Proceedings of the IROS 2013 Workshop
on Neuroscience and Robotics

Towards a robot-enabled, Neuroscience-guided healthy society

Editors

Emre Ugur, Erhan Oztop

Jun Morimoto, Shin Ishii

Tokyo Big Sight, Japan

November 3rd, 2013

Preface

We are experiencing fast paced developments in robotics and neural sciences. Robots are becoming more and more part of our daily lives; in the near future they will be with us as companions, caregivers, smart prosthetics, and nano-robots in our bodies. The progress in neural sciences accelerated by brain imaging, clever behavioral experimentation and technical advancements such as multi-electrode recordings, better analysis techniques and neuroinformatics tools. Compared to twenty years ago, now the existing neuroscientific data and knowledge are more easily accessible, and thus available for building robotic systems that can exhibit the robustness, adaptability and intelligence of humans. Reciprocally, significant developments in robotics and machine learning put robotics in the service of Neuroscience as experimental platforms or test-beds of brain models.

We are witnessing the growth of a solid interdisciplinary research frontier, which on one hand uses Neuroscience for better robotics and intelligent systems, and on the other hand, uses robotics to better understand human cognition and intelligence. With this workshop we aim to bring in the pioneers in this frontier for further fostering this interdisciplinary effort by facilitating the exchange of ideas among researchers from diverse fields. The ultimate goal is to disseminate the current state of the art and set the research targets that need to be reached to ensure a robot-enabled, Neuroscience-guided healthy society.

We aimed to bring together researchers from both the robotics and Neuroscience in order to explore how to maximize the progress at the multidisciplinary frontier of robotic-for Neuroscience and Neuroscience-for-robotics. Five regular papers were accepted as contributions to the workshop after peer-reviewing. We expect that the twelve invited talks from distinguished scientists on Neuroscience and Robotics, together with the contributed talks elucidate initial answers for the questions posed above, and emphasize the challenges ahead.

We thank all submitting authors for choosing this workshop to disseminate their work. We thank keynote speakers who considerably contributed to the quality and the impact of the workshop. Needless to say, the program committee members have a big role in making the workshop a success; we thank them for their fine reviewing efforts. Finally, we would like to thank the IROS 2013 Organization Committee for facilitating the workshop execution.

This workshop was partially supported by European Communitys Seventh Framework Programme FP7/2007-2013 (Specific Programme Coop-

eration, Theme 3, Information and Communication Technologies) under grant agreement no. 270273, Xperience; by European Communitys Seventh Framework Programme FP7/2007-2013 under the grant agreement no. 321700, Converge. It was also partially funded by a contract in H23 with the Ministry of Internal Affairs and Communications, Japan, entitled 'Novel and innovative R&D making use of brain structures'.

December 2013

Emre Ugur
Erhan Oztop
Jun Morimoto
Shin Ishii

Organizing committee

- Emre Ugur, Innsbruck University, Innsbruck, Austria.
- Erhan Oztop, Ozyegin University, Istanbul, Turkey.
- Jun Morimoto, ATR, Kyoto, Japan.
- Shin Ishii, Kyoto University, Kyoto, Japan.

Program committee

- Tosif Ahamed, OIST, Japan
- Jan Babic, JSI, Slovenia
- Giuseppe Lisi, NAIST, Japan
- Jun Morimoto, ATR, Japan
- Yukie Nagai, Osaka University, Japan
- Bojan Nemec, JSI, Slovenia
- Erhan Oztop, Ozyegin University, Turkey
- Takeshi Ogawa, ATR, Japan
- Ludovic Righetti, USC, USA
- Archana Singh, University of Tokyo, Japan
- Barkan Ugurlu, ATR, Japan

Contact

Emre Ugur, PhD
University of Innsbruck
Institute of Computer Science, IIS
Technikerstr. 21a
6020 Innsbruck, Austria
email: emre.ugur@uibk.ac.at
phone: +43 512 507 53330

Table of Contents

Regular Papers

Recurrent Slow Feature Analysis for Developing Object Permanence in Robots 1
Hande Celikkanat, Erol Sahin, and Sinan Kalkan

Humanoids Learning to Crawl based on Natural CPG-Actor-Critic and Motor Primitives 7
Cai Li, Robert Lowe, and Tom Ziemke

A Bio-inspired Modular System for Humanoid Posture Control 16
Vittorio Lippi, Thomas Mergner, and Georg Hettich

Zero-calibration BMIs for sequential tasks using error-related potentials 22
Jonathan Grizou, Inaki Iturrate, Luis Montesano, Manuel Lopes, and Pierre-Yves Oudeyer

Detection of event-less error related potentials 28
Jason Omedes, Inaki Iturrate, and Luis Montesano

Invited Talk Abstracts

Is artificial emotion really emotional? 34
Minoru Asada

Shared body for self and others in the brain 35
Akira Murata

Cognitive Interaction Technology for Helpful Robots 36
Helge Ritter

Human-derived sensor fusion principles used to control biped balancing of external disturbances in a humanoid robot 37
Thomas Mergner

Motor primitives and central pattern generators: from biology to robotics 39
Auke Ijspeert

Mind-Controlled Humanoid Robots and Physical Embodiment 40
Abderrahmane Kheddar

Robots under Neural Control: How to create a neuron-based learning and memory system for behaving machines? 41
Florentin Worgotter

Adaptive robot skill synthesis through human sensorimotor learning 42
Jan Babic

Brain-Machine-Interface Improves Recovery Time from Perturbation in Flight Attitude on a Novel Complex Piloting Task 43
Daniel Callan

Brain Exoskeleton-Robot Interface for Rehabilitation Assistance 45
Tomoyuki Noda

A Waypoint-based Framework and Data-driven Decoder for Brain-Machine Interface in Smart Home Environments 46
Motoaki Kawanabe

Brain and body machine interfaces for assistive robot technology 47
Joern Vogel

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Recurrent Slow Feature Analysis for Developing Object Permanence in Robots

Hande Çelikkanat
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: hande@ceng.metu.edu.tr

Erol Şahin
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: erol@ceng.metu.edu.tr

Sinan Kalkan
KOVAN Research Lab.
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
Email: skalkan@ceng.metu.edu.tr

***Abstract*—In this work, we propose a biologically inspired framework for developing object permanence in robots. In particular, we build upon a previous work on a slowness principle-based visual model (Wiskott and Sejnowski, 2002), which was shown to be adept at tracking salient changes in the environment, while seamlessly "understanding" external causes, and self-emerging structures that resemble the human visual system. We propose an extension to this architecture with a prefrontal cortex-inspired recurrent loop that enables a simple short term memory, allowing the previously reactive system to retain information through time. We argue that object permanence in humans develop in a similar manner, that is, on top a previously matured object concept. Furthermore, we show that the resulting system displays the very behaviors which are thought to be cornerstones of object permanence understanding in humans. Specifically, the system is able to retain knowledge of a hidden object's velocity, as well as identity, through (finite) occluded periods.**

I. Introduction

Humans are born into persistent worlds. Through years and countless interactions, we come to understand the world as a place that makes temporal and spatial sense. Objects do not appear out of nowhere, nor vanish into thin air, and as they move from point A to point B, they indeed have to exist for some time at every point in between. However, it is difficult to claim that we have so far built robots that truly make use of these basic axioms. Given that this understanding is a basis for us humans to act effectively in our persistent world, in this study we propose a model for building an understanding of object permanence in terms of a higher-order internal representation of the environment. Our ultimate goal is to build effective environment manipulation capabilities on top of this basis later on. But first, the robot needs to "understand" what it is to exist in a persistent world.

Against the complexity of a world abundant with continuously changing sensory signals, we take refuge in the "slowness principle" [1]: While the sensory signals are noisy and erratic, their underlying physical causes are relatively persistent in time. For instance, retinal signals can vary greatly from one moment to another due to lightning conditions, as well as saccadic movements of the eye, however the object which the eye sees is constant. Therefore we must be able to process these erratic sensory signals to extract meaningful high-level representations, which are characteristic of varying more "slowly", thereby containing more valuable "information", than the readily-available sensory signals.

In [1], Wiskott and Sejnowski propose Slow Feature Analysis (SFA). They show that sensory signals can be processed through successive steps of principal component analysis to extract optimally slow signals, which summarize the meaningful event in the scene. The solutions are guaranteed to be optimally slow within a predefined family of functions, while still conveying meaningful information. In forthcoming work, Wiskott et al. design a hierarchical visual architecture which can recognize objects through translational, orientational, and scaling transformations [2], distinguish known and novel objects, predict the type of the solutions if the transformation is known a priori [3], survive multiple co-occuring transformations, and even adapt themselves to behave like simple and complex visual neurons when trained with natural-life scenes [4]. As is, this architecture develops the object concept very plausibly. However, it is reactive in time, responding momentarily to inputs; and not being able to retain information through time, it cannot survive the object permanence problem. We propose a prefrontal cortex inspired extension to serve as a working memory.

The contributions of this paper are threefold: First, we apply SFA to real world images to demonstrate that the invariant object recognition capabilities can indeed survive real world data. (Note that with the exception of Zhang and Tao [5] and Berkes and Wiskott [4], SFA has not been used for real world images before. Furthermore, in these two studies, it has not been utilized for object recognition.) Second, we propose a quantitative method to estimate the sufficient number of slowly varying signals to represent a certain event. Finally, and most significantly, we propose an extension to develop an understanding of object permanence.

Our fundamental claim in building our extension on top of the SFA framework is that the object permanence can be regarded as a stage which develops on top of an already developed reactive object concept. In this sense, we claim that the SFA architecture fulfills the initially maturing object concept understanding, on top of which the object permanence understanding develops later in time on a par with the maturation of the prefrontal cortex. Last but not least, we show that the proposed framework demonstrates similar characteristics (and pitfalls) like an infant learning permanence of objects.

H. Celikkanat gratefully acknowledges support of TUBITAK 2211 program.

II. Related Work

A. Object Permanence

Piaget famously proposed that the cognitive functions of an infant progresses in developmental "stages" [6]. Within the first stage, he singles out the *object permanence* understanding as one of the cornerstones, at the end of which objects come to be identified as independent entities. Furthermore, he also claimed that object permanence similarly develops in substages. The object concept forms in the second substage (1–4 months), indicated by the infant starting to follow their movements. She reaches for partially hidden objects by 4–8 months, and for fully hidden objects 8–12 months. By this time, she makes the A-not-B error [1], which disappears by 12–18 months.

Once fully developed, Michotte identifies two indicators of the object permanence understanding [7]:

The Tunnel Effect is the infant's capability of judging when an object, having previously disappeared behind a screen, will reappear again. This indicates the ability to track the object's position even while it's not directly observable. It also depends on the length of the occluded period: Young infants (of 4 months) can track the objects behind sufficiently short screens (< 14.8 cm), but their performance degrades to chance level as the occluded time gets longer. Older infants (of 6 months) can handle longer periods.

The Screen Effect is the surprise of the infant when object A enters behind a screen, and reappears not as itself, but having transformed into object B, indicating she understands the integrity of the object's identity.

B. Neurological Bases of Object Permanence

The close relationship between the maturation of the prefrontal cortex, and the emergence of object permanence understanding, has attracted much attention from neuroscientists [8]–[12]. Diamond and Goldman-Rakic [8] are one of the first to the demonstrate the link between the maturation (or integrity) of the dorsolateral prefrontal cortex and successful performance at the A-not-B task. They conduct a longitudinal study of infants performing the A-not-B task, as well as of adult rhesus monkeys with bilateral prefrontal and parietal ablations. They note a significant performance increase between 7.5-9 and 12 months, since the delay necessary to elicit the A-not-B error increases from 2-5s to 10s. In addition, monkeys with bilateral ablations of DL-PFC perform at the level of 7.5-9-month-olds, while unoperated and parietally operated monkeys are as successful as 12-month-olds; showing the direct dependence of A-not-B task on DL-PFC maturation.

Imaruoka et al. [11] and Saiki [12] introduce a novel paradigm, called the multiple object permanence tracking task, in which objects are moving in a display. They are distinguishable by their features, such as color and shape. The participants are required to track the objects, while also maintaining their features mentally. In this dynamic environment, Saiki [12] shows that our ability to keep bindings of objects color, shape and spatiotemporal locations was significantly impaired when objects move. Even though the visual short-term memory is generally assumed to be capable of maintaining 3-5 feature bound object representations, when the objects are on the move, this ability regresses down to 1 or 2 objects. Employing the same paradigm in an fMRI experiment, Imaruoka et al. [11] demonstrates the activation of anterior prefrontal cortex.

One thing significant about the prefrontal cortex is that it is abundant with recurrent loops, both intrinsic [13], and through other brain areas [14]. The generally accepted hypothesis is that these recurrent loops are the key structure to keep track of time concept in sequential events [15].

C. Robotics Studies

Chen and Weng [16] propose a value-based behavior to develop a rudimentary object permanence. The system is hard-coded to (1) be "surprised" when events are incongruent with its predictions, and (2) gaze longer upon surprising events. After habituation, it gazes longer at events which violate object permanence principles. Roy et al. [17] propose a mental imagery system for the robot, with a global physical model of itself, the objects, and the human partner. The system has an object tracking module, which maintains invisible objects for some time, and dropping ones that are hidden for too long.

A highly relevant work is the MTRNN model by Yamashita and Tani [18]. MTRNN is composed of two groups of contextual neurons, one group with a slow learning timescale, and one with a fast learning timescale. The fast neurons adapt themselves to rapid changes in the environment, thus discovering motion primitives, while the slow neurons learn to discern the context, thereby learning the sequence of necessary primitives to perform a certain behaviour. The major downside is that the slow neurons must be set to a certain discriminative initial state, both to learn, and to reproduce a certain sequence. Therefore, even though it uses the same idea of separating fast and slow signals, MTRNN needs a level of supervision that is not available in the classic object permanence scenario.

D. Slow Feature Analysis

Wiskott and Sejnowski [1] take a novel approach to visual perception. Through a rigorous mathematical procedure called Slow Feature Analysis, they extract the slowest signals carrying most information about the scene. These signals have a total ordering, allowing the selection of the slowest and most informative ones. The resulting system turns out to be highly robust, with some extra features emerging as well. Many identical SFA modules can be stacked together hierarchically, enabling feasible processing and parallelization of high-dimensional images. The system develops invariant object recognition [2]. It can withstand (possibly multiple) transformations such as translation, rotation, and scaling, distinguishing known and novel objects, while also providing insight about the transformation. For instance, in case of multiple moving objects, the system not only distinguishes different objects, but it can also identify the position of any of them. It is not negatively affected by multiple co-occuring transformations (translation, rotation, and/or scaling), rendering it suitable for real-life scenarios, where transformations do not generally occur in isolation. It is also mathematically treatable [3], and it is possible to predict the exact shape of outputs that will result from each of these transformations.

An interesting feature is the biologically plausible prop-

[1] In the A-not-B task, an object is first hidden at a location A several times, until the infant learns to retrieve it successfully. Afterwards, it is visibly taken away from A and moved to a second hidden location B, however infants at this stage still try to retrieve it from the previously learnt location A.

erties that emerge. For instance, the nodes self-organize to behave like simple and complex cells of the visual cortex [4]. When trained with natural life-like scenes, they adapt to prefer Gabor-filter-like inputs, responding maximally to certain directions, and minimally to others. In addition, certain nodes self-specialize to display end-inhibition or side inhibition, again similarly to specialized V1 complex cells. These adaptations are purely due to the input characteristics: Since the visual sequences are natural, they bear spatial and temporal continuity, resulting in these preferences. In yet another study, Franzius et al. [19] show the emergence of hippocampal place cell, head direction cell, and spatial view cell-like formations, which have specialized in rodents to represent its spatial state. Again, they emerge completely due to the nature of the input.

SFA has also been successfully employed for practical purposes. Zhang and Tao [5] propose an SFA-based system to recognize human actions. They also introduce three variants: (1) supervised, (2) discriminative, and (3) spatial discriminative. Kompella et al. [20] devise incremental and online deduction of slow features, while retaining computational feasibility in the face of high-dimensional input. The original SFA approach requires a (costly) offline learning phase, therefore this is an important step for real-time applications.

To the best of our knowledge, there has been no a priori studies to enhance this system with recurrence, nor any attempts to carry a trace of activation through time. The previous studies of slow feature analysis are purely reactive in time.

III. Methods

A. Slow Feature Analysis

Wiskott and Sejnowski [1] formalize the following optimization problem: Given an I-dimensional input signal, $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_I(t)]^T$, the objective is to find a set of input-output functions, $\mathbf{g}(x)$, which will produce a J-dimensional output signal $\mathbf{y}(t) := \mathbf{g}(\mathbf{x}(t))$, whose components vary as slowly as possible, while still containing information. The objective is to minimize $\langle \dot{(y)}_j^2 \rangle$, $\forall j \in 1, ..., J$, with:

$$\begin{aligned} \langle y_j \rangle &= 0 \quad \text{(zero mean)}, & (1)\\ \langle y_j^2 \rangle &= 1 \quad \text{(unit variance)}, & (2)\\ \forall j' < j : \langle y_{j'} y_j \rangle &= 0 \quad \text{(decorrelation)}. & (3) \end{aligned}$$

The angular brackets indicate averaging over time.

The unit variance constraint avoids the trivial solution with zero information content. The decorrelation constraint ensures non-redundant signals. It also enforces a total order: The smaller the index j is, the more optimal is the solution y_j.

This optimization problem is difficult to solve, but it can be simplified by constraining the output functions to be linear combinations of a finite set of nonlinear functions, that is, $y_j(t) = g_j(\mathbf{x}(t)) := \mathbf{w}_j^T \mathbf{z}(t)$. The nonlinear functions $\mathbf{z}(t)$ can be obtained via applying a set of functions $\mathbf{h} = [h_1, ..., h_K]$ on the input signals, thus expanding them nonlinearly: $\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t))$. After this nonlinear expansion, the problem can be treated as linear in the expanded signal components $z_k(t)$, similar to using a kernel to linearize the classification problem.

Then the problem reduces to finding the weight vectors $\mathbf{w}_j = [w_{j1}, ..., w_{jK}]^T$ to minimize $\langle \dot{y}_j^2 \rangle = \mathbf{w}_j^T \langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle \mathbf{w}_j$.

Assuming that the functions h_k are chosen such that the expanded signal $\mathbf{z}(t)$ has zero mean and unit covariance matrix ($\langle \mathbf{z} = 0 \rangle$ and $\langle \mathbf{z}\mathbf{z}^T = I \rangle$), the constraints:

$$\begin{aligned} \langle y_j \rangle &= \mathbf{w}_j^T \langle \mathbf{z} \rangle = 0,\\ \langle y_j^2 \rangle &= \mathbf{w}_j^T \langle \mathbf{z}\mathbf{z}^T \rangle \mathbf{w}_j = \mathbf{w}_j^T \mathbf{w}_j = 1,\\ \forall j' < j : \langle y_{j'} y_j \rangle &= \mathbf{w}_{j'}^T \langle \mathbf{z}\mathbf{z}^T \rangle \mathbf{w}_j = \mathbf{w}_{j'}^T \mathbf{w}_j = 0, \end{aligned}$$

are fulfilled if and only if the weight vectors form an orthonormal set. Therefore the set of eigenvectors of $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle$ gives us the weight vectors that satisfy the constraints. From these eigenvectors, we choose the ones with the smallest eigenvalues as the weight vectors, $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle \mathbf{w}_j = \lambda_j \mathbf{w}_j$,with $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_J$, resulting in the input-output functions: $g_j(x) = \mathbf{w}_j^T \mathbf{h}(x)$.

In other words, to find the slowest signal, we use the eigenvector of the smallest eigenvalue, corresponding to the direction of the least variance in the time derivative of the input. For other signals, orthogonal directions can be used, given by eigenvectors of increasing eigenvalues. They are found by a principle component analysis on the matrix $\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \rangle$.

For nonlinear expansion, Wiskott et al. use the first and second-degree monomials of the input: $\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t)) = [x_1(t), \cdots, x_I(t), x_1(t)x_1(t), x_1(t)x_2(t), \cdots, x_I(t)x_I(t)]^T$. Higher order expansions are possible, but not necessary, since a hierarchical architecture results in increasing complexity in higher-levels, performing this expansion in every layer.

Notice that the outputs signal are computed instantaneously, i.e., they are not a result of simple temporal low-pass filtering. Hence, the optimization problem is being solved by instantaneously calculating a higher level representation.

B. Recurrent SFA

We use a hierarchical architecture composed of SFA nodes (Figure 1) [2]. The input images have a resolution of 65x65. The bottom layer reads from the input, and is formed of 15x15 SFA nodes, each with a 9x9 receptive field, among which 5 pixels overlap. The higher 3 levels have 7x7, 3x3, 1x1 nodes respectively, all but the last one with 3x3 receptive fields. This part of the architecture is proposed in previous studies [2], and called thereupon **Feed-forward SFA** for clarity.

We extend Feed-forward SFA with an extra 1x1 layer on top, which feeds its output at time t is back to itself at $t + \Delta t$. The new architecture is called **Recurrent SFA**. The input to the top (n^{th}) layer at time t, $\mathbf{x}^n(t)$, becomes:

$$\begin{aligned} \mathbf{x}^n(t) = [&y_1^{n-1}(t), y_2^{n-1}(t), ..., y_J^{n-1}(t),\\ &y_1^n(t - \Delta t), y_2^n(t - \Delta t), ..., y_J^n(t - \Delta t)]^T. \end{aligned}$$

where $\mathbf{y}^n$ is the output of the n^{th} layer.

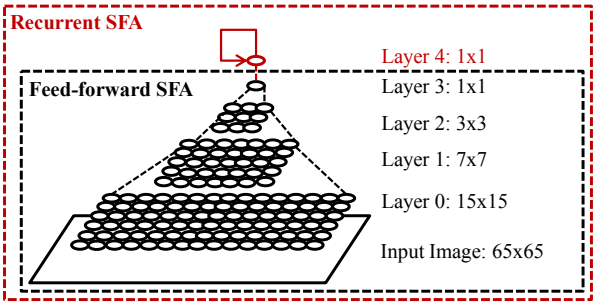


Fig. 1: The architecture of Feed-forward and Recurrent SFAs.

