period of time.

The designer may opt to use both overt and covert strategies within the same system. The logical strategy is to use covert adaptation to subtly reach target states and opt for overt forms to manage extreme states.

Example: The adaptive strategy of our Tetris game was relatively simple. The speed of the falling blocks could be increased or slowed. The relationship between the detection of user states and the type of adaptation was also straightforward; speed would be increased during boredom, decreased during overload and sustained during zone and engagement (Figure 4).

Designing the adaptation was principally an issue of magnitude (how much should speed of falling blocks change at each increment) and feedback (should the user be provided with objective feedback of difficulty/drop speed). After a series of pilot tests, we opted for very small adjustments in drop speed and no feedback to the player. This strategy corresponds to the covert designs that were described earlier. The experience of playing the game was that adjustments to drop speed were very subtle and the effect was cumulative and only apparent after playing for several minutes. This strategy was adopted in order to focus players' attentions on the game rather than the activity of the biocybernetic loop.

3.6 Stage 6: Evaluation

The benefits of biocybernetic adaptation for the user may be direct and tangible or nebulous. Like all categories of emerging technology, it is important for the biocybernetic loop to demonstrate some kind of benefit for the user experience, e.g. better performance, greater engagement, reduced frustration.

The final stage of the design process represents the challenge of evaluating and demonstrating the benefits of the biocybernetic loop. There are at least two levels of evaluation for a physiological computing system. The first is designed to answer the issue of relative benefit; in other words, if a biocybernetic system is better, then we can only make this assessment in relative terms, i.e. better compared to what? In the case of adaptive systems, the logical 'control' case is a manual version of the same system, but this kind of comparison tends to be asymmetrical because: (1) manual control systems offer gross control compared to automatic systems, and (2) there is a large novelty effect associated with nascent technology. The issue of novelty may be resolved by longitudinal testing but the nature of the comparison is an intractable issue. The second level of evaluation is to compare different versions of the biocybernetic control loop, i.e. different criteria or control dynamics, random adaptation [31]. This second type of evaluation offers a controlled comparison in a way that the first type does not, but it is only useful for differentiating the impact of different kinds of biocybernetic loop on user perceptions and behaviour.

Once the comparative frame has been established, the designer must decide how to measure outcomes from the evaluation. With respect to comparing the performance of the adaptive system with a 'control' case, there are several options available, such as: performance quality (productivity, efficiency, error rate), psychophysiology (use of measures to index psychological concepts, provided they are subject to direct influence of the loop) and measures related to the properties of the interaction (questionnaires, interviews). The primary choice of outcome measure will be determined by the conceptual basis of the loop as defined at stage one. If the biocybernetic loop is designed to reduce mental workload or reduce errors, these implicit goals set the context for system evaluation. For example, evaluation work conducted on the original biocybernetic loop [31-33] considered a range of outcome measures, including performance quality, subjective mental workload and EEG responses; see [34] for a summary.

When the scope of evaluation includes different kinds of working biocybernetic loops, the designer may also wish to compare the adaptive response of the loop, e.g. frequency of adaptive responses, relative count of different types of adaptive response.

Example: We decided to evaluate the basic prototype with respect to two main questions: (1) does biocybernetic adaptation promote improved player experience compared to manual adjustment of game demand, and (2) how does the reactivity of the biocybernetic loop (i.e. frequent vs. infrequent adjustments) impact on system performance and player experience. The first hypothesis is designed to contrast a covert, automated process of adjustment with a condition where adjustments are overt and at the discretion of the player. The second hypothesis is concerned with the design of trigger events (that activate adjustments) and how psychophysiological criteria impact on the system adaptation and the player experience.