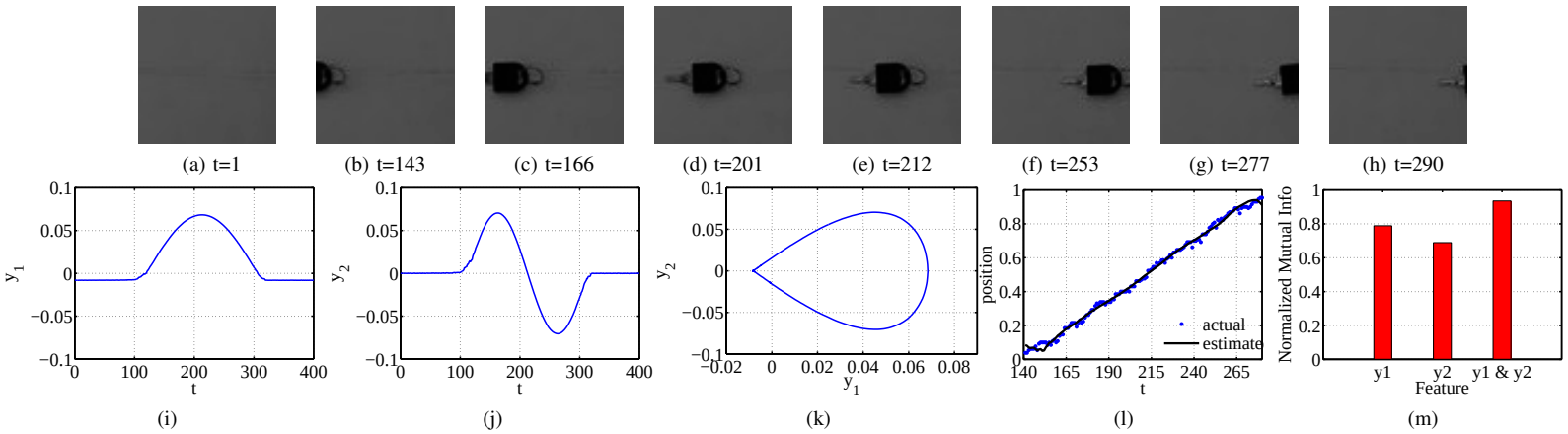


Fig. 2: The feed-forward SFA response to a single moving object. (a-h) Object enters into view from left at t=110, leaves from right at t=320. (i-j) The slowest two responses of the system, y_1 and y_2. (k) Phase diagram y_1 vs. y_2. Every point on the phase diagram corresponds to a unique location of the object on the retinal plane. (l) Actual versus estimated position values of the object. (n) Mutual information between the (actual) position values of the object, and y_1, y_2, and $y_1\&y_2$.

The time difference Δt, as the single parameter of the system, determines the maturity of the emulated prefrontal cortex, and hence called the "maturity" parameter. The smaller it is, the "younger" the system will be, and recall only the near past. As the system gets more mature, it can be increased to allow a longer window.

Note that a hierarchical structure: (1) maintains feasibility by restricting the input matrices of each node to a constant size, (2) enables parallel processing, (3) forms a biologically accurate model of the visual cortex, with strongly position-dependent lower-level cells, and position-independent higher-level cells. Furthermore, higher-level cells can represent increasingly more complicated input-output functions (starting with degree of 2 at the lowest-layer, and increasing as 4, 8, 16, and so on.)

IV. Experimental Results

The experiments are divided into two sets to distinguish capabilities that are already offered by feed-forward SFA, versus the newly introduced ones. In the first set, we demonstrate feed-forward SFA in various cases, such as a single moving object, a single object that disappears and reappears again, and multiple objects moving around. These are also interesting as a proof-of-concept that the original SFA approach is feasible for object recognition in real-world images. The second set demonstrates recurrent SFA in an object permanence scenario. Specifically, we show that, when recurrence is introduced, the tunnel and screen effects emerge. We further demonstrate how it is possible to model an increasingly mature prefrontal cortex, by manipulating the single parameter. For each set, same object and behavior was used for both training and testing.

A. Feed-forward SFA

The first experiment shows the response of the feed-forward SFA to an object traversing the x-axis from left to right (Figure 2a-h, data was grayscaled to remove the color cue, which makes classification too easy for different objects.) This set is important for establishing a basis of the output shapes. Figures 2i and 2j show the slowest two signals, whose shapes are exactly as predicted by the theoretical analysis [3]. Let $[t_A, t_B]$ denote the whole experiment duration, and $[t_a, t_b] \in [t_A, t_B]$ a time interval in the experiment during which the object is visible. A single pattern is visible during $[t_a, t_b]$, and is out of the view during $[t_A, t_B]\backslash[t_a, t_b]$ ($\backslash$ indicating set difference). The case with $t_a \neq t_A$ and $t_b \neq t_B$, is called a *bounded* case, since the output must equal to a constant c_1 all during the interval $[t_A, t_B]\backslash[t_a, t_b]$, given that the system sees the (approximately) same background all the while. Due to the zero mean constraint (Equation 1), c_1 tends to 0 in the limit $(t_B - t_A) \to \infty$. The analysis predicts that the slowest signal (y_1) should be a half cosinus, with the second slowest signal (y_2) being a sinus of a single oscillation. (The other signals which are not shown here are cosinuses and sinuses of increasing oscillations.)

The slowest two signals have a significance: They predict the object's 1D position uniquely. On the phase diagram of y_1 vs. y_2 (Figure 2k), every point corresponds to a single position on the x axis. This is because, as shown previously, the SFA outputs reflect the main underlying free variable causing the change in the system, called the **configuration variable**, which in this case is the position. Exact position values can be estimated via a simple regression [2]: Figure 2l depicts the actual and regressed position values.

Ideally, one would like to predict the states of the configuration variables based on the outputs. However which output combination would be necessary or sufficient is not automatically given by the network. For instance, in this case, notice that y_1 on its own is not enough to retrieve the position values, and neither is y_2, due to nonlinearities of both signals. Here, a combination of the two is sufficient. However different transformations need different outputs to be combined. When there is more than one configuration variable, this can be even more complicated: In one case in [1], where both position and identity are changing, a combination of y_1 and y_3 estimate the position, while y_2 and y_4 estimate the identity. So far, a qualitative (human-supervised) assessment have been used to decide. We propose using mutual information for a quantitative assessment, without supervision. Specifically, we calculate the mutual information between all the output combinations, and

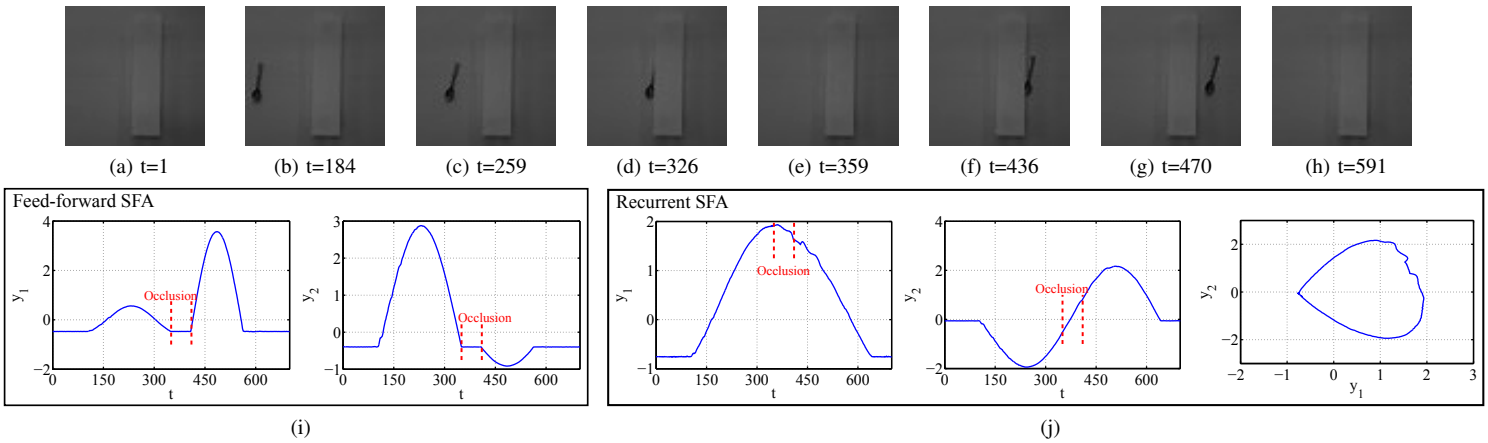


Fig. 3: A single moving object which is occluded for some time (between t=350 and 410) during the trial. (a-h) Snapshots from the input images. (i) Feed-forward SFA. (j) Recurrent SFA and the "Tunnel Effect".

the position values, then perform a thresholding to select the minimum sufficient number of outputs. Figure 2m displays the mutual information provided by y_1, y_2 and $y_1 \& y_2$. As expected, $y_1 \& y_2$ is sufficient for this case.

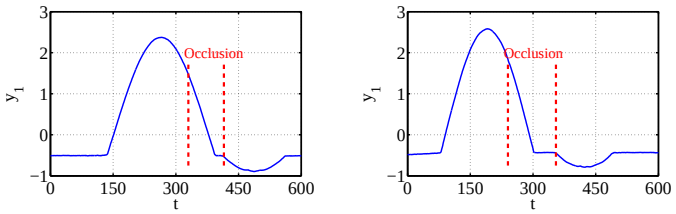
The second experiment stands as a proof-of-concept: As the object is traversing the retinal plane, it disappears behind a screen. It continues to move behind the screen with a constant velocity, and reappears in due time (Figure 3a-h). As expected, due to the reactive nature SFA, as soon as it disappears, the SFA outputs diminish to 0, and on its reappearance they increase again (Figure 3i).

A final issue is the response of the system to more than one object. In this case, there are two objects, the first one in view at t=100-360; the second one at t=710-880. Both are occluded shortly, the first between t=230-250, and the second between t=820-830 (Figure 4a-k). As predicted, the system develops highly object-dependent outputs (4l-n). It is still possible to estimate the position of the objects, but in addition, the outputs also code the identity of the object at any time. For instance, a positive y_1 response during the first visible interval distinguishes the first object from the second one, which has a negative y_1 response for that interval. As shown in [2], a kNN classifier with $\approx 95\%$ success rate can be trained to estimate the identity (Figure 4o).

B. Recurrent SFA

When a recurrent input is added, the system begins to behave similarly with infants with maturating prefrontal cortex. The first indicator is an ability of tracking the position of an object behind a screen. This is demonstrated by the child's ability to guess when it will become visible again (the **tunnel effect**). Figure 3j demonstrate the occluded object case with recurrent input. Recurrent SFA is able to retain its activation throughout occlusion, giving a comparable phase diagram y_1 vs. y_2 with the visible case. This means we can "track" the position of the object uniquely, even through occlusion.

Psychological studies indicate that the tunnel effect depends on the length of the "tunnel". Younger infants are successful for short tunnels only, while older infants can manage increasingly longer ones. A similar effect is observed in Figure 5 with two longer tunnels. Keeping the maturity parameter constant, there is a limit to the occluded period which can be compensated, similar to infants.



(a) Tunnel with size 10 cms. (b) Tunnel with size 14 cms.

Fig. 5: As the tunnel gets longer, Recurrent SFA is no more able to sustain the signals from vanishing.

The final indicator of a mature understanding of object permanence is the **screen effect**, in which the infant maintains the object's identity. She is surprised if a different object reappears from behind the screen. To demonstrate the effect, we show that the architecture has difficulty adjusting when a different object reappears, in which case its predictions collide with the apparent stimuli, resulting in a "surprise". Figure 6a demonstrates the feed-forward case: When Object A disappears behind the screen, and reappears having changed into Object B at time 150, the system responds immediately. Figure 6b demonstrates the recurrent case, with maturity parameters of $\Delta t = 20$ and $\Delta t = 40$, where the system needs time to adjust itself to the changed object. The delays, in which the system insists on seeing Object A, indicate an expectation that the object's identity should have been preserved.

V. Conclusion

We have shown how slow feature analysis, previously shown to develop the object concept, can be extended with a recurrent loop to retain information through time. The proposed extension mimics an important developmental stage, the object permanence understanding. We argue that the building of one ability on top of another is reminiscent of the way humans maturate. The resulting system can predict an occluded object's movements, as well as keeping in mind its identity. These abilities are not infinitely powerful: After a long enough occlusion, they give in, just as in infants. Our study also serves as a minor contribution to the SFA framework: We demonstrate

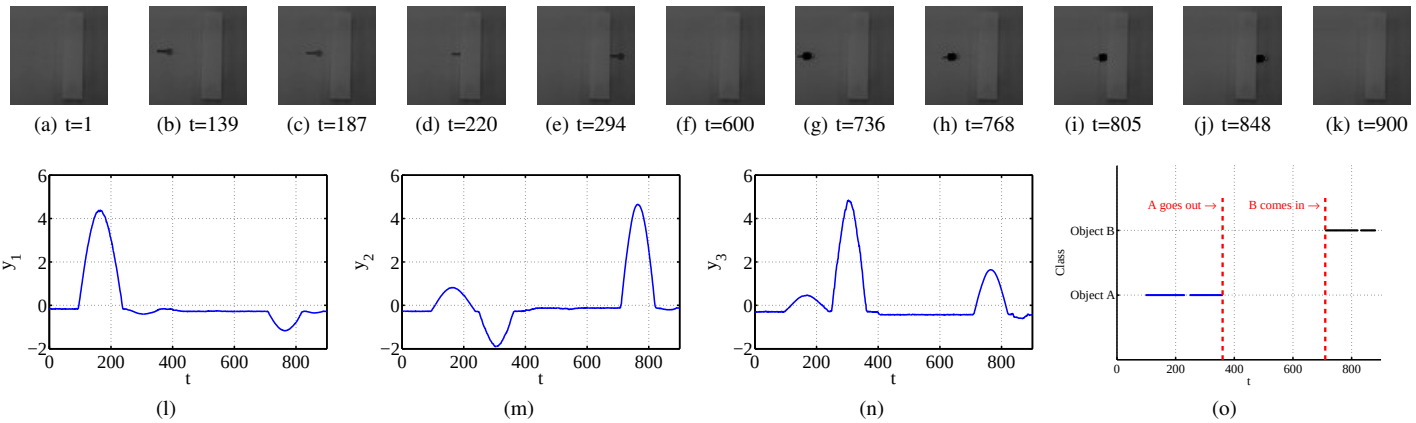
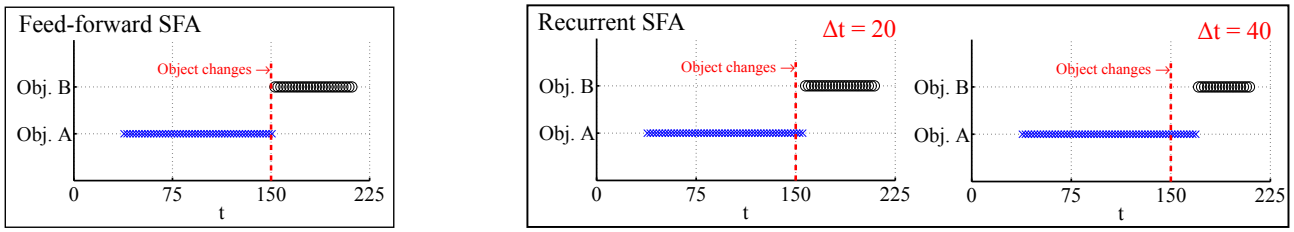


Fig. 4: The feed-forward SFA response to two objects presented sequentially, both of which are occluded for some time. (a-k) The first object is in view from t=100 to 360, occluded between t=230 and 250; the second object is in view between t=710 and 880, occluded between t=820 and 830. (l-n) The slowest three responses. (m) kNN classification of object identity.



(a) Feed-forward case

(b) Recurrent case

Fig. 6: The "Screen Effect". The identity classification made by the recurrent SFA. Object A is changed with object B at t=150. (a) Feed-forward case. (b) Recurrent case with maturity parameters $\Delta t = 20$ and $\Delta t = 40$. Notice that as the maturity parameter increases, the system takes more time getting over its reluctance to accept the chance.

how mutual information can estimate the sufficient outputs, as well as validating SFA for recognizing real-world objects.

An interesting question is whether a gradually increasing maturation parameter will boost cognitive development, since as shown repeatedly, initial limitations of our body promote development by restricting the complexity. The effect of working memory restrictions, other levels of recurrence, and real robotic applications are all promising future directions.

References

[1] L. Wiskott and T. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comp.*, vol. 14, pp. 715–770, 2002.

[2] M. Franzius, N. Wilbert, and L. Wiskott, "Invariant object recognition and pose estimation with slow feature analysis," *Neural Comp.*, vol. 23, pp. 2289–2323, 2011.

[3] L. Wiskott, "Slow feature analysis: A theoretical analysis of optimal free responses," *Neural Comp.*, vol. 15, pp. 2147–2177, 2003.

[4] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *Journal of Vision*, vol. 5, pp. 579–602, 2005.

[5] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *PAMI*, vol. 34, no. 3, pp. 436–450, 2012.

[6] J. Piaget, *The origins of intelligence in children*, 1952.

[7] A. Michotte, "Perception and cognition," *Acta Psychologica*, vol. 11, pp. 69–91, 1955.

[8] A. Diamond and P. Goldman-Rakic, "Comparison of human infants and rhesus monkeys on piaget's ab task: evidence for dependence on dorsolateral prefrontal cortex," *Exp. Brain Res.*, vol. 74, 1989.

[9] M. A. Bell, "Brain electrical activity associated with cognitive processing during a looking version of the a-not-b task," *Infancy*, vol. 2, 2001.

[10] A. Baird, J. Kagan, T. Gaudette, K. Walz, N. Hershlag, and D. Boas, "Frontal lobe activation during object permanence: Data from near-infrared spectroscopy," *NeuroImage*, vol. 16, pp. 1120–1126, 2002.

[11] T. Imaruoka, J. Saiki, and S. Miyauchi, "Maintaining coherence of dynamic objects requires coordination of neural systems extended from anterior frontal to posterior parietal brain cortices," *NeuroImage*, vol. 26, no. 1, pp. 277–284, 2005.

[12] J. Saiki, *Multiple Object Permanence Tracking: Maintenance, Retrieval and Transformation of Dynamic Object Representations*, August 2008, pp. 277–284.

[13] M. Pucak, J. Levitt, J. Lund, and D. Lewis, "Patterns of intrinsic and associational circuitry in monkey prefrontal cortex," *Journal of Comparative Neurology*, vol. 376, pp. 614–630, 1996.

[14] G. Alexander, M. Delong, and P. Strick, "Parallel organization of functionally segregated circuits linking basal ganglia and cortex," *Annual Review of Neuroscience*, vol. 9, pp. 357–381, 1986.

[15] P. F. Dominey, M. Hoen, J.-M. Blanc, and T. Lelekov-Boissard, "Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and erp studies," *Brain and Lang.*, vol. 86, 2003.

[16] Y. Chen and J. Weng, "Developmental learning: A case study in understanding object permanence," in *Fourth Int. Workshop on Epigenetic Robotics*, 2004, pp. 35–42.

[17] D. Roy, K.-Y. Hsiao, and N. Mavridis, "Mental imagery for a conversational robot," *IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics*, vol. 34, pp. 1374–1383, 2004.

[18] Y. Yamashita and J. Tani, "Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment," *PLoS computational biology*, vol. 4, 2008.

[19] M. Franzius, H. Sprekeler, and L. Wiskott, "Slowness and sparseness lead to place, head-direction, and spatial-view cells," *PLoS Computational Biology*, vol. 3, no. 8, pp. 1605–1622, 2007.

[20] V. R. Kompella, M. Luciw, and J. Schmidhuber, "Incremental slow feature analysis," in *Proc. of* 22^{nd} *Int. Jnt. Conf. on Artificial Intelligence*, vol. 2, 2011, pp. 1354–1359.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Humanoids Learning to Crawl based on Natural CPG-Actor-Critic and Motor Primitives

Cai Li, Robert Lowe and Tom Ziemke
Interaction Lab
University of Skovde
Skovde, Sweden
Email: gauss.lee, robert.lowe, tom.ziemke@his.se

***Abstract*—In this article, a new CPG-Actor-Critic architecture based on motor primitives is proposed to perform a crawling learning task on a humanoid (the NAO robot). Starting from an interdisciplinary explanation of the theories, we present two investigations to test the important functions of the layered CPG architecture: sensory feedback integration and whole-body posture control. Based on the analysis of the experimental results, a generic view/architecture for locomotion learning is discussed and introduced in the conclusion.**

I. Introduction

Bio-inspired approaches have been widely applied to model locomotion capabilities in robotic applications[1]. Especially in soft robotics[2], how the neural controller, the body morphology and the environment interact is a challenging theme for locomotion modelers. Firstly, the neural controller has to be able to deal with a number of sensory inputs. Secondly, the neural controller embedded in different bodies has to be morphology-independent, in which case the generic neural controller is transferrable to different morphologies. Finally, it is arduous to evaluate the interaction of a locomotion capability with the environment. In this article, we propose a CPG architecture that aims to deal with all three challenges.

Using central pattern generators (CPGs) is one mainstream bio-inspired solution to modelling quadrupedal locomotion[3][4][5]. In most work, CPGs are used as sensory-input-dependent neural networks of which the output is considered as a force or trajectory generator. According to Ijspeert et al[6], there are two modelling objectives for locomotion capabilities: One is to identify a baseline behavior which contains fundamental patterns for a type of motor ability, for example the coordination of joints. After this is accomplished, the other is to identify how the baseline patterns can adapt to complex and dynamical changes pertaining to the environment or the physical body. Hence, based on this theory, the CPG network should be extensively reformed into a generic architecture where not only the two modelling aims can be fulfilled but also the network's adaptation/learning abilities can be strengthened. In this article, a layered CPG architecture is proposed on the basis of Rybak et al's neuroscientific research[7] and follows the same modelling ideology to the similar purposes. With this architecture, it is possible to utilize typical machine-learning methods to learn the adaptive dynamics.

Reinforcement learning (RL) is an effective mechanism for locomotion learning. From the neuroscienfic perspective, Grillner et al[8] highlight the connection of the brainstem and basal ganglia structures, to which RL is closely related, to CPGs. Based on experiments by Adolph et al[9], the infants learn locomotion through self-correction via thousands of failures. Clearfield et al[10] experimentally illustrate that infant locomotion learning is related to social/emotional interaction. Reinforcement learning might be an appropriate approach to modeling locomotion learning processes as it theoretically describes sorts of emotion-related learning algorithms applicable to the agent/robot interacting with the environment.

In this article, a new layered CPG-Actor-Critic architecture based on motor primitives[6] is proposed to learn crawling on the NAO robot. In section II, the detailed architecture of the layered CPG network and learning algorithm is introduced. In section III, the CPG-Actor-Critic is tested on the simulated NAO robot and the results are transferred to the physical robot for validation. Learned joint dynamics for crawling is demonstrated for analysis. In the final section, the conclusion about a generic neural structure of locomotion learning is drawn for the purpose of implementing locomotion learning in a robot.

II. Methods and Theories

A. The layered CPG architecture

Central pattern generators have been investigated to model locomotion in a lot of robotic applications[1]. However, the adaptation/learning capabilities of CPGs used on those robotic platforms are limited. Firstly, the approaches to involving adaptation/learning in most applications can only deal with several specific types of adaptation and based on several methods such as sensor-information extraction[11][12], neural connectionism alternation[3], demonstrated information[13] and so forth. Secondly, the CPG structure lacks a generic compatibility with which any above-mentioned adaptive methods can work together. According to Grillner et al's research[8], CPGs biologically are able to assimilate two functions: sensory feedback integration and posture control. If each degree of freedom of the robotic joint is considered as a stable limit cycle, sensory feedback integration is the function of reshaping the limit cycle and the posture control is applied in order to shift the oscillation center.

Rybak et al[7] uncovered the possible biological mechanics of CPGs (Figure 1 left). In this schema, the rhythmic generator (RG) layer provides a primitive source of oscillatory signals. The pattern formation (PF) layer is a level on which all the RGs are mutually connected to form the phase-separated output. This layer contains fundamental characteristics of one type of locomotion. For example, crawling is experimentally

observed to be one type of locomotion featured in anti-phase movement of the ipisilateral limbs and in-phase movement of diagonal limbs[14]. The dynamics adaptation represents the functions of motorneurons of which the output is sent directly to muscle spindles. In this layer, the output of PF layer is adapted into distinct dynamics in order to adapt to different environment or interactions. After reshaping the output of PF layer in dynamic adaptation (DA) layer, the RG itself turns out to be a "clocking" driver for CPGs. The above three-layer architecture has been implemented to model walking behaviors in explaining locomotion models[12][15]. However, the difficulty of implementing this structure is that there is no generic models of the DA layer so that the robot without proper sensors loses the ability to locomote. In this article, this problem might be solved with the proposed solutions presented below.

Mathematically, RGs can be modelled by different oscillators (e.g. Hopf oscillators[15]). In this work, a phase oscillator is chosen:

$$\begin{aligned}
\dot{r_i} &= a_i(R_i - r_i) \\
\dot{W_i} &= 2\pi\omega_i + K_i \\
K_i &= \sum_j w_{ji} \cdot r_j sin(W_j - W_i - P_{ji}) \\
\omega_i &= \frac{\omega_{1i}}{e^{-100 \cdot Aex_i} + 1} + \frac{\omega_{2i}}{e^{100 \cdot Aex_i} + 1} \\
Aex_i &= r_i \cdot sin(W_i + \frac{\pi}{2}) \\
A_i &= r_i \cdot sin(W_i)
\end{aligned}$$

where A_i is the output of this phase oscillator and Aex_i is the frequency control output. r_i and W_i are the amplitude and phase variables. ω_i is the frequency of the oscillator with ω_{1i} and ω_{2i} controlling the ascending and descending frequency. K_i is the connection term from the other oscillators to oscillator i. wji is the connection weight of from oscillator j to i. W_j is the phase of oscillator j and P_{ji} is the phase difference from oscillator j to i. a_i and R_i are the convergence rate and converged value of amplitude. In our work, the parameters settings are as follows: $a_i = 50$, $R_i = 1.0$, $w_{1i} = w_{2i} = 1.0$. In the PF layer, a four-cell CPG network is utilized as a CPG core to drive the motion of each joint (details are in Figure 1). This CPG core has the capabilities to maintain structural stability according to group theory[16] and been verified to generate basic patterns of both crawling and walking by adapting parameters w_{1i} and w_{2i}[5][17][18].

In the DA layer, the model of motor primitives is applied to adjust the output of correspondent PF-layer neurons. The motor primitive model is:

$$\begin{aligned}
\tau\dot{z_i} &= \alpha(\beta(g_i - y_i) - z_i) + amp \cdot A_i + f \\
\tau\dot{y_i} &= z_i \\
f(W_i, p) &= \frac{\sum_{j=1}^{N} \psi_j v_j}{\sum_{j=1}^{N} \psi_j} p_i \qquad (1) \\
\psi_j &= exp(h_j(cos(W_i - c_j) - 1)) \qquad (2) \\
\tau\dot{g} &= \alpha_g(g_0 - g) \\
\alpha = 8.0, \beta &= \frac{\alpha}{4}, \alpha_g = \frac{\alpha}{2} \qquad (3) \\
c_j &\sim (0, 2\pi)
\end{aligned}$$

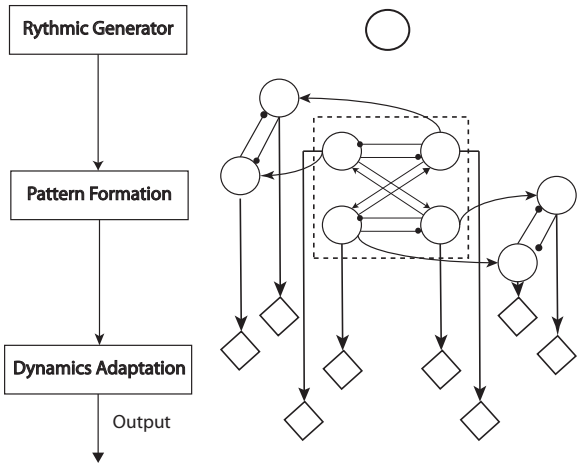


Fig. 1. Left: The functional structure of CPG anatomy and each block represents one-layer functionality. Right: The neural structure of CPG employed for crawling. The single circle above most represents the RG layer as an oscillator. The recurrent neural network composed of connected circles represents the function of the PF layer. The diamonds represent functions of the DA layer. Within the PF-layer network, the four-cell network in the dash-line frame controls the rhythms of pitch motion for Shoulders and Hips. The other four outside the dash-line frame controls the roll motion of Shoulders and Hips. The arrow-head lines represent in-phase oscillation (the phase difference between two oscillators is 0 or 2π). The dot-head lines represent anti-phase oscillation (the phase difference between two oscillators is π)

where z_i, y_i and g_i are the variables of the motor primitive. amp is the weight of correspondent input from PF layer and set to 27 which makes the output of motor primitive to oscillate between -1 and 1. A_i is the "clocking" input from the PF layer. τ is the time constant which is equal to the period ($\frac{1}{\omega_i}$) of input A_i. f is the forcing term in which ψ_j are fixed basis functions, v_j are the weights and p_i is the amplitude which is equal to amp. $N = 50$ represents the number of basis functions. Using nonlinear arbitrary functions in f is a well-defined approach in machine learning[19] for nonlinear regression and analogous to population coding for computational neuroscience[20]. In equation (2), h_j is a constant equal to $2 \cdot N$ and W_i is the phase input from the PF layer. c_j is vector containing N separations of the scope in $(0, 2\pi)$. g_0 is the anchor point ($g_0 = 0$). Equation group (3) guarantees the damping convergence of the motor primitives.

Motor primitives are widely used to model discrete motor learning[21][22] and rhythmic movement[6]. In terms of periodic movement learning, Gams et al[23] and Nakanishi et al[24] employed demonstrated signals to learn motor primitives of rhythmic motion with local weighted regression. However, supervised learning might not always be the case in locomotion learning. Infants learn to crawl by interacting with the environment rather than being demonstrated how each joint moves dynamically[25][26]. Locomotion learning based on RL without demonstrated signals and motor primitives are also popular[27][28]. However, as the motor primitives model has a good learnability, in this article, a new approach of using motor primitives and RL for locomotion learning without demonstrated signals is proposed.

B. Natural CPG-Actor-Critic

The CPG-Actor-Citic architecture has been used for exploring and learning complicated locomotion patterns for both

bipedal[29][30][15] and quadrupedal robots[31][32]. Inspired by Grillner et al[8], the functions of CPG-Actor-Critic connects the layered architecture to an affective learning (RL) process in which the optimal parameters of CPGs are determined. The actor generates actions by exploring the state space and the critic evaluates the actions taken by observing rewards. Policy gradient is a well-established method used to update the parameterized action space. Since the normal "vanila" gradient suffers the slow learning rate, natural policy gradient is employed to speed up the learning in the CPG-Actor-Critic architecture. This is the so-called Natural CPG-Actor-Critic.

As previously mentioned, learning locomotion is a repetitive task. As such, episodic natural policy gradient (eNAC) is selected. eNAC is well-known for its learning efficiency on searching optima in a continuous parameterized space. Compared to Cacla (continuous action space learning automaton)[33], another efficient continuous-space RL, eNAC might suffer in possible failures by updating a parametrized model into an uncertain action space[33]. So a positive eNAC is applied in our work (details in II-C)to rule out the potential failures. NAC is proposed by Kakade et al[34] and further developed and used in motor learning by Peters et al[22][35]. It transforms the traditional RL problem of solving the Bellman equation to an explorative process of linear/non-linear regression. As an efficient policy gradient approach, the basic principles of NAC are as follows:

Assume the stationary policy is $\pi^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{u})$ which can determine action space $\boldsymbol{u}$ based on state space $\boldsymbol{x}$ with a static distribution $d^{\pi}(x)$. The immediate reward is $r(x_t,u_t)$ and baseline value is b. According to policy gradient theorem[22], the monte-carlo expected reward $J(\boldsymbol{\theta}) = \int_{\boldsymbol{x}} d^{\pi}(\boldsymbol{x}) \int_{\boldsymbol{u}} \pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x})(\sum_{t=1}^{T} \alpha_t r(x_t,u_t) - b)$ and its normal policy gradient can be written as:

$$\bigtriangledown_{\theta} J(\boldsymbol{\theta}) = \int_{\boldsymbol{x}} d^{\pi}(\boldsymbol{x}) \int_{\boldsymbol{u}} \pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x}) \bigtriangledown_{\theta} log(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x})) (Q(\boldsymbol{u},\boldsymbol{x}) - b) d\boldsymbol{x} d\boldsymbol{u}$$

$$Q(\boldsymbol{u},\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t r(x_t,u_t)$$

where the policy $\pi^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{u})$ is derivable at the policy parameters $\boldsymbol{\theta}$, namely $\bigtriangledown_{\theta}\pi^{\boldsymbol{\theta}}$ exists. $Q(\boldsymbol{u},\boldsymbol{x})$ is the action-state function. By using function approximation approaches, $Q(\boldsymbol{u},\boldsymbol{x}) - b$ can be approximated by $\bigtriangledown_{\theta} log^T(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x}))\boldsymbol{w}$, the policy gradient can be rewritten as:

$$\bigtriangledown_{\theta} J(\boldsymbol{\theta}) = \int_{\boldsymbol{x}} d^{\pi}(\boldsymbol{x}) \int_{\boldsymbol{u}} \pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x}) \bigtriangledown_{\theta} log(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x})) \bigtriangledown_{\theta} log^T(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x}))\boldsymbol{w} d\boldsymbol{x} d\boldsymbol{u} \quad (4)$$

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + \alpha \bigtriangledown_{\theta} J|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \quad (5)$$

In equation (4), $\bigtriangledown_{\theta} log(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x}))$ is the basis function vector related to state space $\boldsymbol{x}$ and $\boldsymbol{w}$. By and large, Equation (5) plot the rudimentary rule of thumb for policy gradient approaches. For searching maximal expected reward $J(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ in a faster speed, natural policy gradient will maximize the Kullback-Leibler distance between $J(\boldsymbol{\theta} + \triangle\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ to update searching policy $\pi^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{u})$ until it converges[22]. n represents the n-th step of update and α is the learning rate (equal to 0.01). By using natural policy gradient, the normal form is turned into:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + \alpha F_{\theta}^{-1} \bigtriangledown_{\theta} J|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} = \boldsymbol{\theta}_n + \alpha \boldsymbol{w} \quad (6)$$

$$F_{\theta} = \int_{T} \pi^{\boldsymbol{\theta}} \bigtriangledown_{\boldsymbol{\theta}} log\pi^{\boldsymbol{\theta}} \bigtriangledown_{\boldsymbol{\theta}} log\pi^{\theta} d\boldsymbol{x} d\boldsymbol{u}$$

where F is the Fisher Matrix (FM) and $\boldsymbol{w}$ is the weight vector. Multiplied by FM, normal policy gradient (Equation (5)) is changed to the steepest one (Equation (6)). Then the RL problem is transformed to figure out the approximation of $Q(x,u)$ function with the basis functions by searching for a proper weight vector. According to the derivation of eNAC[22], the weights, obtained by least square learning, can be employed to update the policy parameters directly and the calculation of the weights is:

$$\begin{bmatrix} \boldsymbol{w} \\ J \end{bmatrix} = (\phi\phi^T)^{-1}\phi R.$$

$$\phi = [\sum_{t=1}^{s} \alpha_t \bigtriangledown log^T(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}_t|\boldsymbol{x}_t))\boldsymbol{w}, 1]_{1:H}^T \quad (7)$$

$$R = [\sum_{t=1}^{T} \alpha_t r(\boldsymbol{x}_t,\boldsymbol{u}_t)]_{1:H}^T \quad (8)$$

where $1:H$ represent H times samplings within one trial (refer to details in the Algorithm). ϕ is the basis vector and constant 1 is used to determine the baseline J avoiding large-variance update. α_t is the theoretical discounting factor. R is the average reward vector in which r is the instant reward (for the detailed eNAC proof, please refer to[22]). In this article, motor primitives are embedded in the eNAC algorithm. Within motor primitives, the weights v_j (in equation (1)) are the parameters and basis functions (equation (2)) represent the states of a motor primitive. Assume that the actor emits the action $\boldsymbol{a}$ according to a gaussian policy, then the action can be written as:

$$\boldsymbol{a} = \boldsymbol{\theta}^T\boldsymbol{\psi_f}(\boldsymbol{x},\boldsymbol{t}) + \boldsymbol{\epsilon_t}$$
$$\epsilon_t^n \sim N(0,(\sigma^n)^2)$$

where $\boldsymbol{\theta}$ is the policy parameter vector reflecting the weights v_j in motor primitives. $\boldsymbol{\psi_f}$ is the vector of normalized basis functions of motor primitives. $\boldsymbol{\epsilon_t}$ is the gaussian exploration vector with deviation $\boldsymbol{\sigma}$ at time t and ϵ_t^n is the exploration for the nth basis function in $\boldsymbol{\epsilon_t}$. On the other hand, the actor can also be altered to another form:

$$\boldsymbol{a} = (\boldsymbol{\theta}^T + \boldsymbol{\epsilon_t}^T)\boldsymbol{\psi_f}(\boldsymbol{x},\boldsymbol{t}) \quad (9)$$
$$\epsilon_t^n \sim N(0,(\sigma^n)^2)$$

Equation (9) is state-dependent exploration[36] since the exploration of the action is $\boldsymbol{\epsilon_t}(\boldsymbol{\psi}) = \boldsymbol{\epsilon_t^T}\boldsymbol{\psi}(\boldsymbol{x},\boldsymbol{t})$. In order to reduce the computational load technically, using equation (9) can save one-time matrix multiplication for exploration. The action follows the new gaussian function $\boldsymbol{a} \sim N(\boldsymbol{\theta^T\psi_f}, \boldsymbol{\psi_f^T\Sigma\psi_f})$ and the RL basis function becomes $\bigtriangledown_{\theta} log(\pi^{\boldsymbol{\theta}}(\boldsymbol{u}|\boldsymbol{x})) = \frac{\boldsymbol{\epsilon_t^T\psi_f\psi_f^T}}{\boldsymbol{\psi_f^T\Sigma\psi_f}}$, where $\boldsymbol{\Sigma}$ is the diagnal deviation matrix. if the same deviation is employed for all the parameters and $\boldsymbol{\psi_f}$ is normalized, $\boldsymbol{\psi_f^T\Sigma\psi_f}$ turns out to be $\boldsymbol{\sigma^2}$.

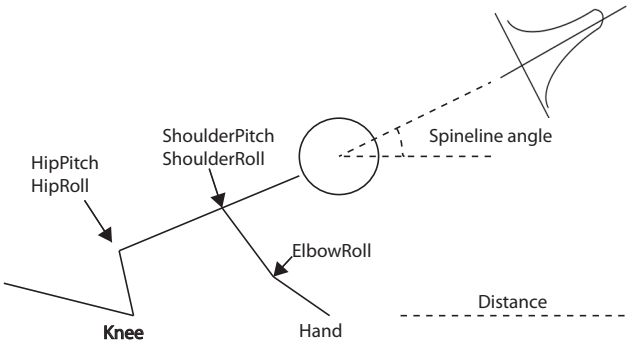


Fig. 2. The standard crawling posture on knees and hands and the main joints controlled by CPGs. The distance and spineline angle indicate the quality of crawling. The spineline angle is controlled by a gaussian function.

C. Learning to Crawl

With the above-mentioned eNAC algorithm, the robot is able to explore the dynamics of each joint on its own according to the specific reward function. Figure 2 shows the standard crawling (crawling on knees and hands[37]). The main joints controlled by CPGs are the ones located at the hip and shoulder. The elbow oscillates with the rhythms of the shoulder pitch. Since crawling is left-and-right symmetric[14], the number of degrees of freedom (DOFs) can be reduced from 8 (left and right joints) to 4 (left or right joints only). Therefore, the parameters for standard-crawling learning is $4 \cdot 50 = 200$ (where 4 is the number of DOFs and 50 is that of basis functions). From the previous work[31], the move distance and spineline angle (Figure 2) are two significant factors to evaluate the quality of crawling behaviors. Accordingly, in the CPG-Actor-Critic architecture, the reward function is composed of two terms ($r_{distance}$ and r_{angle}) as two evaluation landmarks for the above-mentioned two variables:

$$\begin{aligned} r_{reward} &= r_{distance} + r_{angle} \\ r_{distance} &= exp(\frac{D}{2}) - 1 \\ r_{angle} &= exp(e) - 1 \\ with\, e &= N(\mathbf{x_0}, \sigma = 0.02) \end{aligned} \tag{10}$$

where D is the distance the robot crawls every episode. e is a gaussian distribution with the center x_0 and variance σ. Using e is possible to maintain the posture of standard crawling without learning some extreme postures[31]. In the case of infants learning to crawl, this function works like parents' hands adjusting or holding up the infant's body when she/he is crawling.
The pseudo-code of the learning-to-crawl algorithm based on natural CPG-Actor-Critic can be summarized as:

eNAC Algorithm:
Repeat M trials each of which includes 10 rollouts (H=10), In each rollout, action is generated by$\boldsymbol{a} = (\boldsymbol{\theta}^T + \boldsymbol{\epsilon_t^T})\boldsymbol{\psi_f}$ where $\boldsymbol{\epsilon_t} \sim N(0, \sigma^2)$ ($\sigma = 0.05$) for t = 1,2,3....s
Calculate:
for each rollout, the episodic return $r_i = \sum_j r_{angle}(j)$
the eligibility $\Psi_i = \sum_{t=1}^{s} \sigma^2(\boldsymbol{\epsilon_t \psi})\boldsymbol{\psi^T}$
after each rollout, $r_i = r_i + r_{distance}$ then the gradient is:

$$\begin{bmatrix} \mathbf{w} \\ J \end{bmatrix} = (\phi\phi^T)^{-1}\phi R.$$

where $R = [r_1, r_2, ...r_H]^T$ and $\phi = [\Psi_1, \Psi_2, ..., \Psi_H]^T$
Updating for each trial:
if $\delta > 0$, with $\delta = R_{avg} - V_{n-1}$ where R_{avg} is the average of R and V_{n-1} is the episodic value function of last updating:
$V_n = V_{n-1} + 0.1 \cdot \delta$ and $\theta_{n+1} = \theta_n + \alpha\mathbf{w}$, otherwise no updating.
Until the convergence condition is satisfied: $\delta < 0$ all the time or $|\delta| < 10^{-4}$.

It is noteworthy that, inspired by Cacla architecture[38], the "positive updating" is used to avoid the inappropriate updating in the parametrized action space. Since the function approximation cannot accurately converge to the real Q function ($\mathbf{w}$ cannot be zero), the convergence condition is necessary to determine the termination of each learning process.

III. Investigations and Analysis

The objective of investigations in this section is to verify the capability of motor primitives based on CPG-Actor-Critic on dynamics modification and postural control. There are two investigations: One is to test the learnability of the CPG architecture by using a generic "reshaping" mechanism (Equation (1)) with the same targeted posture (the same spineline angle). The other is to test if the generic motor primitives can also adjust the joint posture (shifting centers of limit cycles) under the condition that the posture control reward is set to two different targets.

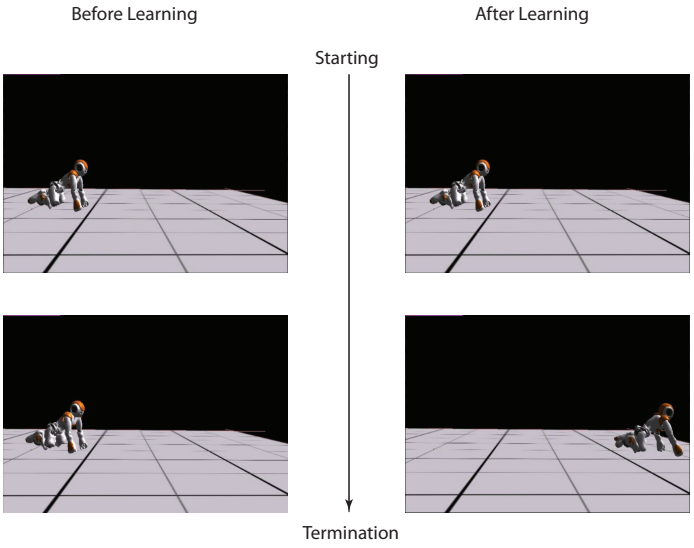


Fig. 3. Left: the beginning and termination snapshots of crawling before learning. Right: the beginning and termination snapshots of crawling after learning

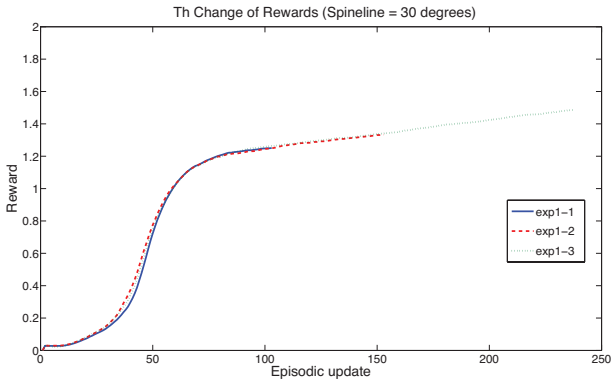


Fig. 4. Learning curves of three trials in investigation 1. run1-1∼3 represent the three independent trials in investigation 1.

A. Investigation 1:Proprioceptive Sensory Feedback Integration

In this investigation, with the average spineline angle fixed at 30 degrees ($x_0 = 1.05$), the simulated robot learns to crawl in three independent runs and finally converges to three different results by balancing the distance maximization and posture maintenance. Every run starts with the same initial posture with ($x_0 = 1.08$, approximately 28 degree) and performs a bad non-crawling behavior with no crawling distance (Figure 3). However, after learning, the standard crawling emerges from the interaction amongst the CPG-Actor-Critic architecture, the humanoid body and the environment (For the detailed performance, please refer to the video[39]). Interestingly, the three learning trials converge with similar smooth reward curves (Figure 4) but different results (Figure 5).
In order to clearly investigate the reasons of the formation of crawling, the joint dynamics are shown separately in Figure 5. Since the standard crawling is a whole-body motion, the CPG-Actor-Critic autonomously decides how to adapt the motion of each joint. The adaptive changes of pitch joints for shoulders and hips focus on the adjustment of their amplitudes (Figure 5. A and B). Especially, the HipPitch joint tends to swing more backward so that robot can crawl forward with more force. Interestingly, extracting from the results in our investigations (Figure 5.C and D), the significant factor determining if the robot can crawl forward properly is the roll motion. Not only the amplitudes of roll joints in shoulders and hips are statistically adapted, but also the phases of CPGs controlling roll joints are shifted compared to the original CPG output without learning. It seems the DA layer modelled as motor primitives has the capability to deal with the integration of sensory information locally and even to adjust the phase difference which is set inappropriately in the PF layer. It can also tune the posture. From the joint dynamics of roll motion, it is clearly observed that the limit cycles of roll joints are shifted, in which case the oscillation centers of roll joints are adaptively adjusted. Compared to the explicit posture-control terms in previous work[15][31], the implicit terms grounded in the motor primitives can integrate two functions: sensory feedback integration and posture control. To verify the functionality of posture control, investigation 2 below is focused on this aim.