For the current study, we contrasted three types of working biocybernetic loop: (a) a conservative system that only made an adjustment when changes in EEG substantially deviated from baseline (greater than 200%), (b) a liberal system that adjusted game demand in response to a small deviation from baseline levels of EEG (100%), and (c) a "normal" system that responded to moderate changes in EEG (150%) relative to both (a) and (b). It was anticipated that the conservative system would be the least reactive version of the loop, i.e. pushing up demand very gradually and responded to overload very slowly. The liberal system should make frequent adjustments and be faster to respond to the overload scenario. The 'normal' version was included for purposes of comparison with the liberal and conservative systems. For the manual control system, we adopted a 'Wizard of Oz' approach where participants were required to speak aloud an instruction to increase ("higher") or decrease ("lower") the speed of the falling blocks. These adjustments were made in real-time by an experimenter sitting behind a screen in the laboratory.

10 participants (6 females) volunteered for the evaluation session. This was a repeated measures design where each participant encountered four versions of the system (conservative/liberal/normal/manual). Each individual game of Tetris was played for 5 minutes. The order of presentation of each system was counterbalanced and participants had a 5-minute rest break between each game. Participants always started each game on the slowest speed setting. If the blocks reached the top of the board and "game death" occurred, the game always restarted with an empty board on the slowest speed setting.

In order to evaluate player experience, two sets of measures were collected. A mood adjective checklist (UMACL) [35] was administered before and after each game session. This questionnaire assessed three components of mood: energetical arousal (EA: tired-alert), tense arousal (TA: relaxed-tense) and hedonic tone (HT: happy-sad). Participants also completed an immersion questionnaire [21] designed to capture the quality of gaming experience.

The results of the experiment are divided into two areas, behaviour of the system and the influence of system behaviour on the user experience. We captured three aspects of system behaviour: (1) mean frequency of adjustments to task difficulty (increases and decreases of task difficulty), (2) mean frequency of "game deaths" or "resets" (i.e. when blocks reached the top of the board and the game had to be reset), and (3) average difficulty of each game (i.e. game difficulty could vary between 1-10 in accordance with speed of falling blocks). These data were calculated for all three versions of the biocybernetic system and the manual system. Mean values are provided in Table 1 below.

System	Increase Demand	Decrease Demand	Mean Reset	Mean Difficulty
Conservative	63.6	43.2	1.3	3.8
Normal	41.7	56.6	0.6	2.4
Liberal	28.6	62.4	0.4	1.9
Manual	9.4	1.7	0.5	3.3

Table 1. Mean values for measures of system adaptation across all four systems (N=10).

The measures were subjected to a MANOVA analysis (System Type x System Measure) to assess statistical significance. A significant interaction was found between both factors that was subjected to post-hoc t-tests. The number of adjustments to increase task demand was significantly higher for the conservative system compared to the liberal system ($p<.01$); unsurprisingly, all three biocybernetic systems exhibited a higher rate of upward adjustment compared to the manual system ($p<.01$). The latter trend was also apparent in the analysis of downward adjustment ($p<.05$). The analysis of mean difficulty level revealed that difficulty was significantly lower for the liberal system compared to all other systems ($p<.05$).

The impact of system adaptation on the user experience was assessed using two types of subjective questionnaire. The mood questionnaire (UMACL) was administered before and after each game session; this allowed us to calculate a change score (post-game minus pre-game) to assess mood changes in three component of mood: EA (alert-tired), TA (tense-relax) and HT (happy-sad). All three components were subjected to a 2 x 3 MANOVA. Mean values are presented in Table 2.

System	EA	TA	HT	IMM
Conservative	4.3	3.1	0.0	64.7
Normal	2.0	2.4	-1.9	65.5
Liberal	0.2	1.1	-2.0	66.1
Manual	1.1	0.7	-1.6	73.9

Table 2. Mean values for subjective data: EA = energetic arousal, TA = tense arousal, HT = hedonic tone, IMM = immersion (N=10).

Post-hoc t-tests revealed a significant effect for EA only ($p<.05$) – participants found the experience of playing the conservative version of the biocybernetic game to be more alerting compared to the liberal version. The analysis of responses to the immersion questionnaire was insignificant, but a trend was observed that participants found the manual version of the game to be the most immersive.