B. Investigation 2: Posture Adjustment

In investigation 2, the objective is to verify the capability of the proposed CPG-Actor-Critic architecture on the adjustment of joint posture. Actually, the spineline angle reward proffers a control signal of limiting the whole-body posture. With a loose control coefficient (e.g. $\sigma > 0.02$ in Equation (10)) or without the spineline restriction, the robot will only consider the maximization of crawling distance, ignoring the maintenance of the posture. This causes a convergence to an extreme crawling behavior. In reality, parents always need to guide a right posture by holding up or lifting the infants' body when they are crawling. Therefore, the posture limitation is necessary.
For testing the posture control abilities, two spinelines are chosen ($x_0 = 1.03$ and $x_0 = 1.08$, approximately 31 degrees and 28 degrees). Two independent learning trials are performed respectively for each of these two spineline-angle controlled postures. With the results obtained, the comparison of limit cycles of joints in 4 learning runs are given in Figure 6. For each group of the results (black and red curves), the crawling joint dynamics converge to similar limit cycles. In terms of the motion of pitch joints (shoulders and hips), from Figure 6.A-B, the deviation between two limit-cycle centers is blurry. However, it is conspicuous for the roll joints, especially hip roll joints (Figure 6.C-D). The limit-cycle centers are both shifted rightwards for shoulder and hip roll motion from posture 2 (28 degree spineline angle) to posture 1(31 degree spineline angle). This limit-cycle-center shifts correspond to the closing-inward and opening-outward posture changes of shoulder and hip joints. This is a typical whole-body motion of lifting the gravity center of the body and increasing the spineline angle. Compared to the explicitly allocated posture-change terms in previous work[15][31], using motor primitives can interactively rule out the unnecessary joints for posture control. For example, in investigation 2, to change from posture 1 to posture 2, the system determines to fixate on altering the posture of roll joints other than pitch joints based on the whole-body motion logic.

C. Transferred Test on the Physical Robot

In this article, with 7 learning results, they are transferred to the physical NAO robot for testing. In all the learning trials, the popular Webots simulator[40] based on ODE (open dynamics engine) is used. In order to successfully test the learned motion from the simulated robot to the physical one, some preconditions have to be realized. As discussed in previous work[31], the possible failures of transferred results on physical robots could be caused by the disparity in physics engines and difference between simulation time and real time. In our work, the frequency of the CPG is doubled while being transferred. 5 out of 7 results can be successfully transferred expect the results for the posture ($x_0 = 1.03$). After the CPG amplitudes of pitch joints are reduced to 70%, the transfer to the physical robot is now successfu. Figure 7 shows the snapshots of one-step crawling on the physical robot (for details, please refer to the video[41]). Compared to the previous implementation with only optimized postures, the left-right curvy motion of the spineline, a typical characteristic of crawling behaviors[14], emerges after learning.

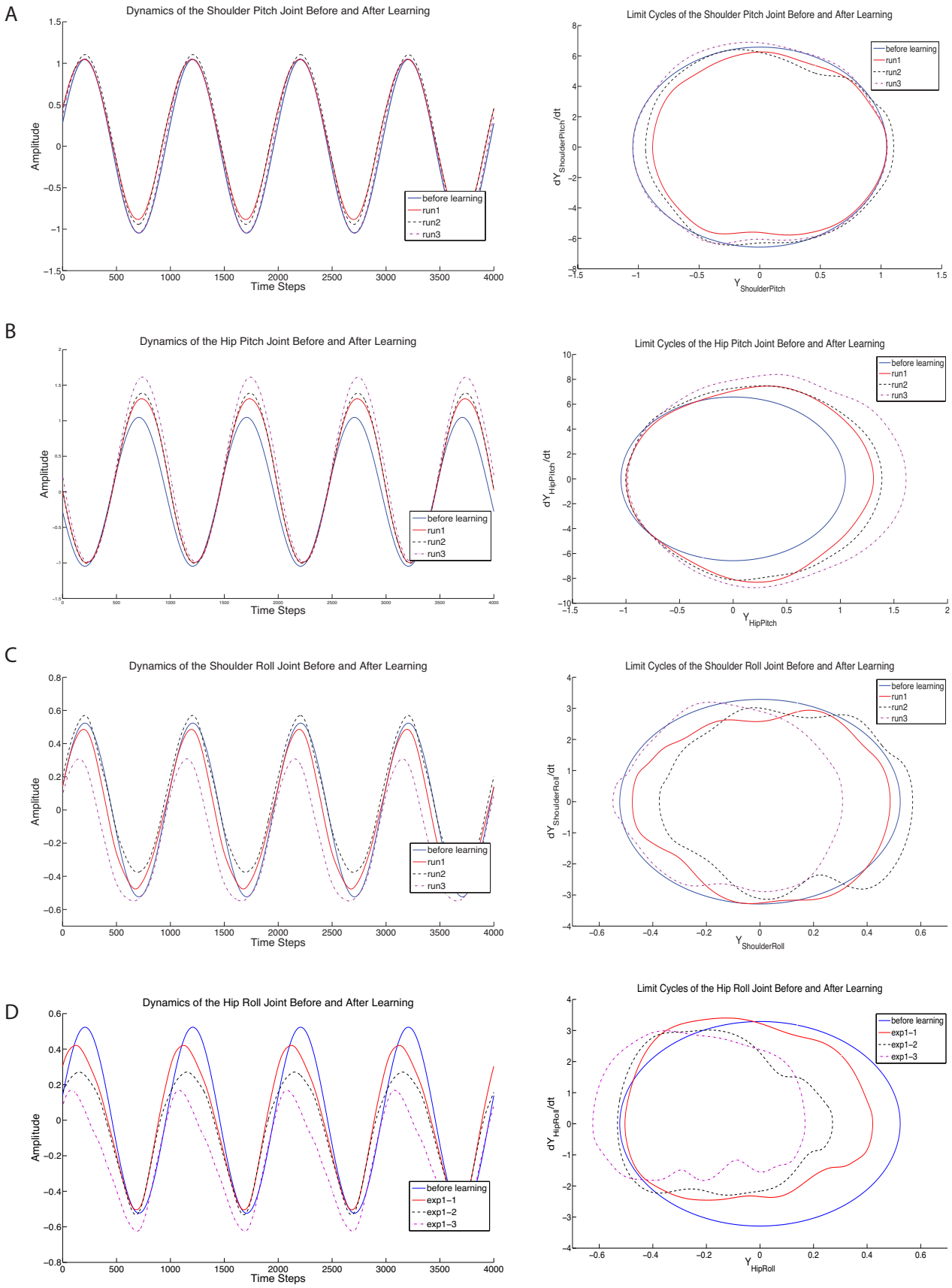


Fig. 5. The dynamics of joints (Shoulder Pitch and Roll, Hip Pitch and Roll) before and after learning. The blue line indicates the original joint motion without learning and the red solid lines, black dashed lines and purple dashed lines show the results of run 1-3 on the right side. The left-side figures are the correpondent limit cycles.

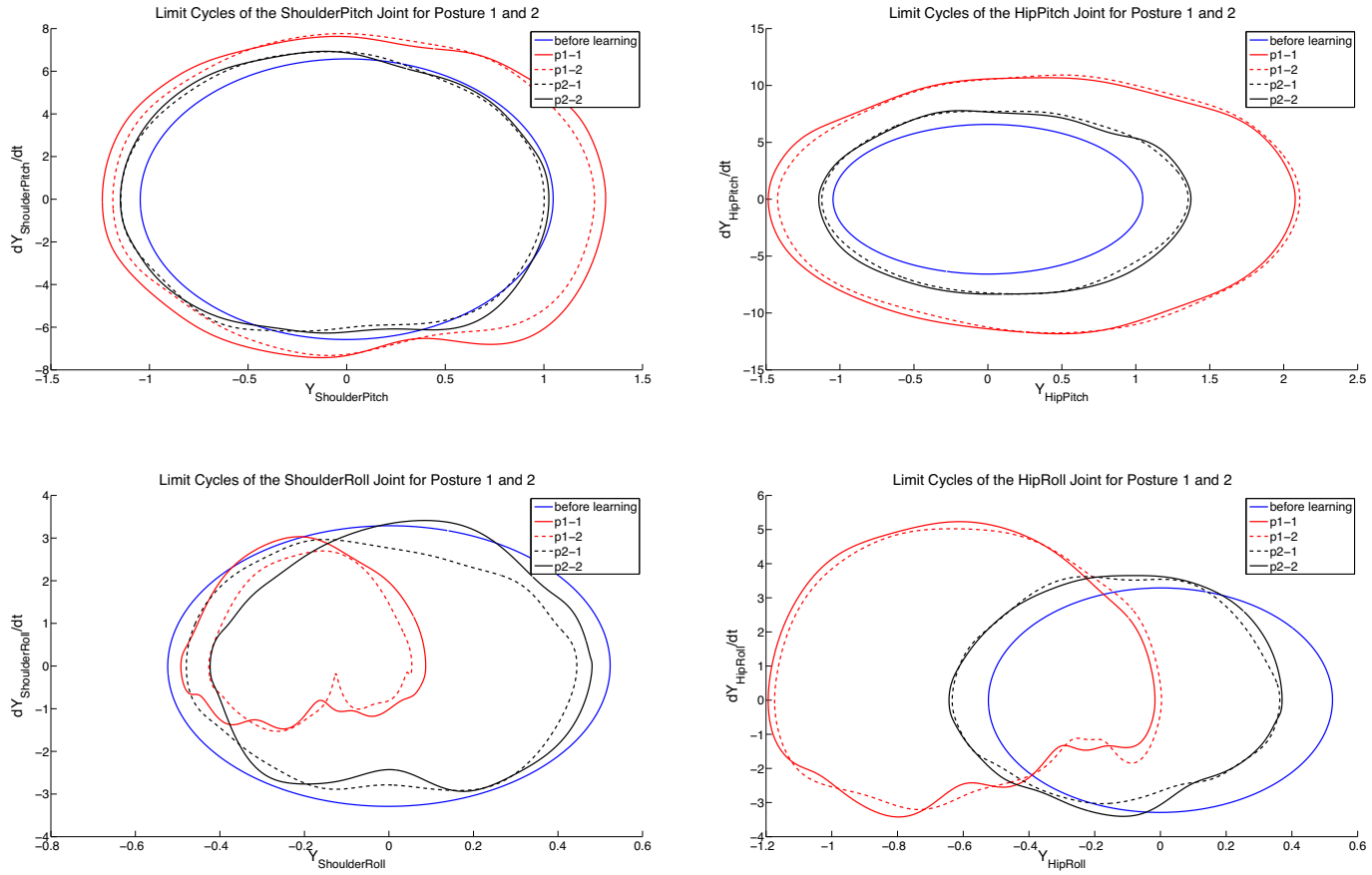


Fig. 6. The figure shows the limit cycles of each controlled joint with 2 independent learning trials for two different body postures. The blue solid lines represent the limit cycles for each joint before learning. Red solid and dashed lines indicate the limit cycles for tested posture 1 ($x_0 = 1.03$, appox.31 degree). Black solid and dashed lines indicate the limit cycles for tested posture 2 ($x_0 = 1.08$, appox.28 degree). pn-n is the abbreviation of posture n-run n. In the investigation, all the outputs of CPGs for Hip-roll, Hip-pitch, Shoulder-roll and Shoulder-pitch joints are resacled by multiplying 0.35, 0.3, 0.3, 0.2.

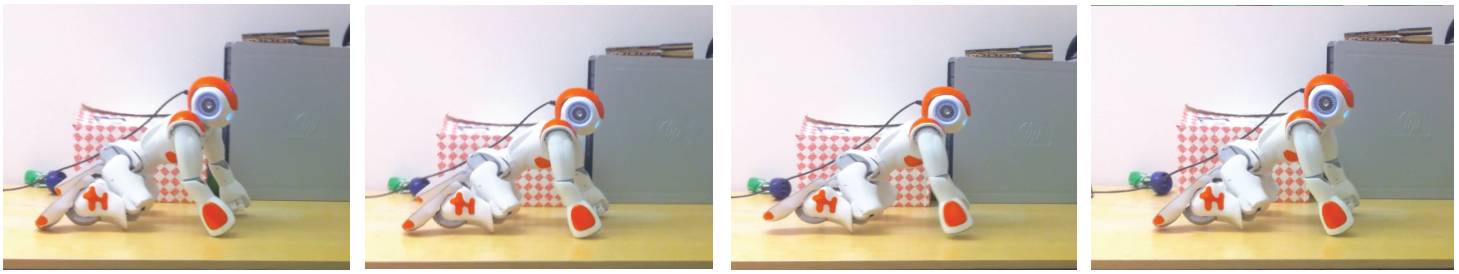


Fig. 7. The implementation on the physical robot. This figure shows the video snapshot of one-crawling-step NAO robot on a wooden flat table (One crawling step means one time alternation of the supporting leg and arm).

D. Summary

In this section, the motor primitives based CPG-Actor-Critic architecture is tested for learning standard crawling from non-crawling behaviors by phase shifting, output reshaping and whole-body posture adjusting. The learned results are transferred to the physical robot. It seems the three-layered CPG model is able to proffer a generic mechanism to seek out the possible answers to the typical problems regarding to CPGs: sensory feedback integration and posture control.

IV. Conclusion

A. Motor Primitives based CPG-Actor-Critic

In this article, the proposed CPG-Actor-Critic based on motor primitives seems to be able to optimize humanoid crawling given an initially rough baseline behavior. The forcing term $f(W_i, p)$ in equation (1) works like a sampling sensor which perceives a large number of proprioceptive points of the CPGs and adapts flexibly them into distinct dynamic patterns on the basis of actor-critic interaction. Even though this implementation of motor primitives with RL approaches instead of supervised learning opens a new page for locomotion learning, this approach still has some disadvantages: Firstly, learning

might be slow. In the above investigation, each run takes about 6-7 hours to complete. Therefore, the learning process might not be transferred to the physical robot. Secondly, the frequency is not adaptive. In the work presented above, all the CPG frequencies are fixed. Even though the motor primitives can innately preserve the learned dynamics when the frequency is changed, it still cannot guarantee that the new-frequency patterns still can work when the whole-body dynamics change with the oscillation frequency. As a matter of fact, after reducing the frequency from 1.0 to 0.5, the robot's body dynamics change and crawling cannot be properly presented. The solution to this problem might be using hybrid learning based on eNAC by counting in the frequency parameters[42]. Thirdly, the implementation of CPGs is not energy efficient. In our work, the CPGs are used as trajectory generators. The layered architecture still lacks of an adaptive approach to altering the stiffness of joints. On the other hand, low energy efficiency might be a natural flaw of rigid body robots. Even though force control might be able to improve the energy issues on a rigid body, the inflexibilities of the body is still a stumbling block stifling a robot from being energy-efficient for locomotions.

In a conclusion, motor primitives based CPGs is able to not only learn demonstrated/supervised signals[23] but also adapt to flexible patterns based RL approaches in our work. It might be possible to combine the two together to form a more general architecture for locomotion-learning tasks.

B. A Generic View of Bio-inspired Locomotion Learning

Even though a lot of inspirations related to locomotion learning/development can be extracted from cognitive science (e.g. Thelen et al[43]), neuroscience (e.g. Schore et al[44], Grillner et al[8]), psychology (e.g. Clearfield et al[26], Adolph et al[9]) and robotics (e.g. Pfeifer et al[2]). From the perspective of Thelen et al[43], locomotion development/learning is focused on the formation and adaptation of the so-called "attractors" in a dynamic system. This assumption offers a baseline for locomotion learning that what should be focused on is not how a static system can be modelled but a dynamic system might be developed. The stagnation is only one "special" attractor of the system. In this sense, motor primitives have been assumed to represent locomotion attractors in Ijspeert et al's work[6]. Schore et al[44] and Grillner et al[8] both imply that locomotion learning might be RL-related from the perspective of neuropsychology and neural structures. From psychological point of view, Clearfield et al[26] indicates the developmental relation of locomotion to spatial memory including the distance. Adolph et al[9] recently explain why the repetitiveness is important for infants to learn locomotion. Finally, Pfeifer et al[2] rethink the locomotion and emphasize the interaction between the body and the environment. Based on the above-mentioned comprehension of locomotion from different angles, locomotion learning is an affective-related, interactive and repetitive process with cognitive cues. Therefore, it seems the motor primitives based CPG-Actor-Critic capsulated the possible potentials to cover the explanations from each view of them. A general CPG architecture in Figure 8 might be able combine the self-learning and supervised learning. In the supervised learning mechanism, the input can be mimicry/demonstrated signals of joint dynamics. The PA layer can be a recurrent neural network

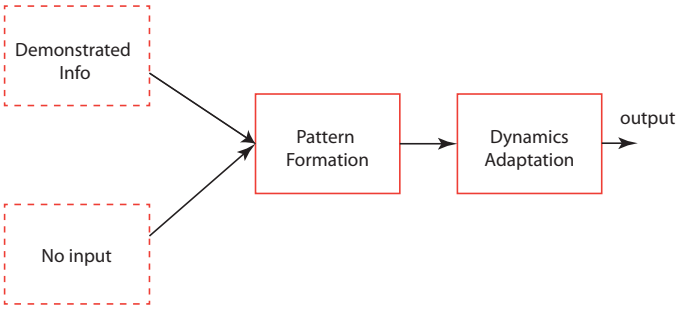


Fig. 8. The generic architecture of locomotion learning. The dashed-line blocks represent the choices of input. The other blocks represent the functions of each layer.

in which each neuron modeled by motor primitives can learn input signals by using local weighted regression. This has been demonstrated in the work by Gams et al[23]. As the approach presented in our approach, the self-learning mechanism does not need supervised input. The PA layer can be constructed as a recurrent neural network composed of oscillators. The main contribution of this paper is that the output of PA layer by both self or supervised learning can be adaptively changed in the DA layer modeled by motor primitives for different tasks or different sensor configuration in the CPG-Actor-Critic architecture. eNAC or PoWER (Policy learning by Weighting Exploration with the Return) learning algorithms[36] can be applied to adapt the motor dynamics.

In future work, the proposed new architecture will be tested on different legged morphologies and even applied to bipedal-walking learning. Meanwhile, the frequency has to be considered as one of the factors for learning. On the other hand, this new model will be modified to integrate the sensory feedback from sensors, which can increase the adaptation of the whole architecture.

ACKNOWLEDGMENT

Authors would like to thank the European RobotDoC project (www.robotdoc.org) for funding and supporting this research.

REFERENCES

[1] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots : a review Neurobiology of CPGs," *Neural Networks*, vol. 21, pp. 642–653, 2008.

[2] R. Pfeifer and J. C. Bongard, *How the Body Shapes the Way We Think: A New View of Intelligence (Bradford Books)*. The MIT Press, Nov. 2006.

[3] N. Harischandra, J. Knuesel, A. Kozlov, A. Bicanski, J.-M. Cabelguen, A. Ijspeert, and O. Ekeberg, "Sensory feedback plays a significant role in generating walking gait and in gait transition in salamanders: a simulation study." *Frontiers in neurorobotics*, vol. 5, no. November, p. 3, Jan. 2011.

[4] X. Zhao, J. Zhang, and C. Qi, "CPG and Reflexes Combined Adaptive Walking Control for AIBO," *2012 11th International Conference on Machine Learning and Applications*, pp. 448–453, Dec. 2012.

[5] S. Degallier, L. Righetti, L. Natale, F. Nori, G. Metta, and A. Ijspeert, "A modular bio-inspired architecture for movement generation for the infant-like robot iCub," *2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 795–800, Oct. 2008.

[6] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: learning attractor models for motor behaviors." *Neural computation*, vol. 25, no. 2, pp. 328–73, Feb. 2013.

[7] I. A. Rybak, N. A. Shevtsova, M. Lafreniere-Roula, and D. A. McCrea, "Modelling spinal circuitry involved in locomotor pattern generation: insights from deletions during fictive locomotion." *The Journal of physiology*, vol. 577, no. Pt 2, pp. 617–39, Dec. 2006.

[8] S. Grillner, P. Wallén, K. Saitoh, A. Kozlov, and B. Robertson, "Neural bases of goal-directed locomotion in vertebrates–an overview." *Brain research reviews*, vol. 57, no. 1, pp. 2–12, Jan. 2008.

[9] K. E. Adolph, W. G. Cole, M. Komati, J. S. Garciaguirre, D. Badaly, J. M. Lingeman, G. L. Y. Chan, and R. B. Sotsky, "How do you learn to walk? Thousands of steps and dozens of falls per day." *Psychological science*, vol. 23, no. 11, pp. 1387–94, Jan. 2012.

[10] M. W. Clearfield, C. N. Osborne, and M. Mullen, "Learning by looking: Infants' social looking behavior across the transition from crawling to walking." *Journal of experimental child psychology*, vol. 100, no. 4, pp. 297–307, Aug. 2008.

[11] L. Righetti and A. J. Ijspeert, "Pattern generators with sensory feedback for the control of quadruped locomotion," *2008 IEEE International Conference on Robotics and Automation*, pp. 819–824, May 2008.

[12] J. Nassour, V. Hugel, F. B. Ouezdou, G. Cheng, and S. Member, "Failure Maps : Applied to Humanoid Robot Walking," *IEEE Transactions on Neural Networks*, vol. 24, no. 1, pp. 81–93, 2013.

[13] J. Buchli, L. Righetti, and A. J. Ijspeert, "Frequency analysis with coupled nonlinear oscillators," *Physica D: Nonlinear Phenomena*, vol. 237, no. 13, pp. 1705–1718, Aug. 2008.

[14] L. Righetti, K. A. Nylén, and A. Ijspeert, "Is the locomotion of crawling human infants different from other quadruped mammals?" Tech. Rep., 2008.

[15] C. Li, R. Lowe, and T. Ziemke, "Humanoids Learning to Walk: a Natural CPG-Actor-Critic Architecture," *Frontiers in Neurorobotics*, vol. 7, 2013.

[16] M. Golubitsky and I. Stewart, *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space*, ser. Progress in Mathematics Series. Birkhäuser Basel, 2003.

[17] C. Li, R. Lowe, B. Duran, and T. Ziemke, "Humanoids that crawl: Comparing gait performance of iCub and NAO using a CPG architecture," *2011 IEEE International Conference on Computer Science and Automation Engineering*, pp. 577–582, Jun. 2011.

[18] C. Li, R. Lowe, and T. Ziemke, "Modelling Walking Behaviors Based on CPGs: A Simplified Bio-inspired Architecture," *From Animals to Animats 12 Lecture Notes in Computer Science*, vol. 7426, pp. 156–166, 2012.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*, 2006.

[20] P. Dayan, *Theoretical Neuroscience: Computational And Mathematical Modeling of Neural Systems*. Massachusetts Institute of Technology Press, 2005.

[21] J. Kober, A. Wilhelm, E. Oztop, and J. Peters, "Reinforcement learning to adjust parametrized motor primitives to new situations," *Autonomous Robots*, vol. 33, no. 4, pp. 361–379, Apr. 2012.

[22] J. Peters, "Machine learning for motor skills in robotics," Ph.D. dissertation, 2007.

[23] A. Gams, A. J. Ijspeert, S. Schaal, and J. Lenarčič, "On-line learning and modulation of periodic movements with nonlinear dynamical systems," *Autonomous Robots*, vol. 27, no. 1, pp. 3–23, May 2009.

[24] J. Nakanishi, J. Morimoto, G. Endo, and G. Cheng, "A framework for learning biped locomotion with dynamical movement primitives." in *IEEE-RAS/RSJ International Conference on Humanoid Robots*, no. Humanoids, 2004, pp. 925–940.

[25] R. V. Kail and J. C. Cavanaugh, *Human Development: A Life-span View*, 2012.

[26] M. W. Clearfield, "The role of crawling and walking experience in infant spatial memory." *Journal of experimental child psychology*, vol. 89, no. 3, pp. 214–41, Nov. 2004.

[27] G. Endo, J. Morimoto, T. Matsubara, J. Nakanishi, and G. Cheng, "Learning CPG-based Biped Locomotion with a Policy Gradient Method: Application to a Humanoid Robot," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 213–228, Feb. 2008.

[28] J. Morimoto and J. Nakanishi, "Poincare Map Based reinforcement learning for biped walking," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 2392–2397.

[29] Y. Nakamura, T. Mori, M.-a. Sato, and S. Ishii, "Reinforcement learning for a biped robot based on a CPG-actor-critic method." *Neural networks*, vol. 20, no. 6, pp. 723–35, Aug. 2007.

[30] G. Endo, J. Morimoto, T. Matsubara, J. Nakanishi, and G. Cheng, "Learning CPG-based Biped Locomotion with a Policy Gradient Method: Application to a Humanoid Robot," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 213–228, Feb. 2008.

[31] C. LI, R. Lowe, and T. Ziemke, "Crawling Posture Learning in Humanoid Robots using a Natural-Actor-Critic CPG Architecture," in *Proceedings of European conference on Artificial Life*, 2013.

[32] N. Kohl and P. Stone, "Policy gradient reinforcement learning for fast quadrupedal locomotion," in *Proceedings of the IEEE International Conference on Robotics and Automation*, May 2004.

[33] M. Wiering and M. van Otterlo, *Reinforcement Learning: State-Of-the-Art*, ser. Adaptation, learning and optimization. Springer Berlin Heidelberg, 2012.

[34] S. Kakade, "A Natural policy gradient," *Advances in neural information processing systems*, 2001.

[35] J. Peters and S. Schaal, "Natural Actor-Critic," *Neurocomputing*, vol. 71, no. 7-9, pp. 1180–1190, 2008.

[36] J. Kober and J. Peters, "Policy search for motor primitives in robotics," *Machine Learning*, vol. 84, no. 1-2, pp. 171–203, Nov. 2010.

[37] Wikipedia, "Crawling (human):http://en.wikipedia.org/wiki/Crawling," 2013.

[38] H. van Hasselt and M. a. Wiering, "Reinforcement Learning in Continuous Action Spaces," *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, no. Adprl, pp. 272–279, Apr. 2007.

[39] C. Li, "crawling before learning and after learning: https://www.youtube.com/watch?v=sb0brwcajcw," 2013.

[40] O. Michel, "Webots TM : Professional Mobile Robot Simulation," *Advanced Robotic Systems*, vol. 1, no. 1, pp. 40–43, 2004.

[41] C. Li, "Crawling Nao robot learned by cpg-actor-critic:https://www.youtube.com/watch?v=6CtKSDQV9sw," 2013.

[42] J. Kober, "Learning Motor Skills: From Algorithms to Robot Experiments," Apr. 2012.

[43] E. S. Thelen, *A dynamic systems approach to the development of cognition and action*. MIT Press, 1996.

[44] A. N. Schore, *Affect Regulation and the Origin of the Self: The Neurobiology of Emotional Development*, 1994.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

A Bio-inspired Modular System for Humanoid Posture Control

Vittorio Lippi, Thomas Mergner, Georg Hettich

Abstract—**Bio-inspired sensorimotor control systems may be appealing to roboticists who try to solve problems of multi-DOF humanoids and human-robot interactions. This paper presents a simple posture control concept from neuroscience, called disturbance estimation and compensation, DEC concept [1]. It provides human-like mechanical compliance due to low loop gain, tolerance of time delays, and automatic adjustment to changes in external disturbance scenarios. Its outstanding feature is that it uses feedback of multisensory disturbance estimates rather than 'raw' sensory signals for disturbance compensation. After proof-of-principle tests in 1 and 2 DOF posture control robots, we present here a generalized DEC control module for multi-DOF robots. In the control layout, one DEC module controls one DOF (modular control architecture). Modules of neighboring joints are synergistically inter-connected using vestibular information in combination with joint angle and torque signals. These sensory interconnections allow each module to control the kinematics of the more distal links as if they were a single link. This modular design makes the complexity of the robot control scale linearly with the DOFs and error robustness high compared to monolithic control architectures. The presented concept uses Matlab/Simulink (The MathWorks, Natick, USA) for both, model simulation and robot control and will be available as open library.**

I. Introduction

Postural adjustments (PAs) allow humans to make their voluntary movements smooth and skillful. The adjustments (1) provide the movement buttress that the action-reaction law of physics prescribes, (2) maintain body equilibrium by balancing the body's center of mass (body COM) over the base of support, and (3) cope with interlink coupling torque disturbances from link acceleration (also due to the action-reaction law). The adjustments require coordination across (1)-(3) and between these and the voluntary movements. They participate in the movement and muscle synergies and sensorimotor 'building blocks' [2-5] that help to simplify the complexity given by the high redundancy in the motor system. PA impairment by damage of the cerebellum or sensory systems tends to produce a severely disabling syndrome called ataxia (jerky and dysmetric movements, postural instability; [6]).

There has been recent progress in understanding the neural mechanisms of human postural control [7-10]. This owes to the use of engineering methods that allow relating measured postural responses to exactly known external disturbances in model-based approaches. These models mostly considered human balancing in the sagittal plane, which predominantly occurs around an axis through the ankle joints, and described its biomechanics as that of a single inverted pendulum, SIP.

Among these models, the DEC (disturbance estimation and compensation) model [1, 10] is unique in that it uses sensory-derived internal reconstruction of the external disturbances having impact on the body posture. Model simulation data for various disturbance scenarios and changes in sensor availability were in good agreement with human data [10-14]. The model was re-embodied into a SIP postural control robot [15] and the robot was successfully tested in the human test bed [14].

Further development of the DEC model comprised its extension to double inverted pendulum (DIP) biomechanics with hip and ankle joints and an investigation of the neural control underlying the coordination between these two joints [16]. This work involved a double inverted pendulum, DIP postural control robot. Furthermore, feasibility tests with a 4 DOF agent involving Matlab's SimMechanics toolbox were successful. This led to the here presented generalized DEC module for the control of multi-DOF robots with a modular control architecture. Using one DEC module for each DOF, control complexity linearly increases with the number of DOF.

The next section gives an overview of the DEC control principles, followed by a description of the implementation in a SIP and the generalized modular control of a multi-DOF DEC system. Finally, the DEC library is briefly described and demonstrated by presenting an application. In Conclusion, outstanding results are emphasized and future improvements are outlined.

II. System Overview

A. The DEC concept

Figure 1 shows a simplified scheme of the DEC module as it was developed for the SIP control. The module controls joint position of a moving link with respect to a supporting link and consists of three parts:

(A) *Proprioceptive negative feedback loop of joint angular position* (box 'Prop.'). A PID (proportional, integral, derivative) controller provides the torque command ($P \approx m \cdot g \cdot h$; m, body mass; h, center of mass= COM height; g, gravitational acceleration). The neural time delay of this loop amounts to ≈60 ms (ankle joint).

(B) *Intrinsic stiffness and damping loop of musculoskeletal system* ('passive stiffness' in box '*Biomech'; it* amounts to ≈15% of reflexive stiffness and damping of (A) and

Manuscript received September 6, 2013. Supported by the European
All authors are with the Neurology Department, Freiburg University Clinics, 79106 Freiburg, Germany.
V. L. Author is corresponding author (phone: +49-0761-27052280; fax: +49-0761-27053100; e-mail: Vittorio.lippi@uniklinik-freiburg.de).
T. M. Author (e-mail: mergner@uni-freiburg.de).
G. H. Author (e-mail: georg.hettich@uniklinik-freiburg.de).

feedback gain of (A) and (B) together is unity). Time delays are virtually zero. (A) and (B) together form a servo that, given appropriate control parameters, actuates the joint such that actual joint position equals the desired position (input is displacement trajectory via the Set Point Signal).

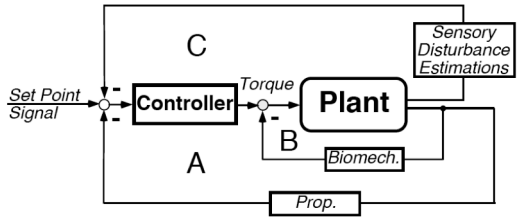


Fig. 1. Simplified scheme of DEC module.

(C) *Disturbance estimation and compensation (DEC) loop*. The DEC loops (four in a complete scheme, see below) estimate the external disturbances through sensor fusions and command via negative feedback the servo (A, B) to produce the joint torques that compensate for the disturbances. Assuming ideal compensation, the servo can function as if there were no disturbances. Note furthermore that no feed forward of plant dynamics is used for the servo (e.g. through an inverse of plant dynamics). Sensory information (mainly vestibular) from the DEC loops upgrades the servo from joint coordinates to space coordinates. Identified lumped human time delay across all three loops (A-C) measured at the controller for the ankle joint amounts to ≈180ms, with the largest share from (C) [9,10].

In the SIP used for the DEC control, the joint (ankle joint) connects the moving link (body) with the supporting link (foot). Postural movements tend to be rather slow, such that centrifugal and Coriolis forces can be neglected. Postural stability is achieved by the DEC feedback for any desired possible joint position, allowing the superposition of voluntary movements with the compensation of external disturbances. The many external events that may have a mechanical impact on body stability are decomposed in, and estimated as external disturbances. Underling these estimations are sensory mechanisms.

Humans use multisensory integration of vestibular signals, vision, touch, and joint proprioception (angle, angular velocity, and force/torque) for their postural control [17]. Studies on human self-motion perception [18] and animal work on sensory processing [19] showed that the central nervous system internally processes physical variables that are not directly available from the sensory organs, but result from sensor fusions.

For example, humans distinguish in the absence of external spatial orientation cues (visual, auditory, and haptic) between body rotation (velocity and position), orientation with respect to the earth vertical, and linear acceleration. They do so by combining input from peripheral vestibular receptor organs (canals, otoliths; see [14]). Furthermore, they may use estimates of the variables for controlling body segments that are distant from the sensor organ in the body. For example, humans may perceive trunk-in-space motion by combining a vestibular head-in-space motion signals with a proprioceptive trunk-to-head (neck) motion signal [18]. Transfer of the space reference also may apply to the other vestibular signals and may be applied to other body segments and even extended to external items that are in firm haptic contact with the body (e.g. the support surface when standing).

The external disturbances and their estimates can be considered from two viewpoints. First, they reflect outside world events that tend to affect the joint torque in certain conditions (e.g. while standing). These events occur in world coordinates and usually in a context dependent way (e.g. ride on "this especially fast escalator"). Corresponding predictions of these estimates may later be re-called from memory and fed forward to the estimation mechanisms where they are fused with the sensory-derived estimates. According to the DEC concept, also self-produced disturbances, such as the gravity effect during voluntary body lean, entail fusions of predicted and sensory derived estimates [1]. The second aspect is that the disturbances affect body stability via the joint torque they produce. The corresponding torque components are referred to as *disturbance torques*. Both aspects will be considered in the next section (III).

The disturbance torques in the SIP scenario are part of the ankle torque

$$T_{\mathrm{A}} = J \cdot \frac{d^2\alpha_{\mathrm{BS}}(t)}{dt^2} \quad , \qquad (1)$$

where J represents the body's moment of inertia about the ankle joint (not including the feet) and α_{BS} the body-space angle (primary position: COM projection on ankle joint be vertical, $\alpha_{\mathrm{BS}} = 0°$). In the absence of any disturbance, T_{A} equals the actively produced muscle torque, T_{a}. The disturbance torques add to T_{a} in the form

$$T_A = (T_g + T_{in} + T_{ext} + T_p) + T_a \quad , \qquad (2)$$

where T_{g} is the gravitational torque, T_{in} the inertial torque, T_{ext} the external torque, and T_{p} the passive joint torque. T_{g}, T_{in}, T_{ext}, and T_{p} challenge the control of T_{A} (exerted by T_{a}) and are compensated for by T_{a} [1]. While T_{p} represents an intrinsic musculoskeletal property, T_{g}, T_{in}, and T_{ext} are produced by the neural feedback.

The following section explains the DEC loops in two steps. First, an explanation is given for the simple case of the SIP balancing about the ankle joints. Then, the generalized form for modular control of multi-DOF systems will be presented. Tables I and II give our designations of the DEC module's inputs and outputs, respectively.

III. Disturbance Estimations

Four physical *external* disturbances need to be taken into account for posture control:

(i) Support surface rotation (platform tilt)
(ii) Gravity and other field forces
(iii) Support surface translation (external acceleration)
(iv) Contact forces (external torque)

(I) DEC OF SUPPORT SURFACE ROTATION

SIP scenario. The support surface tilt produces the foot-space excursion α_{FS} (primary position, level; α_{FS}= 0°). Body inertia tends to maintain the primary body-space orientation, which is upright (primary body-foot angle $\alpha_{BF} \approx \alpha_{BS}$ = 0°). On the other hand, the servo tends to take the body with the platform in relation to α_{BF}. This applies to the passive ankle torque T_p, in the form

$$T_p = -K_{P'} \cdot \alpha_{BF} - K_{D'} \cdot \frac{d\alpha_{BF}(t)}{dt} \quad , \quad (3)$$

with $K_{P'}$ representing the passive stiffness (proportional) factor and $K_{D'}$ the passive damping factor. The reflexive part of the servo, loop (A), is commanded by the estimate of α_{FS}, $\hat{\alpha}_{FS}$, to maintain the body orientation upright.

To this end, $\hat{\alpha}_{FS}$ combines vestibular and proprioceptive information by a *down channeling* of the vestibular derived space reference from the body to the feet. According to [10], humans achieve this by using the derivatives of α_{BS} and α_{BF} in the form

$$\dot{\alpha}_{FS} = \dot{\alpha}_{BS} - \dot{\alpha}_{BF} \quad . \quad (4)$$

According to [10], there is a subsequent processing of the estimate by a velocity threshold (0.18°/s; with level support, $\dot{\alpha}_{FS}$ tends to be subthreshold, and its noise, mainly from the $\dot{\alpha}_{BS}$ signal, is prevented from entering the control [14]), a scaling factor (G=0.75), and a mathematical integration.

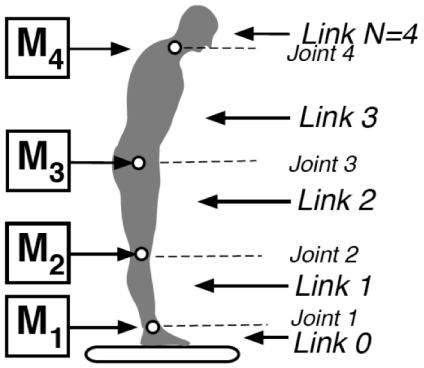


Fig. 2. Conventions used for labeling links, joints and control modules (M).

Implemented in the SIP control, $\hat{\alpha}_{FS}$ 'upgrades' the control from joint to space coordinates.

Generalized case. For the nth link in a multi-DOF system (Fig. 2), the link orientation in space is given by α_n^{SPACE}. This information is obtained from vestibular input that is down-channeled analogous to (4) through the fusion of vestibular and proprioceptive signals by

$$\alpha_n^{SPACE} = \alpha_{n+1}^{SPACE} - \alpha_{n-1}^{JOINT} \quad . \quad (5)$$

The down-channeling proceeds from the upper most segment α_{HEAD}^{SPACE} that contains the vestibular organs. The tilt of the lowest link in the system (most often the foot), which provides the support for the upper links, is given by

$$\alpha_0^{SPACE} = \alpha_{HEAD}^{SPACE} - \sum_{k=1}^{N} \alpha_k^{JOINT} \quad . \quad (6)$$

Recent evidence from humans [20] suggests that, while the down-channeling to the supporting link occurs through velocity signals as in (4), the further processing in terms of thresholding (dead band discontinuity), gain scaling and integration is performed for the lowest link α_0^{SPACE}, from where the tilt estimate is then up channeled for controlling the tilt responses of the upper links. The position of the common *COM* of all the links above the respective joint is calculated (see below). This accounts for the fact that the location of the COM above the joint may change when the configuration of the links changes. The control of each joint can now be viewed as if dealing with a SIP.

(II) DEC OF GRAVITY DISTURBANCE

SIP scenario. Body lean evokes the gravitational torque T_g that is related to α_{BS} by

$$T_g = m \cdot g \cdot h \cdot \sin(\alpha_{BS}) \quad . \quad (7).$$

The estimate $\hat{T}_g$ uses a vestibular derived α_{BS} signal and includes a detection threshold and gain scaling. For use of $\hat{T}_g$ in the DEC feedback loop, small angle approximation reduces (7) to $T_g = m \cdot g \cdot h \cdot \alpha_{BS}$. Furthermore, $\hat{T}_g$ is divided by $m \cdot g \cdot h$ to obtain an angle equivalent of the torque. As an alternative to use $\hat{T}_g$ one may use directly α_{BS} (in the form of an estimate $\hat{\alpha}_{BS}$).

Note that during support surface tilt with compensation of $\hat{\alpha}_{BS}$ (or $\hat{T}_g$) alone, the body is tilted with the platform; it is the compensation of $\hat{\alpha}_{FS}$ that maintains the body upright.

Generalized case. In a DIP or multi-DOF body, τ_g is calculated by

$$\tau_g = m_n^{UP} \, g \, CoM_{nx} \quad , \quad (8)$$

where CoM_{nx} is the horizontal component of the position of the center of mass $\boldsymbol{CoM}_n$ of all the segments above the controlled joint. CoM_n is computed performing the weighted average

$$\boldsymbol{CoM}_n = [\boldsymbol{CoM}_{n+1} + L_n \cos(\alpha_n^{SPACE})] m_{n-1}^{UP} + m_n h_n \cos(\alpha_n^{SPACE}) \quad , \quad (9)$$

where L_n is the length of the link controlled by the joint and h_n is the distance of the COM of the nth link from the nth joint.

Analogous to (I), the gravity compensation in each joint comprises all links above this joint, as if dealing with a SIP.

(III) DEC OF SUPPORT LINEAR ACCELERATION

SIP scenario. Human perception of support surface acceleration may involve various sensory systems and may include sensing of shear forces in the foot soles. However, the corresponding biological knowledge base is still limited. Reference [1] used vestibular information to estimate support surface acceleration, having in mind that vestibular-loss subjects have major problems during such stimulus conditions in the absence of external orientation cues. Support surface acceleration evokes the disturbance torque T_{in} in the form

$$T_{in}' = - \hat{a}_{FS} \cdot m \cdot h \cdot \cos(\alpha_{BS}) \quad , \quad (10)$$

where $\hat{a}_{FS}$ is the estimate of support surface acceleration. $\hat{a}_{FS}$ can be computed from the difference between two vestibular

signals, the one of head linear acceleration and the one of the head acceleration due to body rotation (also derived from vestibular input; see [1]. The processing for $\hat{a}_{FS}$ also comprises thresholding and gain scaling.

Generalized case. In the case of a multi-DOF system, the external acceleration is computed for each joint. Analogous to the SIP scenario, the part of the vestibular head acceleration signal that is not explained by trunk rotation at the hip or at any joint below is taken to stem from support surface acceleration. This is expressed as

$$\boldsymbol{a}^{EXTERNAL}=\boldsymbol{a}^{VESTIBULAR}-\boldsymbol{a}_n^{SELF} \quad , \quad (11)$$

where the acceleration produced by joint movements is:

$$\boldsymbol{a}_n^{SELF} = \boldsymbol{a}_{n+1}^{SELF} + L_n \frac{d^2}{dt^2}\begin{bmatrix}\sin(\alpha_n^{SPACE})\\ \cos(\alpha_n^{SPACE})\end{bmatrix} \quad . \quad (12)$$

The disturbance torque then results from

$$\tau_{acc}=a_x^{EXTERNAL} CoM_{ny} m_n^{UP} + a_y^{EXTERNAL} CoM_{nx} m_n^{UP}. \quad (13)$$

TABLE I.
MODULE INPUTS

Symbol	Description	Source
Control signal	Desired α_n^{SPACE}, Com_n or α_n^{JOINT} (3 options)	Desired position
$\boldsymbol{CoM}_{n+1}$	Center of mass of the robot over the n+1th joint	n+1th module
m_{n+1}^{UP}	Mass of the robot from head to the n+1th joint	n+1th module
α_{n-1}^{SPACE}	Up-channeled α_{n-1}^{SPACE}	n-1th module
$\alpha_n^{SPACE-DOWN}$	Down-channeled α_n^{SPACE}	n+1th module
J_{n+1}^{UP}	Moment of inertia of the robot from head to the nth joint	n+1th module
$\boldsymbol{a}_{n+1}^{ang}$	Head angular acceleration with respect to the joint n+1th	n+1th module

(IV) DEC OF CONTACT FORCE DISTURBANCE

SIP scenario. Led the disturbance torque T_{ext} of equation (2) be the results of a horizontal force F_{ext} exerted on the body by a pull on the clothes at the height *h,* which is above the COM (such that foot-support shear forces may be neglected). T_{ext} is then related to F_{ext} and T_{BS} in the form

$$T_{ext}= F_{ext} \cdot h \cdot \cos(\alpha_{BS}) \quad . \quad (14)$$

An estimate of the external disturbance may be obtained from sensing the amount and location at the body of F_{ext} (and α_{BS}). However, having in mind that humans tend to sense centre of pressure (COP) shifts under the feet during such stimuli, studies from our laboratory [1,10,11,13] derived the estimate $\hat{\boldsymbol{T}}_{ext}$ from a sensory measure of COP, as represented in T_a, in the form of

$$T_{ext} = T_a - (T_g + T_{in} + T_p - T_A) \quad , \quad (15)$$

accounting for T_A, T_g, and T_{in} by equations (1), (7), and (10), respectively, and neglecting T_p, because it is relatively small. Processing of the $\hat{\boldsymbol{T}}_{ext}$ estimate includes again a detection threshold and, because T_a provides positive feedback, a gain clearly <1 and a low-pass filtering. Humans appear to restrict the use of $\hat{\boldsymbol{T}}_{ext}$ to situations where the contact force stimulus endangers postural stability [13]. Possibly, co-contraction of antagonistic muscles across the involved joints may additionally help T_{ext} compensation as long as the COP shift does not exceed the base of support.