The goal of the evaluation was to assess how reactivity of the biocybernetic system influenced system performance and user experience. The conservative EEG criteria tended to "push" the player, making the least number of downward adjustments of difficulty and prompting a relatively high frequency of "game death" (Table 1). By contrast, the liberal system required only a small deviation from baseline levels of EEG activity in order to initiate an intervention and made a high number of downward adjustments of difficulty compared to the conservative system. As a result, the liberal system has the lowest level of mean task difficulty out of all four systems. The "normal" system represented a middle path between both extremes of EEG criteria. As expected, the manual system had a much lower frequency of adjustments as participants tended to increase difficulty to a desirable level and make no subsequent adjustments. In terms of average difficulty (Table 1), it was noted that the preferred level of demand was rather low (i.e. maximum difficulty = 10) and fell slightly short of the average demand imposed by the conservative system.

Despite differences in system behaviour, we found little evidence for any substantial effects on the user experience. The "push" provided by the conservative system resulted in enhanced alertness but there were no other significant effects on mood. We were surprised that experience of the game led to an increase of negative affect as this was completely unanticipated. We did speculate that the automated adjustment provided by the biocybernetic loop may have enhanced immersion in the game, compared to the manual version where the person had to 'break' from the experience to indicate an adjustment, but no evidence was found for this effect; in fact, immersion was maximized in the manual condition, presumably because this condition provided an optimal level of challenge.

In hindsight, we felt that the range of EEG criteria adopted to create three versions of the biocybernetic loop was too narrow to create significantly different kinds of user experience. Also, we adopted a 2sec time window for data processing and adaptive responses for all biocybernetic loops. The combination of a 2sec response and small, incremental adjustments to drop speed meant that the system was often unable to prevent "game death" when the game board was filled above two thirds of its height (and drop speed was high). We also felt that players needed longer exposure to the system, preferably in the form of repeated sessions, in order to differentiate the different types of adaptive control

4. SUMMARY

The paper has described the design process of the biocybernetic loop with a case study example. Each stage in the design process contains specific categories of challenge that span psychology, signal processing and human-computer interaction. The most significant challenge for the designer is addressing the multidisciplinary nature of the biocybernetic loop in sufficient detail.

5. ACKNOWLEDGMENTS

This research was performed as part of the REFLECT project (http://reflect.pst.ifi.lmu.de/) and funded by the EU under the Future and Emerging Technologies (FET) Initiative.

6. REFERENCES

[1] Allanson, J. and Fairclough, S. H. 2004. A research agenda for physiological computing. *Interacting With Computers*, 16, 857-878.

[2] Fairclough, S. H. 2009. Fundamentals of Physiological Computing. *Interacting With Computers*, 21, 133-145.

[3] Allison, B. Z., Wolpaw, E. W. and Wolpaw, J. R. 2007. Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4, 463-474.

[4] Pope, A. T., Bogart, E. H. and Bartolome, D. S. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40, 187-195.

[5] Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J. and Pope, A. T. 2000. A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology*, 10, 4, 393-410.

[6] Wilson, G. F. and Russell, C. A. 2007. Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors*, 49, 6, 1005-1018.

[7] Solovey, E. T., Chauncey, K., Lalooses, F., Paraso, M., Weaver, M., Scheutz, M., Schermerhorn, P., Sassaroli, S., Fantini, A., Girouard, A. and Jacob, R. J. K. 2011. Sensing cognitive multitasking for a brain-based adaptive user interface. In *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'11* (Vancouver, 2011). ACM Press.

[8] Van Laerhoven, K. 2011. ISWC 2010: The latest in wearable computing research. *IEEE Pervasive Computing*, 10, 1 2011), 8-10.

[9] Poh, M.-Z., McDuff, D. and Picard, R. 2011. A medical mirror for non-contact health monitoring. In *Proceedings of the ACM SIGGRAPH Emerging Technologies* (Vancouver, 2011). ACM Press.