TABLE II
MODULE OUTPUTS

Symbol	Description	Destination
τ_n	Torque produced in the joint n	Joint servo loop
$\boldsymbol{CoM}_n$	Center of mass of the robot over the nth joint	n-1th module
m_n^{UP}	Mass of the robot from head to the nth joint	n-1th module
α_n^{SPACE}	Up-channeled α_n^{SPACE}	n+1th module
$\alpha_n^{SPACE-DOWN}$	Down-channeled α_n^{SPACE}	n-1th module
J_n^{UP}	Moment of inertia of the robot from head to the nth joint	n-1th module
$\boldsymbol{a}_n^{ang}$	Head angular acceleration with respect to the nth joint	n-1th module

Generalized case. For convenience, the subscripts in (15) were changed to superscripts, allowing to denote the number of the module by the subscript. With this modification, (15) takes the form

$$\boldsymbol{\tau}_{ext}= \boldsymbol{\tau}_n^{a} - \boldsymbol{\tau}_n^{g} - \boldsymbol{\tau}_n^{in} - \boldsymbol{\tau}_n^{p} + \boldsymbol{\tau}_n^{A} \quad . \quad (16)$$

With the term J_n^{UP} representing the moment of inertia of all the segments over the controlled joint, equation (1) takes the form

$$\boldsymbol{\tau}_n^{A} = \frac{d}{dt}(\dot{\boldsymbol{\alpha}}_n^{SPACE} J_n^{UP}) \quad , \quad (17)$$

which takes into account that also J_n^{UP} may change in the case of a robot with several DOFs.

In order to keep the computation of J_n^{UP} as simple as possible by exchanging only one variable between blocks, the moment of inertia J_n^{UP*} is computed around the axis passing through the center of mass of the whole group of segments from the n^{th} segment to the *head* in the form

$$J_n^{UP*}=(J_{n+1}^{UP*} + m_{n+1}^{UP}||\boldsymbol{CoM}_{n+1}-\boldsymbol{CoM}_n||^2)+ \\ + J_n+m_n||\boldsymbol{CoM}_n^{LINK}-\boldsymbol{CoM}_n||^2 \quad , \quad (18)$$

where $\boldsymbol{CoM}_n^{LINK}$ is the center of mass of the n^{th} link equal to

$$\boldsymbol{CoM}_n^{LINK} = l_n\begin{bmatrix}\sin(\alpha_n^{SPACE})\\ \cos(\alpha_n^{SPACE})\end{bmatrix} \quad . \quad (19)$$

J_n^{UP*} is then down-channeled to the n-1^{nt} block, while J_n^{UP}, used in (17), is computed as

$$J_n^{UP}=J_n^{UP*} + m_n^{UP}||\boldsymbol{CoM}_n||^2 \quad . \quad (20)$$

IV. SOFTWARE LIBRARY

The software library consists of a Matlab/Simulink block that implements a single DEC module containing the servo and the above described disturbance estimates. An interactive mask allows the user to specify the block position in a multi-link system (as top or bottom/supporting, or intermediate link). The physical features such as link height, link mass, height of link COM, etc. are input into the block as anthropometric parameters. Each block has input and output ports to exchange data with the neighboring blocks above and below (see Tables I and II) and allows modifying the processing of the estimates in terms of thresholding, gain scaling, etc. The PID controller parameters and the passive stiffness and damping parameters of the servo are set after deciding the target variables (COM-joint orientation of the above link, or link orientation with respect to earth vertical, or joint angle).

The software will be made available as an open library in the internet.

V. CASE STUDY

Simulations of the modular control concept were performed with a four link humanoid agent in Matlab/Simulink. The links were feet (fixed to the ground), shank, thigh and trunk (here HAT; head, arms and torso), interconnected by the ankle joint, the knee joint and the hip joint, respectively. Stimuli were applied in the sagittal plane,

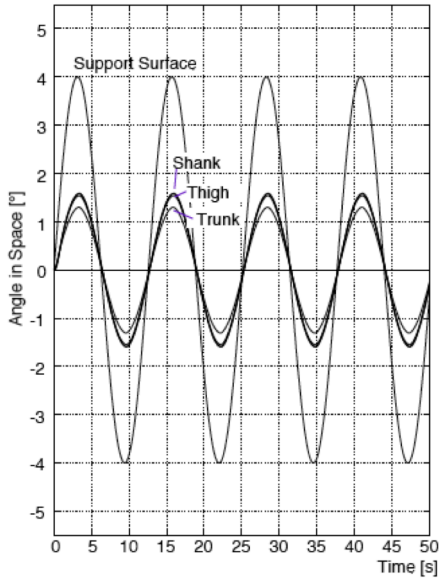


Fig. 3. Responses of the simulated four link humanoid agent to sinusoidal support surface rotation (0.08 Hz). The combined action of the gravitational torque compensation, support tilt compensation and support linear acceleration compensation tends to maintain the body upright, but this only partially due to human-like compensation gains (<1).

which allowed us to use a planar triple inverted pendulum biomechanical model. The humanoid's biomechanical parameters corresponded to human anthropometric measures [21]. The above-described DEC modules were used to control the ankle, knee and hip joints in a modular way. The control parameters were adapted from [20]. Accounting for the fact that humans tend to stiffen the knee joints during our tests (see below), a high level of passive stiffness was used for this joint.

Two experiments were performed. In the first experiment, the agent balanced sinusoidal ±4° support surface tilts (Fig. 3). In this test, the agent used the gravitational torque compensation and the support surface tilt compensation for controlling the ankle joint, and these two compensation together with the support surface acceleration compensation for controlling the knee and the hip joints. Compensation gains were set to human-like values (<1). This entailed the under-compensation shown in Fig. 3, with the trunk being compensated slightly better than the two leg segments.

In the second experiment, the agent performed a voluntary forward trunk lean of 4° (Fig. 4). This experiment tested whether the control would produce the human hip-ankle coordination, which consists of a compensatory backward lean of the leg segments such that the body COM is maintained over the ankle joint (see [20]). Additionally, this

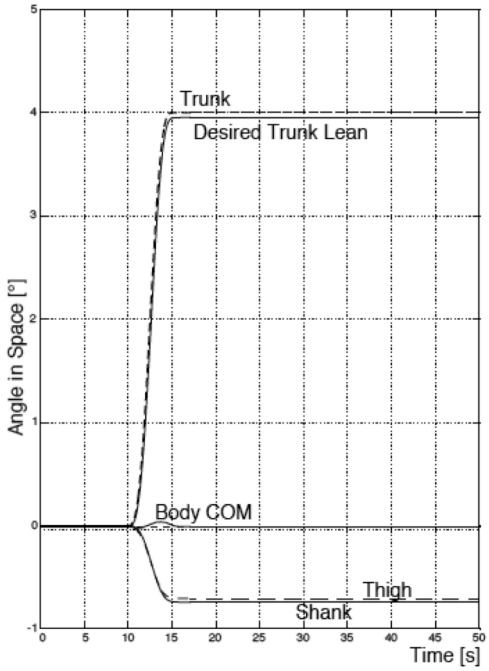


Fig. 4. Behavior of the four link humanoid agent performing a voluntary forward trunk lean in the hip joint. Note that the leg links Thigh and Shank are leaning backwards, so that the Body COM remains above the ankle joint. This ankle-hip coordination automatically arises from the interaction between the hip and ankle control modules and the agent's biomechanics.

coordination neutralizes most of the coupling torques exerted by the trunk bending on the leg segments. The voluntary movement is associated with predictions of the disturbance estimates, which have unity gain, and are fused with the sensory-derived disturbance estimates (see [1]). This explains the almost perfect performance in Fig. 4.

VI. CONCLUSION

Upgrading the DEC concept from using one module in the SIP control to the modular control architecture in a multi-

DOF system is possible, because the DEC concept allows controlling each joint as if it were dealing with a SIP. In particular:

(I) The support surface tilt DEC estimates take the rotation of each supporting link as a tilt disturbance for the upper links. Controlled is the orientation of the links' COM above the supporting joint with respect to earth vertical.

(II) The gravity DEC estimates compensate for the gravitational torque produced by the upper links' total COM.

(III) The support surface linear acceleration DEC estimates compensate for the acceleration effect occurring at the top of any supporting link. The effect is produced as joint torque by the inertial force of all upper links. Noticeably, this compensation also includes up-going coupling forces effects (concerning down-going coupling forces, see [20]).

(IV) The contact force DEC estimates compensate for the evoked torque in the supporting joint, taking into account the moment of inertia of the above (supported) links.

The present concept attributes postural responses to unforeseen external disturbances to sensory mechanisms and feedback. A sensory network of down- and up-going spinal pathways from the brainstem and back to it and to higher CNS centers (e.g. cerebellum) carrying vestibular signals and receiving spinal proprioceptive input has been demonstrated in animal work [22]. Principles of how predictions of disturbance estimates, centrally derived and fed forward during voluntary movements, may be fused with sensor-derived disturbance estimates have been suggested in Reference [1]. Thresholding and gain scaling of estimates have been attributed to sensory noise in Reference [14].

Human movement coordination such as the hip-ankle coordination occurring during voluntary trunk bending (Fig. 4) or balancing of support surface tilt [20] may emerge as automatic 'postural adjustments' from DEC mechanisms. Given moderate disturbances and full foot support, ankle and hip responses fulfill the criteria of the human ankle and hip strategy (see [17]).

In a previous study, we compared the DEC concept with the classical control approach that uses extended observers for disturbance estimation [23]. This solution worked in simulations, but had problems to deal with inaccuracies of sensors and actuation when implemented into the robot. A later solution that included the vestibular system into a standard engineering approach was more successful in terms of stability, but not in terms of human-like responses [24].

Empirically, model simulations and experiments with the robot demonstrated stability of the system [20]. A mathematical generalized demonstration is beyond the scope of the present paper and will be postponed to a specific treatment.

Current work on DEC tries to include visual information into the sensory fusions, to combine sagittal and frontal plane DEC modules, to deal with the four-bar linkage of biped stance (2 legs-ground-pelvis) in the frontal plane, to deal with the shifting of body weight between legs during walking, and to explore further examples of human movement coordination.

REFERENCES

[1] T. Mergner, "A neurological view on reactive human stance control," *Annu Rev Control*, vol. 34, pp. 177–198, 2010.

[2] M. C. Tresch, P. Saltiel, and E. Bizzi, "The construction of movement by the spinal cord," *Nature Neuroscience*, vol. 2, 162–167, 1999.

[3] Y. P. Ivanenko, G. Cappellini, N. Dominici, R. E. Poppele, and F. Lacquaniti, "Coordination of locomotion with voluntary movements in humans," *Journal of Neuroscience.* vol. 25, pp. 7238–7253, 2005.

[4] L. H. Ting, "Dimensional reduction in sensorimotor systems: a framework for understanding muscle coordination of posture," *Progress in Brain Research*, vol. 165, pp. 301–325, 2007.

[5] A. d'Avella and D. K. Pai, "Modularity for sensorimotor control: evidence and a new prediction. *Journal of Motor Behavior*, vol. 42, pp. 361-369, 2010.

[6] A. J. Bastian, "Mechanisms of ataxia," *Physical Therapy*, vol. 77, pp. 672-675, 1997.

[7] R. Johansson and M. Magnusson, "Human postural dynamics," *CRC Crit. Rev. Biomed. Eng*., vol. 18, pp. 413-437, 1991.

[8] H. van der Kooij, R. Jacobs, B. Koopman, and H. Grootenboer, "A multisensory integration model of human stance control," *Biological Cybernetics*, vol. 80, pp. 299–308, 1999.

[9] R. J. Peterka, "Sensorimotor integration in human postural control," *Journal of Neurophysiology*, vol. 88, pp. 1097–1118, 2002.

[10] C. Maurer, T. Mergner, and R. J. Peterka, "Multisensory control of human upright stance," *Experimental Brain Research*, vol. 171, pp. 231–250, 2006.

[11] T. Mergner, C. Maurer, and R. J. Peterka "A multisensory posture control model of human upright stance," *Progress in Brain Research*, vol. 142, pp. 189-201, 2003.

[12] G. Schweigart and T. Mergner, "Human stance control beyond steady state response and inverted pendulum simplification. *Experimental Brain Research*, vol. 185, pp. 635-653, 2008.

[13] C. Cnyrim, T. Mergner, and C. Maurer, "Potential role of force cues in human stance control," *Experimental Brain Research*, vol. 194, pp. 419–433, 2009.

[14] T. Mergner, G. Schweigart, and L. Fennell, "Vestibular humanoid postural control," *Journal of Physiology - Paris*, vol. 103, pp. 178–194, 2009.

[15] T. Mergner, F. Huethe, C. Maurer, and C. Ament, C, "Human equilibrium control principles implemented into a biped robot," In T. Zielinska and C. Zielinski (eds.) *Robot Design, Dynamics, and Control* (Romansy 16), CISM Courses and Lectures, vol. 487, pp. 271-279, 2006.

[16] G. Hettich, T. Mergner, and L. Fennel, "Double inverted pendulum model of reactive human stance control." In: Multibody Dynamics 2011, ECCOMAS Thematic Conference., 2011. Brussels, Belgium, 4-7 July 2011. (http://www.posturob.uniklinik-freiburg.de).

[17] B. Horak and J.M. Macpherson, "Postural orientation and equilibrium', in L. Rowell and J. Shepherd (ed.) Handbook of Physiology, New York: Oxford University Press, 1996, pp 255-292.

[18] T. Mergner, W. Huber, and W. Becker, "Vestibular-neck interaction and transformations of sensory coordinates," *Journal of Vestibular Research*, vol. 7, pp. 119–135, 1997.

[19] G. Bosco and R.E. Poppele, "Representation of multiple kinematic parameters of the cat hindlimb in spinocerebellar activity," *Journal of Neurophysiology*, vol. 78, pp.1421–1432, 1997.

[20] G. Hettich, L. Assländer, A. Gollhofer, and T. Mergner, "Human hip-ankle coordination emerging from multisensory feedback control," Submitted.

[21] D. A. Winter, *Biomechanics and motor control of human movement. (2nd ed.)*. New York: Wiley, 1990.

[22] J. D. Coulter, T. Mergner, and O. Pompeiano, "Effects of static tilt on cervical spinoreticular tract neurons," *Journal of Neurophysiology*, vol. 39, pp. 45-62, 1976.

[23] K.A. Tahboub and T. Mergner, "Biological and engineering approaches to human postural control," *Journal of Integrated Computer Aided Engineering*, vol. 14, pp. 15-31, 2007.

[24] K.A. Tahboub, "Biologically-inspired humanoid postural control," *Journal of Physiology - Paris*, vol. 103, pp. 195–210, 2009.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Zero-calibration BMIs for sequential tasks using error-related potentials

Jonathan Grizou*, Iñaki Iturrate†, Luis Montesano†, Manuel Lopes*, Pierre-Yves Oudeyer*
* INRIA Bordeaux Sud-Ouest, France
{jonathan.grizou, manuel.lopes, pierre-yves.oudeyer}@inria.fr
†Departamento de Informática e Ingeniería de Sistemas (DIIS), Universidad de Zaragoza, Spain
{iturrate, montesano}@unizar.es

***Abstract*—Do we need to explicitly calibrate Brain Machine Interfaces (BMIs)? Can we start controlling a device without telling this device how to interpret brain signals? Can we learn how to communicate with a human user through practical interaction? It sounds like an ill posed problem, how can we control a device if such device does not know what our signals mean? This paper argues and present empirical results showing that, under specific but realistic conditions, this problem can be solved. We show that a signal decoder can be learnt automatically and online by the system under the assumption that both, human and machine, share the same a priori on the possible signals' meanings and the possible tasks the user may want the device to achieve. We present results from online experiments on a Brain Computer Interface (BCI) and a Human Robot Interaction (HRI) scenario.**

I. Motivation

EEG-based brain-machine interfaces (BMIs) (see [1] for a review) provide a communication channel between humans and machines using only brain activity. Since the first demonstration of brain-controlled devices, research on BMIs has emerged as one of the most growing research whose ultimate goal is to endow people with severe motor disabilities with communication and control capacities. Yet, the practical promise of this technology remains unfilled with BMIs remaining confined to the laboratory or limited to clinical studies or home demonstrations that require close technical oversight [2], [3].

Among the existent signals used to develop a BMI, recent works have shown that it is possible to decode information related to human error processing, namely the error-related potentials [4] appearing for instance when the device action does not match the user's expectations. This potential has been used mainly to improve the BMI decoder [5] and, more interestingly, as feedback information to solve sequential tasks [6], [7]. In fact, they can be interpreted as instructions given by a teacher to a learner under the implicit assumption that the two agents share a mutual understanding of feedback meaning (e.g. a decoder that translates raw signals into feedback such as speech into words).

In practice, BMI solves the meaning problem using an open-loop calibration procedure to train a decoder in a supervised manner. This calibration phase hinders the deployment of out-of-the-lab applications [1], due to the need of a specific calibration for each task and session. However, this phase is required mainly due to the non-stationary nature of the EEG [8]; the large intra- and inter-subject variability [9], and variations induced by the task [10].

It is worth noting that this problem can be generalized to many kind of human-machine interaction (HMI) scenarios. In modern human-robot collaboration, there is a need for intuitive interfaces that allow non-technical users to teach robots. A major obstacle is that of requiring a pre-defined set of instruction signals. As a result, in the human to robot interaction (HRI) community, an important part of the work consists of building classifiers to translate human communicative signals (speech, gestures, facial expression) to symbolic meanings understandable by the robot. Such procedure requires a costly offline gathering of signals. A machine able to automatically understand such symbols could improve the usability and ease of use of such interactive system and even make use of involuntary signals, e.g. prosody, to better exploit the information provided by the user.

While research on robot learning from human interaction has flourished in the last ten years [11], most work has focused on how to extract statistical *task models* from human teachers following a fixed pre-defined teaching protocol. Thus, a usual assumption is that the learner and the teacher share a mutual understanding of the meaning of each others' signals. The question of how a robot can learn to interpret personalized and potentially noisy teaching signals, i.e. learn *instruction models*, has been much less explored. In a preliminary work [12], we presented a computational approach addressing this problem by considering a finite space of symbolic teaching signals in simulation while bootstrapping the system with known symbols. Later [13], we released the need for bootstrapping and allow the teacher to use signals that can be represented as fixed length feature vectors, which is better suited for HRI scenarios.

Interestingly, one similar approache have been developed in the BMI community. For non-invasive P300 signals, Kindermans et al. proposed a method to learn from scratch in closed loop a decoder by exploiting multiple stimulations and prior information [14], [15]. However, the approach needed for a warm-up period.

In this paper we present a different approach to calibration procedures. We argue that, for some problems, a calibration procedure is not explicitly needed. The system could learn, by interacting with the user, both what to do and how to map brain signals into meaningful instructions. The innovation of this work is not about new machine learning development

but resides in a new way to use them. Our method combines and exploits two sources of information: task constraints and spacial organization of instruction signals in the feature space. We report results from a BCI online experiments and a HRI scenario. The results show that the proposed method is able to learn good instruction models while solving the task efficiently without any prior calibration procedure.

II. Problem

This section describes the problem of executing a task when the instruction signal's meaning is unknown or uncertain. For practical reasons, we will only refer to feedback driven BMI scenarios for the description of the problem and method. Nonetheless, the idea presented here holds for other HRI scenarios by replacing brain signals with other feedback instructions (e.g., speech). BMI control based on feedback signals (illustrated in figure 1) differs from classical brain-machine interfaces in the sense that the user does not actively deliver commands to the device, but only delivers feedback about actions performed by the device. In this setting, the device needs to actively execute actions to solve the tasks and to be able to learn an intelligent behavior from the feedback. This idea can be seen as a shared-control strategy [16], where both the user and the device help each other to solve a task.

Fig. 1: In this BMI scenario, the user is watching the agent moving on the screen and assess the agent actions with respect to its own objective illustrated by the red state.

Essentially, this BMI control follows an iterative sequential process where, in a particular state s, the device performs an action a and the user assesses the action using brain signal $e \in \mathbf{R}^n$. These assessments generate error-related potentials, i.e. signals elicited in the brain when the outcome observed by the user differs from the expected one [4]. Thus, collected data are in the form $\{(e_i, s_i, a_i),\ i = 1, \dots, N\}$, i.e. a sequence of states, actions and teaching signals triplets, with N the number of steps. In a common scenario, the system has been fed with a classifier, parameterized by θ, that translates brain signals e into symbolic instructions z that belong to one of two classes (correct or incorrect assessment), $z \in \{c, w\}$. The model parameters θ are in practice learnt using a calibration procedure. The device then learns from this symbolic feedback.

This control can be exemplified for a reaching task (Fig. 1), where the user wants to reach a target position unknown by the system. The device performs several discrete actions (e.g. moving left or right), and learns from the feedback given by the user. After several steps, the device knows which is the desired target and how to reach it. However, the control can become intractable as the task complexity increases. Furthermore, only binary feedback is available and there is a large percentage of misdetected assessments. Given a set of possible tasks $\Xi = \{\xi_1, \dots, \xi_T\}$, with T the number of task hypothesis, it is possible to speed up the inference by precomputing the agent's optimal behavior $\pi_\xi = p(a|s, \xi)$ for each task and using the feedback signal as a likelihood for the task. For instance, in the previous example, the possible tasks are given by the number of targets. This way, an error (negative feedback) after a particular action will decrease the posterior probability of those targets whose optimal policy agrees with the action.

In this work, we address the problem of removing the need for calibration. Therefore we do not have access to the negative or positive nature $z \in \{c, w\}$ of the brain signals e beforehand. We propose an algorithm that simultaneously calibrates the feedback decoder and executes in closed loop a sequential task only known by the user $\hat{\xi} \in \Xi$. Our method combines and exploits two sources of information: task constraints, namely optimal policies π_ξ, and spatial organization of brain signals in the feature space. The underlying assumptions are:

1) a finite number of possible task hypothesis, i.e. ξ_l for $l \in 1, \dots, \xi_T$, can be defined 2) the inputs signals have some hidden labels z corresponding to their meaning 3) the set of possible meanings is finite, e.g. $z \in \{c, w\}$ 4) given the ground truth labels $\hat{z}$ of the signals, a classifier of sufficient accuracy could be trained to control the device.

These assumptions may look constraining but are actually common ones in most current BMI scenarios. For example, consider the case of a robotic arm assistant helping to grasp objects on a table. Such robotic assistant could be controlled by a user assessing the robot actions. For instance, the robot could start reaching for an object having the user validating or not the decision of the robot. In such scenario, the usual method would be to start a calibration procedure to map ERP brain signals into symbolic feedback instructions for the robot (correct and incorrect). Once enough data are collected, a classifier would be trained and we could start assessing the robot's actions using brain ERP. This simple scenario, which follows a calibration procedure, already includes all the assumption we defined earlier. We have 1) a finite set of hypothesis represented by the finite number of object on the table 2) a user that is told to assess the robot's actions 3) a finite set of possible meanings (correct and incorrect) 4) signals that can be classified as the calibration procedure was able to generate a usable classifier.

III. Method

This section describes our proposed solution to the previously defined problem of executing a task when the instruction meaning is unknown or uncertain and under the given assumption. The main idea is depicted in Fig. 2 for a toy 1D example. The user wants the device to reach the right-most state. However, neither the target $\hat{\xi}$ nor the true feedback labels $\hat{z}$ are known. The feedback signals e are generated as a response to the execution of an action a in state s according to the true unknown task $\hat{\xi}$ the user wants to solve. The key point is that these signals are generated from an underlying model that for binary signals has two different classes. Given sufficient feedback signals, it is possible to build the underlying distributions for each possible target. Only the right task will provide the right meanings (or labels) to each of the feedback signals (Fig. 2 Left), while the other tasks will gradually mix

both classes as the task gradually differs more from the original task (Fig. 2 Middle-Right), up to the point of almost mirroring the labels when the target is mirrored. In the remainder of this section we show how this property can be exploited to estimate the task and the model generating the feedback signals.

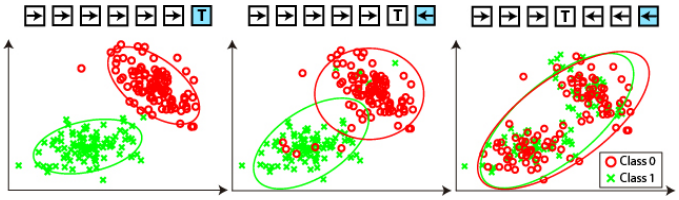


Fig. 2: Task-dependent labels for a 1-D grid world. For the represented example, the arrows indicate for each state what action should elicit a positive feedback to reach the target marked with T (i.e., the optimal policies). 2D Gaussian distributions of binary feedback signals for three possible targets are shown below. While for the correct target the distributions shows a large separability (Left), the overlaps increases as the believed target moves away from the real one (Middle, Right).

Following the literature [17], we will model the EEG signals using independent multivariate normal distributions for each class, $\mathcal{N}(\mu_c, \Sigma_c), \mathcal{N}(\mu_w, \Sigma_w)$. Here, the model parameters θ account for $\{\mu_c, \Sigma_c, \mu_w, \Sigma_w\}$.

Regarding the tasks, the system has access to a set of task hypotheses Ξ which includes the task $\hat{\xi}$ the user wants to solve[1]. We do not make any particular assumption on how the task is represented given that for each particular task ξ we are able to compute a policy π_ξ which represents the probability of choosing a given action a in state s, $\pi_\xi(s,a) = p(a|s,\xi)$. These policies, conditioned on the task, provide meanings to the signals of an action-state pair (e.g. in a reaching task, progressing towards the goal will generate correct answers while moving apart from it will generate wrong ones). We define Z the function that, given a state s, an action a, and a task ξ return the probability of the user intended meaning z, i.e. $Z(s,a,\xi) = p(z|s,a,\xi)$. With $\sum_{f=c,w} p(z=f|s,a,\xi) = 1$. Instead of a binary meaning estimate, we add a noise term to cope with those situations were the user assessment may be wrong. For example, the probability that the user, having task ξ in mind, provides a signal of meaning correct c if the device execute action a in state s is:

$$p(z=c|s,a,\xi) = \begin{cases} 1-\alpha & if\ a = \text{argmax}_a\, \pi_\xi(s,a) \\ \alpha & \text{otherwise} \end{cases}$$

with α modeling error rate of the user. In our case, only two signal meanings are possible, i.e. correct (c) and incorrect (w), therefore: $p(z=w|s,a,\xi) = 1 - p(z=c|s,a,\xi)$

Following the discussion of Fig. 2, a sensible option to estimate the task $\hat{\xi}$ is to measure the coherence of the signal model θ_ξ computed using the virtual meanings, given by Z, provided by the target policy. In other words, the best (ξ, θ_ξ) pair would provide the lowest predictive error (perr) on the observed signals $p(e|s,a,\xi,\theta)$. One possible way of solving this problem is to maximize the expected predictive classification rate:

$$\hat{\xi}, \hat{\theta} \quad = \quad argmax_{\xi,\theta}\ E_e\left(\delta(Z(s,a,\xi), Y(e,\theta_\xi))\right) \qquad (1)$$

where $\delta()$ being an indicator function. And $Y(e,\theta_\xi)$ is the predicted label z for signal e under the model parameterized by θ_ξ, i.e. $Y(e,\theta_\xi) = p(z|e,\theta_\xi)$. With $\sum_{f=c,w} p(z=f|e,\theta_\xi) = 1$. In practice, it is just the probability of the meaning under the Gaussian model provided by θ_ξ. For example, the probability that signal e is of meaning correct (c) under θ can be expressed as:

$$\begin{aligned} p(z=c|e,\theta) &= \frac{p(e|z=c,\theta)p(z=c)}{\sum_{k=c,w} p(e|z=k,\theta)p(z=k)} \\ &= \frac{\mathcal{N}(e|\mu_c,\Sigma_c)p(z=c)}{\sum_{k=c,w} \mathcal{N}(e|\mu_k,\Sigma_k)p(z=k)} \end{aligned} \qquad (2)$$

In our case, only two signal's meaning are possible, i.e. correct (c) and incorrect (w), therefore: $p(z=w|e,\theta) = 1 - p(z=c|e,\theta)$.

The expected predictive error can be explicitly written dependent on the task and decoder model:

$$\begin{aligned} E_e\left(\delta(Z(s,a,\xi), Y(e,\theta))\right) \quad = \\ \sum_{f=c,w} p(z=f|s,a,\xi)p(z=f|e,\theta) \end{aligned} \qquad (3)$$

Note that the optimization process has been factored using the fact that given a task ξ, the estimation of θ under the Gaussian model is trivial. It basically requires to compute the maximum-likelihood estimate θ_ξ^{ML} for each task ξ.

Concretely, given a set of task hypothesis Ξ of size T, we can assign, for each hypothesis, probabilistic labels to the signals received from the user (Z). This provides one dataset of signals with T sets of labels. For each task hypothesis and given its associated hypothetic label set, we can now compute the maximum-likelihood model θ^{ML}. By comparing the fitted model prediction (Y) with the initially assigned labels (Z), we can compute a score, here the expected predictive error, that account for the coherence of the spacial organization of brain signals in the feature space with the associated hypothetic labels. The idea is that only the right task will provide the right meanings (or labels) to each brain signals, while the other tasks will gradually mix both classes (Fig. 2). The correct task should therefore have a lower expected predictive error.

IV. Results

In this section we present online results from a BCI scenario as well as a pick and place HRI scenario to illustrate the wider potential application of our approach. For the remaining of this section, we will consider the error rate of the user α equals to 0.1.

A. BCI Control Task

1) Control task: As illustrated in figure 1, we consider a 5x5 grid world, where an agent can perform five different discrete actions: move up, down, left, right, or a target-reached action. The user goal is to teach the agent to reach one, yet unknown to it, of the 25 discrete positions which represent the set of possible tasks. We thus consider that the agent has access

[1]If this is not the case, the system will find the most suitable task.

to 25 different task hypothesis (one with goal location at each of the cells). We use *Markov Decision Processes* (MDP) to represent the problem [18]. From a given task ξ, represented as a reward function, we can compute the corresponding policy π_ξ using, for instance, Value Iteration [18].

2) EEG-based feedback signals: EEG signals were recorded with a gTec system (2 gUSBamp amplifiers) with 32 electrodes distributed according to the 10/10 international system, with the ground on FPz and the reference on the left earlobe. The EEG signals were digitized with a sampling frequency of 256 Hz, common-average-reference (CAR) filtered and band-pass filtered at $[0.5, 10]$ Hz.

During operation, the role of the users was to assess the agent actions as good or bad, obtaining this way potentials associated to correct or erroneous actions. Previous studies have demonstrated that these signals can be detected online [4] and even be used as binary feedback signals [6]. Following these studies, features were extracted from two fronto-central channels (FCz and Cz) within a time window of $[200, 700]$ ms (being 0 ms the action onset of the agent) and downsampled to 32 Hz. This leaded to a vector of 34 features. This feature vector served as the input for our algorithm

3) Zero-calibration BCI Control with Human Subjects: This experiment will evaluate the main claim of our algorithm, that we can identify the task desired by the user even without an explicit calibration phase and without any knowledge of the brain signals. The experiments were conducted with four subjects (aged between 25 and 28). Each subject performed 5 runs of learning from scratch how to reach a target (chosen randomly).

Figure 3 summarizes the results. The probability of the correct task (averaged across subjects and tasks) is shown in Fig 3a. Figure 3b shows the run by run results. We can conclude that the algorithm is very robust as all the subjects were able to identify the correct task. There are strong variations among subjects, but we note that in previous works the calibration phase used between 300 and 600 examples [6], [19]. Thus, even for the worst subject, it is still possible to start controlling the system without calibration and in less iteration than required by such calibration procedure.

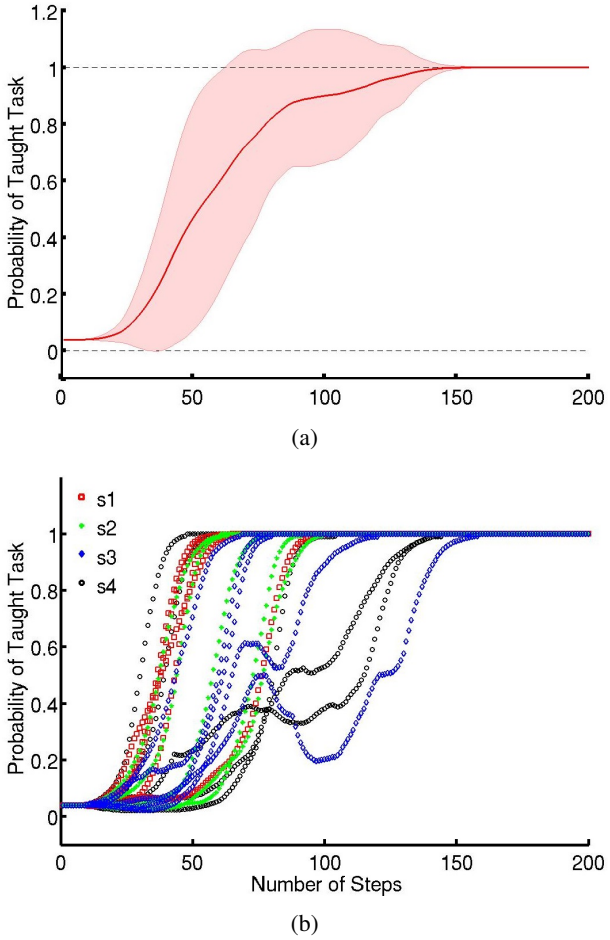


Fig. 3: Results from the online BCI experiment for identifying the task: a) Evolution of the probability of the taught task averaged for all subjects; b) Evolution of the probability of the taught task for each subject and run

B. HRI Pick and Place scenario

In this section, we illustrate the broad range of possible application for our approach with a small size pick-and-place task with a real robot. This robot is going to be programmed using a natural speech interface whose words have an unknown meaning and are not transformed into symbols via a voice recognizer. The interaction between the robot and the human is a turn taking social behavior, where the robot performs an action and waits for a feedback instruction signal to continue. This allows to synchronize a speech wave with its corresponding pair of state and action.

1) Experimental System: We consider a six d.o.f. robotic arm and gripper that is able to grasp, transport and release cubes in four positions. We used a total of three cubes that can form towers of two cubes. The robot has 4 actions available: *rotate left*, *rotate right*, *grasp cube* and *release cube*. The state space is discrete and defined as the location of each object, including being on top of another or in the robot's hand. So for a set of 3 objects we have 624 different states. Figure 4 shows the robot grasping the orange cube.

As for the BCI control task, MDP is used to represent the problem. For this particular representation we assume that the reward function is sparse and so we can generate possible tasks by sampling sparse reward functions. In other words the task is to reach one, yet unknown, of the 624 states of the MDP.

2) Speech processing: We consider speech as the modality for interacting with the robot. After each action we record the teaching word pronounced by the user. This data is mapped into a 20 dimensional feature space using the methodology described next.

A classical method for representing sounds is the *Mel-Frequency Cepstral Coefficients* (MFCC) [20]. It represents a sound as a time sequence MFCC vectors of dimension 12. Comparing sounds is done via *Dynamic Time Warping* (DTW) between two sequences of feature vectors [21]. This distance is a measure of similarity that takes into account possible insertions and deletions in the feature sequence and is adapted

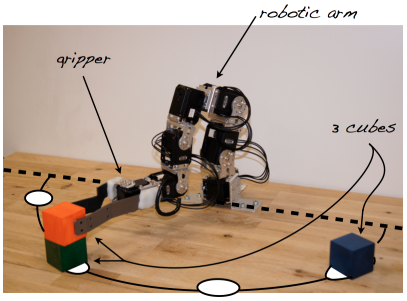


Fig. 4: Robotic System. A six d.o.f robotic arm and gripper learning to performing a pick-and-place task with three cubes.

for sounds comparison of different length. Each recorded vocal signal is represented as its DTW distance to a base of 20 predefined spoken words which are not part of words used by the teacher.

By empirical test on recorded speech samples, we estimate that a number of 20 bases words were sufficient and yet a relatively high number of dimensions to deal with a variety of people and speech. This base of 20 words has been randomly selected and is composed of the words: *Error, Acquisition, Difficulties, Semantic, Track, Computer, Explored, Distribution, Century, Reinforcement, Almost, Language, Alone, Kinds, Humans, Axons, Primitives, Vision, Nature, Building*.

It should be made explicit that this is not state of the art speech processing technics but is not the concern of our research. This representation allows to represent spoken words in a relatively low dimensional space with good accuracy.

3) Zero-calibration HRI online pick and place experiment: This brief experiment demonstrates the transferability of our approach to other domains. In addition, it briefly illustrates the ability of our algorithm to reuse acquired knowledge. Once the robot has understood the first task, we can freeze the classifier corresponding to the identified task and start learning a new task faster as this time the signal to meaning mapping is known.

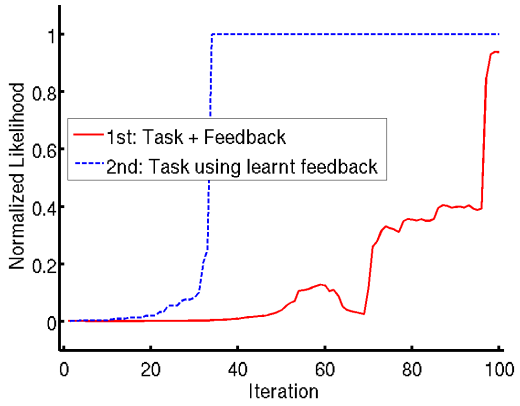


Fig. 5: Evolution of the probability of the taught task. 1) The robot learns a task from unlabeled speech feedback. 2) By freezing the classifier corresponding to the best task estimate, the user teaches the robot a new task faster.

Fig. 5 shows results from one online interactive session with a user using speech to teach the robot what configuration of cube it wanted to the robot to build. In the first run it takes about 100 iterations for the robot to learn the task. Whereas in the second run, when reusing knowledge from the first one, the robot is able to learn a new task faster, in about 30 iterations, meaning that it has well found the two clusters in our feature space as well as the mapping to their corresponding meanings.

V. Discussion

For communication to be successful, the human and the machine need to share some common background which is usually the meaning of the signals received by the device. In practice, such signal to meaning mapping is represented by a specific classifier learnt using a calibration procedure. In this work we have seen that this signal-to-meaning classifier can be learnt automatically and online by the system under the assumption that both, human and machine, share the same a priori on the possible meanings of the signals and the possible task the user may want the device to achieve. We presented a learning algorithm able to associate meaning with unknown signals by reasoning about their relation to previous signals and their relation to the environment itself. The intuition for our method is that the classification of the brain/speech signals is easier when they are interpreted according to the task desired by the user. The method thus relies on finding which pair of classifier-task has the smaller expected prediction error in the signals. We considered the case of brain signals but of particular interest is the possibility to use the same system with other modalities, such as speech. This allows different users to use the system according to their own preferences, skills and limitations. Finally, we showed that, once the system has identified a first task, it can reuse the acquired knowledge about the user instruction signals for learning of a new task faster.

An important challenge for such interactive systems is to deal with non-expert humans. Several studies discuss the different behaviors naive teachers use when instructing robots [22], [23]. An important aspect is that the feedback is frequently ambiguous and deviates from the mathematical interpretation of a reward or a sample from a policy. For instance, in the work of [22] the teachers frequently gave a positive reward for exploratory actions even if the signal was used by the learner as a standard reward. Also, even if we can define an optimal teaching sequence, humans do not necessarily behave according to those strategies [23]. Such aspects were not further considered in this work than by modeling the error-rate of the user.

We believe working without calibration procedure is a novel challenge that can make human to machine interaction more practical to use. Future work will consider how the device can act in order to disambiguate faster the different hypotheses. An important direction is to push this method towards more advanced robotic scenarios by considering, for example, continuous state-action spaces, asynchronous interactions and more complex types of instructions.

ACKNOWLEDGEMENTS

The authors from INRIA are with the Flowers Team, a INRIA / ENSTA-Paristech joint-lab. The research reported was (partially) supported by INRIA, Conseil Régional d'Aquitaine and the ERC grant EXPLORERS 24007.

REFERENCES

[1] J. Millán *et al.*, "Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges," *Front Neurosci*, 2010.

[2] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic lateral sclerosis*, vol. 11, no. 5, pp. 449–455, 2010.

[3] M. A. Lebedev and M. A. Nicolelis, "Brain–machine interfaces: past, present and future," *TRENDS in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.

[4] P. Ferrez and J. Millán, "Error-related EEG potentials generated during simulated Brain-Computer interaction," *IEEE Trans Biomed Eng*, 2008.

[5] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, "Adaptive classification for brain computer interfaces," in *EMBC*, 2007.

[6] R. Chavarriaga and J. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces," *IEEE Trans Neural Syst Rehabil Eng*, 2010.

[7] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4822–4829.

[8] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Trans Biomed Eng*, 2011.

[9] J. Polich, "On the relationship between EEG and P300 : individual differences, aging, and ultradian rhythms," *International Journal of Psychophysiology*, 1997.

[10] I. Iturrate, L. Montesano, and J. Minguez, "Task-dependent signal variations in eeg error-related potentials for brain-computer interfaces," *Journal of Neural Engineering*, vol. 10, no. 2, p. 026024, 2013. [Online]. Available: http://stacks.iop.org/1741-2552/10/i=2/a=026024

[11] B. Argall, S. Chernova, and M. Veloso, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, 2009.

[12] M. Lopes, T. Cederborg, and P.-Y. Oudeyer, "Simultaneous acquisition of task and feedback models," in *IEEE - International Conference on Development and Learning (ICDL'11)*, 2011.

[13] J. Grizou, M. Lopes, and P.-Y. Oudeyer, "Robot learning simultaneously a task and how to interpret human instructions," *Proceedings of ICDL-Epirob*, 2013.

[14] P.-J. Kindermans, D. Verstraeten, and B. Schrauwen, "A bayesian model for exploiting application constraints to enable unsupervised training of a P300-based BCI." *PloS one*, 2012.

[15] P. Kindermans and H. Verschore, "A P300 BCI for the Masses: Prior Information Enables Instant Unsupervised Spelling," in *NIPS*, 2012, pp. 1–9. [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_0332.pdf

[16] I. Iturrate, L. Montesano, and J. Minguez, "Shared-control brain-computer interface for a two dimensional reaching task using eeg error-related potentials," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013.

[17] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. Müller, "Single-trial analysis and classification of ERP components: A tutorial," *Neuroimage*, 2010.

[18] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998, vol. 28.

[19] I. I, M. L, and M. J, "Single trial recognition of error-related potentials during observation of robot operation," in *EMBC*, 2010.

[20] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of mfcc," *Journal of Computer Science and Technology*, vol. 16, no. 6, pp. 582–589, 2001.

[21] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 1, pp. 43–49, 1978.

[22] A. L. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners," *Artificial Intelligence Journal*, vol. 172, pp. 716–737, 2008.

[23] M. Cakmak and A. Thomaz, "Optimality of human teachers for robot learners," in *Proceedings of the International Conference on Development and Learning (ICDL)*, 2010.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Detection of *event-less* error related potentials

Jason Omedes, Iñaki Iturrate, Luis Montesano

***Abstract*—Recent developments in brain-machine interfaces (BMIs) have proposed the use of error-related potentials as a type of cognitive information that can provide a reward or feedback to adapt the BMI during operation, either to directly control devices or to teach a robot how to solve a task. Due to the nature of these signals, all the proposed error-based BMIs work under the assumption that the response is time-locked to the known onset of the event. However, during the continuous operation of a robot, there may not exist a clear event that elicits the error potential. Indeed, it is not clear whether such a potential will appear and whether it can be detected online. Furthermore, calibrating such a system is not trivial due to the unknown instant at which the user detects the error. This paper presents a first study towards the detection of error potentials from EEG measurements during continuous trajectories performed by a virtual device. We present a experimental protocol that allows us to train the decoder and detect the errors in single trial. Further analyses show that the brain activity used by the decoder comes from brain areas involved in error processing.**

I. Introduction

Brain-machine interfaces (BMI) aim to decode brain activity to control devices or provide a communication channel to the user [1]. Recently, several works in BMIs have started to use cognitive information decoded from the user's brain activity in an attempt to improve and extend the capabilities of this type of systems. The underlying idea is to use and decode natural brain activity directly related to the task. For instance, the intention of motion can be used to trigger a robotic device [2] or to anticipate braking during driving [3]. Or, the perception of an error can be used to correct the output of the BMI [4] or to adapt the behavior of the system [5].

In this context, error-related potentials (ErrPs) [6] have gained considerable attention as a cognitive signal that can be incorporated in BMI systems. Error-related potentials are a special kind of event-related potential elicited when the actual outcome of an event differs from the user's expected one. They are elicited in different situations: when the user realizes that he has committed a mistake [7], after the observation of a mistake committed by another person [8], or even when he observes a machine commit a mistake [6], [9]. The latter case is of particular interest for BMIs, and has been used to adapt the BCI classifier [10] or prevent from executing missclassified commands [4].