[10] Baldwin, C. L. and Penaranda, B. N. 2012. Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *NeuroImage*, 59, 1, 48-56.

[11] Novak, D., Mihelj, M. and Munih, M. 2012. A survey of

methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting With Computers*, 24, 153-172.
[12] Cacioppo, J. T. and Tassinary, L. G. 1990. Inferring psychological significance from physiological signals. *American Psychologist*, 45, 16-28.
[13] Christensen, J. C., Estepp, J. R., Wilson, G. F. and Russell, C. A. 2012. The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59, 57-63.
[14] Picard, R. W., Vyzas, E. and Healey, J. 2001. Towards machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 1175-1191.
[15] Fairclough, S. H. 2007. *Psychophysiological inference and physiological computer games*. Workshop on BCI and Gaming at Advances in Computer Entertainment (ACE), Salzburg.
[16] Fairclough, S. H. 2008. BCI and Physiological Computing: Similarities, Differences and Intuitive Control. In *Proceedings of the Workshop on BCI and Computer Games: CHI'08* (Florence, 2008).
[17] Gilleade, K. M., Dix, A. and Allanson, J. 2005. *Affective videogames and modes of affective gaming: assist me, challenge me, emote me*. Paper presented at DIGRA, 2005.
[18] Rani, P., Sarkar, N. and Liu, C. 2006. *Maintaining optimal challenge in computer games through real-time physiological feedback*. In Schmorrow, D. D. (Eds.) Task Specific Information Processing in Operational and Virtual Environments, Foundations of Augmented Cognition, Erlbaum, pp. 184-192.
[19] Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. Harper Perennial, New York.
[20] Chen, J. 2004. Flow in games (and everything else). *Communications of the ACM*, 50, 4 2007), 31-34.
[21] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T. and Walton, A. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66, 641-661.
[22] Gilleade, K. M. and Dix, A. *Using frustration in the design of adaptive videogame*. Paper presented at ACE, 2004. ACM Press, New York, 2004.
[23] Nacke, L. and Lindley, C. A. 2008. Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the Conference on Future Play: Research, Play, Share* (Toronto, 2008). ACM, New York.
[24] Ryan, R. M., Rigby, C. S. and Przybylski, A. 2006. The motivational pull of video games: a self-determination approach. *Motivation and Emotion*, 30, 347-363.
[25] Wright, R. A. 2008. Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Social and Personality Psychology Compass*, 2, 682-701.
[26] Richter, M., Friedrich, A. and Gendolla, G. H. 2008. Task difficulty effects on cardiac activity. *Psychophysiology*, 45, 869-875.
[27] Richter, M. and Gendolla, G. H. E. 2009. The heart contracts to reward: monetary incentives and preejection period. *Psychophysiology*, 46, 451-457.
[28] Klimesch, W. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29, 169-195.
[29] Gevins, A. and Smith, M. E. 2003. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomic Science*, 4, 113-121.
[30] Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R. and Rush, G. 1998. Monitoring working memory load during computer-based tasks with EEG pattern recognition models. *Human Factors*, 40, 79-91.
[31] Freeman, F. G., Mikulka, P. J., Scerbo, M. W., Prinzel, L. J. and Clouatre, K. 2000. Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking system. *Applied Psychophysiology and Biofeedback*, 25, 103-115.
[32] Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J. and Scott, L. A. 2006. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors*, 48, 693-709.
[33] Prinzel, L. J., Parasuraman, R., Freeman, F. G., Scerbo, M. W., Mikulka, P. J. and Pope, A. T. 2003. *Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centred adaptive automation design*. NASA/TP-2003-212442, NASA, 2003.
[34] Scerbo, M. W., Freeman, F. G. and Mikulka, P. J. 2003. A brain-based system for adaptive automation. *Theoretical Issues in Ergonomic Science*, 4, 1-2, 200-219.
[35] Matthews, G., Jones, D. M. and Chamberlain, A. G. 1990. Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81, 17-42.