Interestingly, these signals have been associated to the dopaminergic neural system and, consequently, to human reinforcement learning processes [11]. Following these connections, error potentials have been used as reward signals during reinforcement learning in virtual cursors [9] and robotic arms [12]. Incorporating the assessment of the user opens the door to develop systems that continuously adapt to the user's preferences during operation. Furthermore, in our recent work we have also used error potentials as feedback during the online control of cursors [5] and mobile robots [13].

Jason Omedes, Iñaki Iturrate and Luis Montesano are with the I3A, DIIS, and Univ. Zaragoza, Spain. eMail: {jomedes, iturrate, montesano}@unizar.es.
This work has been supported by Spanish projects DPI2011-2589, and DGA-FSE (grupo T04).

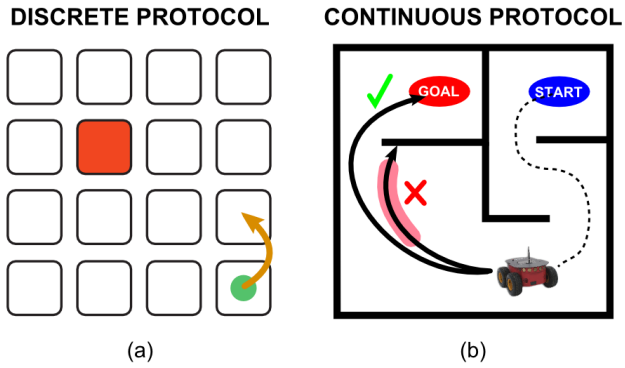


Fig. 1. (a) Discrete scenario, where a device performs discrete actions within a grid to reach a goal location (shadowed in red). (b) Continuous scenario, where a mobile robot is moving continuously through a maze. The error may be detected all along the shadowed part of the trajectory.

Despite the previous achievements, there are still strong limitations to use error-related potentials during the control or the learning of a robotic device. One of the major difficulties, which is shared with all event-related potentials, is that these signals are a response to an event (either exogenous or endogenous) that elicits them. Indeed, there is no study in the literature of error potentials in which the event marking the onset of the potential is not clearly defined and measured. Consequently, the works mentioned before have used discrete worlds (e.g. grids) with instantaneous actions with a clear onset (see Figure 1a), in which case they can be detected in single trial [9], [14]. On the other hand, real applications (such as executing a trajectory with a robotic arm or a mobile robot) imply the use of continuous actions where the error can appear at any moment of the trajectory being executed and not only at the beginning (see Figure 1b). Furthermore, being the error potential a cognitive process its elicitation will depend on the subjective evaluation of each user. (i.e., different elicitation times for each user).

This paper presents the first attempt to detect error potentials during the continuous operation of a device (in our case a cursor on a screen) when there is no clear event that should elicit the potential (hence the name of *event-less*). There are several challenges. First, it is unclear whether these signals are actually elicited under continuous actions, and if they are, whether it is possible to detect them in single trial. Second, the calibration process is not trivial due to the lack of a clear onset. We developed an experimental protocol for a target-reaching task where a device moved continuously while the user assessed the actions performed by it. Two conditions were tested. In the first one, the error was clearly marked

as a sharp turn, while in the second one, the trajectory had no clear event for the error (*event-less*). The experimental results show that it is possible to detect errors in single trial in the *event-less* condition using the first condition as training data to calibrate the BMI. Furthermore, we provide electrophysiological evidence that supports the fact that the used brain activity is originated in brain areas related to error processing.

II. Methods

A. Data recording

Electroencephalographic (EEG) and electrooculographic (EOG) activity were recorded using a commercial gTec system consisting of 32 electrodes distributed according to an extended 10/20 international system (FP1, FP2, F7, F8, F3, F4, T7, T8, C3, C4, P7, P8, P3, P4, O1, O2, AF3, AF4, FC5, FC6, FC1, FC2, CP5, CP6, CP1, CP2, Fz, FCz, Cz, CPz, Pz and Oz), with the ground on FPz and the reference on the left earlobe; for the EOG, 6 monopolar electrodes were recorded (placed above and below each eye, and from the outer canthi of the left and right eyes [15]), with the ground on FPz and the reference on the left mastoid. The EEG and EOG signals were digitized with a sampling frequency of 256 Hz, power-line notch filtered, and band-pass filtered at $[1, 10]$ Hz. The EEG was also common-average-reference (CAR) filtered. Additionally, the horizontal, vertical, and radial EOG were computed as in [15] to remove the EOG from the EEG using a regression algorithm [16]. The data acquisition and on-line processing was developed under a self-made BCI platform.

B. Experimental setup

Two healthy subjects (mean age 28 years) participated in the study recorded in a laboratory of the University of Zaragoza. Participants were asked to restrict blinks to the specific resting periods. The experimental setup consisted of a virtual cursor that had to reach a target position by moving at a fixed speed towards it. The initial cursor and target positions of each trial were the same for both subjects. They were randomly generated forcing a distance of at least 200 pixels between them. One trial consisted of a trajectory performed by the device and lasted a maximum of 5 seconds. Trajectories were correct 70% of the trials. Correct trajectories consisted on straight lines between the start and goal locations, Figure 2a. Erroneous trajectories started as the correct ones but changed direction in a random instant between the 20% and 80% of the path. Two different conditions were tested: (i) a sharp change of direction, analog to a typical ERP protocol where the event onset is present (Figure 2b), and (ii) a smooth change of direction (i.e. a curved movement always of the same duration but with different angles, Figure 2c). We denote this condition as event-less, since the perception of the error is not clearly defined and depends on the subject assessment. Each round was composed of 40 trials, with a break of few minutes between rounds. Six rounds of each condition, alternating between error types (sharp/smooth), were recorded, obtaining around 70 erroneous trials per experiment and participant.

C. Electrophysiology analysis of time-locked error potentials

In order to determine whether the error potentials were present for this protocol, we firstly analyzed the ErrP under the sharp condition, where the onset for erroneous movements was clear. Notice that this case resembles the standard error potentials protocol. The onset of the erroneous event is selected from the time instant in which the device performs the abrupt change of direction. On the other hand, correct trials did not have a specific event. Thus, the onsets of these events were selected at a random instant of time within the execution of a correct trial.

For the electrophysiology analysis, the time-locked averaged potentials were computed for the error, non-error and difference (error minus non-error averages), and averaged for all participants at channel FCz [17]. Scalp topographies at the most relevant peaks of these potentials were also computed. Additionally, the error potentials were also analyzed on the frequency domain by means of the power spectral density (PSD). The PSD was computed from each one-second trial using the Welch's method with a Hamming window, and a window overlap of 250 ms. Then, the error, non-error and difference average PSDs for all participants were computed at channel FCz. Finally, a source localization analysis was performed on the obtained signals with sLoreta [18].

D. Feature extraction

Previous studies in standard protocols have mainly detected error potentials using features from the temporal domain [9], [14]. This type of features are not so well suited for continuous detection, since EEG oscillations can easily resemble ErrP patterns thus resulting in a large number of false positive. Thus, in this work we combined temporal and frequency features extracted from the most relevant common spatial patterns associated (CSPs) to the ErrP. CSPs have been used in the past for the continuous classification of EEG signals, such as motor imagery [19] or slow cortical potentials [20]. CSPs were extracted using the training data (see subsection II-E). The first two spatial patterns were retained as the most discriminant activity between erroneous and correct trajectories. For each chosen spatial pattern, the temporal features were extracted as the EEG voltages within a time window of $[0, 800]$ ms (being 0 ms the direction change onset) downsampled to 64 Hz, forming a vector of 78 features. Regarding to frequency features, the PSD was firstly computed as presented in section II-C. Then, frequency features were selected as the power values of each channel from the theta band ($[4, 8]$ Hz) $\pm$ 1 Hz, as previous studies suggested that the error potentials are generated within this band [21], leading to a vector of 50 features. Finally, both set of features were concatenated and normalized within the range $[0, 1]$.

E. Single-trial continuous classification

The previous features were fed into a support vector machine (SVM) classifier with a radial basis function (RBF) kernel [22]. To avoid SVM sensitivity to imbalanced datasets, the minority class (i.e. the error class) was oversampled to match the number of trials of the majority class (i.e. the non-error class) [23]. During classification, the classifier output was the probability that the current EEG data was an error, p_e.

For the single-trial classification, both conditions (sharp and smooth direction changes) were tested. Since the smooth condition did not have an onset, it was not possible to

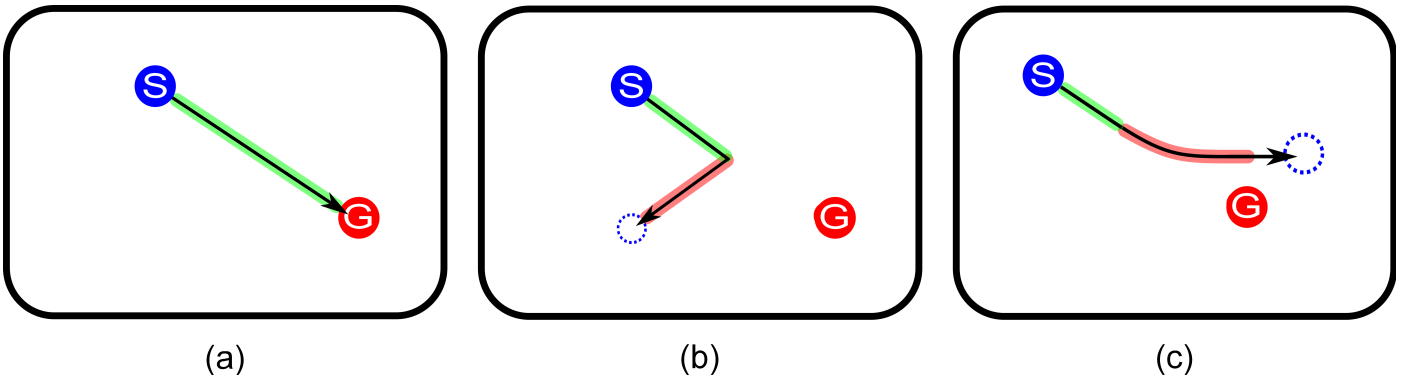


Fig. 2. Designed experimental setup. Starting (S) and goal (G) positions of the device are marked in blue and red respectively. (a) The device performs a correct movement by executing a movement at continuous speed towards the target. (b)-(c) Erroneous trials where the device changes its direction, either (b) abruptly, or (c) smoothly. Correct and wrong directions of movements are shadowed in green and red respectively.

extract meaningful information to train the classifier. Thus, the training-testing sets were composed as follows: for the sharp condition, we performed a 6-fold cross-validation where each fold was composed by a complete recorded round. For the smooth condition, the training set was composed by all the rounds from the sharp condition, whereas the testing set were all the rounds from the smooth condition. For both cases, events used for training were extracted using the onset as described in section II-C.

Once the classifier was trained, we continuously classified every 62.50 ms the testing sets using a sliding window of one-second width. In order to ensure a low false positive rate, the detection of error events was only considered when $p_e > 0.8$. Additionally, any possible eye activity was automatically removed by setting to zero the classifier output any time the EOG signal exceeded 40 μV. For the sliding window, all inter-trial data was removed since it was considered as a resting period and the subject could have been performing muscular activity.

The performance of the computed sliding window was determined as follows: those erroneous trials where the sliding window detected an error, as long as the error was detected after the change of direction (either smooth or abrupt), where considered as true positives (TP). When the error was not detected the trial was considered a false negative (FN). Those correct trials where the classifier did not detect any error were considered true negatives (TN). When an error was detected on correct trials, they were considered false positives (FP). To obtain a more intuitive representation of the performance achieved, we also displayed the trajectories followed by the device for each one of the four possible cases. The goal position of each trial has been repositioned to the center of the image for a better representation.

F. Post-hoc electrophysiology analysis for the event-less condition

The analysis of Section II-C can only be carried out when the onset of the error potential is known (i.e. only for sharp changes). In order to evaluate whether the detection of error potentials in the *event-less* condition uses brain activity related to the error, we performed a post-hoc analysis using the output of sliding window classifier as an artificial onset of the error potential. Only correctly detected erroneous trials (i.e. true positives) were used in the analysis that was identical to that of Section II-C.

III. Results

A. Electrophysiology time-locked analysis

Figure 4 shows the error, correct and difference grand averages for channel FCz averaged for the two subjects in temporal and frequency domain. The difference average is characterized by a sequence of a positive peak at 150 ms, followed by a negative peak at 210 ms and two larger positive and negative peaks at 280 and 400 ms, and finally a positive peak at 600 ms. The topographic interpolation of these peaks can also be seen on Figure 4, showing how they are localized on fronto-central scalp areas. These results agree with previous studies using error potentials under standard conditions (i.e., discrete device actions) studies [6], [9]. Regarding to the frequency averages, a power increment in the theta band was observed for erroneous trials with respect to the correct ones, which also agrees with previous works [21]. Finally Figure 4, Bottom shows the source localization results for the most prominent negative peak (400 ms) of the time-locked difference grand average. The main activation areas were Brodmann areas 6 and 24 (premotor cortex and ventral anterior cingulate cortex), which is in accordance with previous works simultaneously recoding error-related activity and fMRI [24].

B. Classification of time-locked error potentials (condition 1)

Regarding to the results of applying the sliding window, Figure 3 displays the detection level obtained for both subjects in a representative round of the sharp experimental condition. Here, the 40 trials that compose the round are concatenated removing the inter-trial resting periods. In the shown example, it can be seen that most of the trials are properly classified. Also notice that the correctly detected ErrPs are delayed with respect the onset of the event (the sharp change in direction of the device). This delay was on average 867.13 ± 99.33 ms after the onset of the erroneous action. This delay was expected, corresponding to the time needed for the appearance of the most relevant peaks and the maximum spectral power activation used as features.

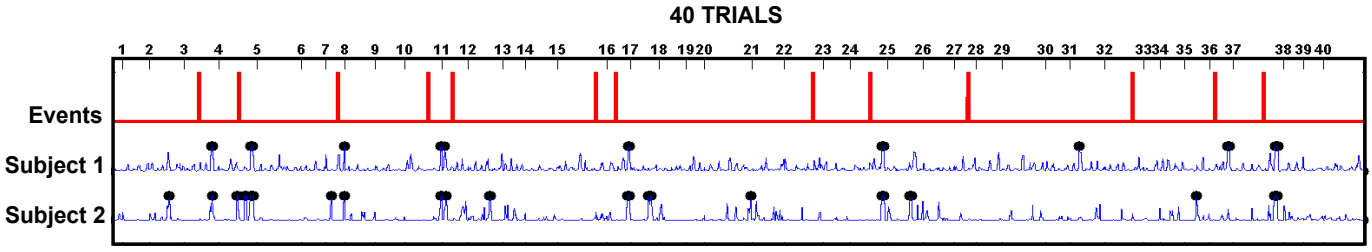


Fig. 3. Representative example of the sliding window results for the sharp condition. The error events are plotted in red indicating a change in direction, while the probability of detecting an error (p_e) for each of the 2 subjects is plotted in blue. Black dots over the probability values indicate the time instant when the classifier detected an error ($p_e > 0.8$).

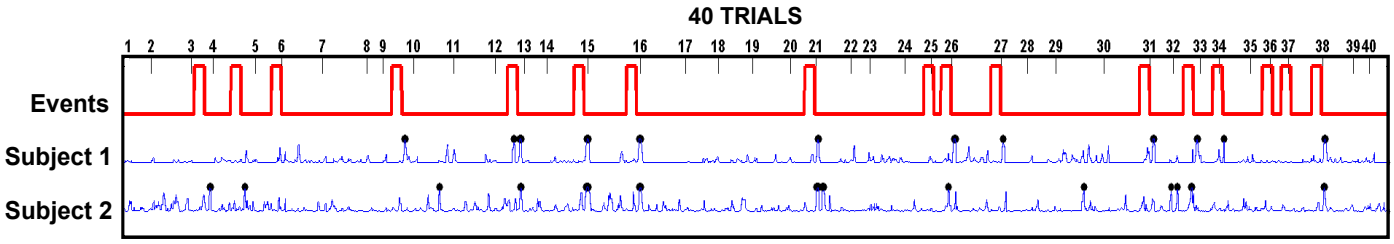


Fig. 5. Representative example of the sliding window results for the *event-less* condition. In this case, the error events (plotted in red) indicate the duration of the turn.

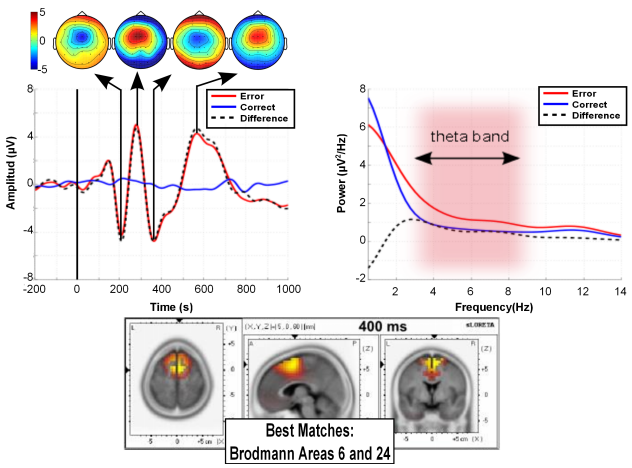


Fig. 4. Electrophysiology results of the experimental condition 1 (sharp turns). Temporal and frequency averages of error and correct trials plus the difference (error minus correct) for channel FCz and scalp topographies at the occurrence of the most relevant peaks for the average of the two subjects.

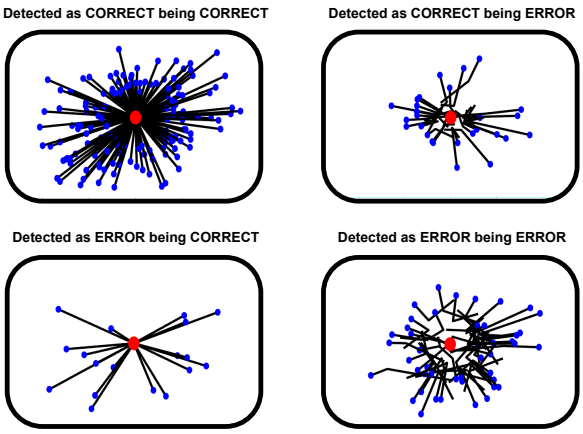


Fig. 6. Confusion matrix of the trajectories performed by the device (black) during the sharp condition. The goal positions have been reprojected to the center of the image (red). The starting position of the cursor with reference to the goal is marked in blue.

The performance rates achieved for the entire test set are depicted in Table I. It can be observed that the number of false positives was reasonably low, which was the main objective of setting a high threshold value ($p_e > 0.8$). At the same time the number of erroneous trials detected as such (true positives), reached 64%. This value was around 10% less performance than those obtained with standard protocols using discrete actions [5], [9].

Finally, Figure 6 displays the trajectories executed by the device according to their classification. Here, it can be seen that correct trials are mostly well detected independently of the direction and distance covered by the device, and only few of them are detected as erroneous. More interestingly, the number

TABLE I. CONFUSION MATRIX CONTAINING THE PERFORMANCE RATE FOR THE EXPERIMENT 1

		Actual Class	
		Correct	Error
Predicted Class	Correct	TN = 89.09%	FN = 36.00%
	Error	FP = 10.91%	TP = 64.00%

of erroneous trials not detected were higher. This was done this way since it was preferable to miss the detection of an error than detect errors where was not intended. Additionally, it can also be observed that many of these trials detected as correct end up very close to the actual goal, which lead us to think that the subjects may have not interpreted them as erroneous.

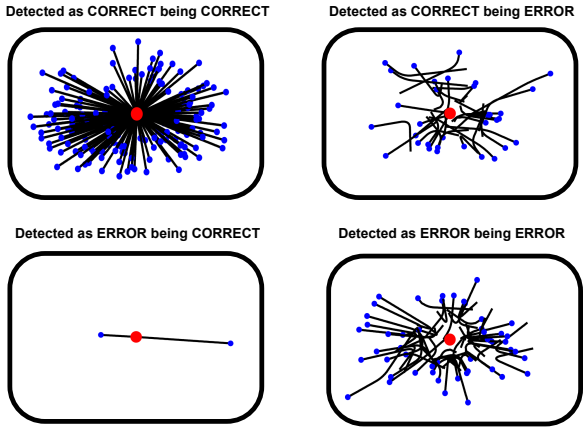


Fig. 7. Confusion matrix of the trajectories performed by the device (black) during the second experimental condition.

TABLE II. CONFUSION MATRIX CONTAINING THE PERFORMANCE RATE FOR THE EXPERIMENT 2

		Actual Class	
		Correct	Error
Predicted Class	Correct	TN = 97.15%	FN = 32.66%
	Error	FP = 2.85%	TP = 67.33%

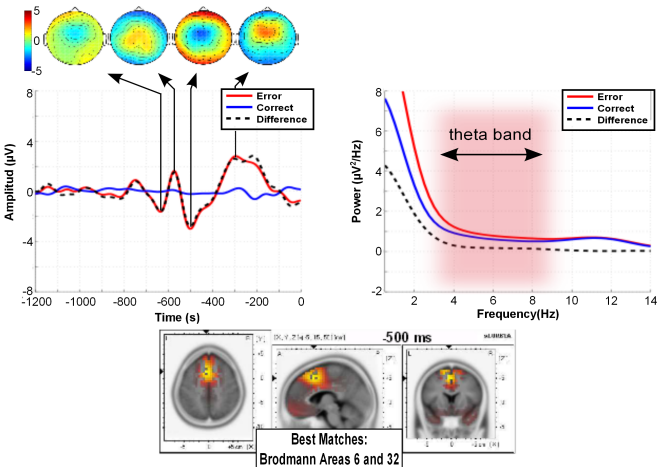


Fig. 8. For the *event-less* condition, temporal and frequency averages of error and correct trials plus the difference (error-minus-correct) for channel FCz and scalp topographies at the occurrence of the most relevant peaks for the average of the two subjects.

C. *Classification of the event-less condition*

The process to analyze this section is similar to the one followed in the previous section III-B. Regarding to the results of applying the sliding window, Figure 5 displays the detection level obtained for both subjects in a representative round of the smooth experimental condition. Once again, the 40 trials composing the round were concatenated after the removal of the inter-trial resting periods. In the example, it can be observed that the classification is carried out successfully, properly detecting most of the events. In this case, it is also noticeable the presence of a delay for correctly detected ErrPs. However, since there not exist a clear onset that elicit the error potential, this delay cannot be computed. Nonetheless, it was possible to compute the delay of detecting an error after the curve finished, which was 297.22 ± 213.72 ms on average. The larger standard deviations obtained indicated that the moment of error detection had larger trial-to-trial variations. On the other hand, assuming a similar error delay as the obtained in the previous case (867.13 ms), the ErrPs always appeared at random points within the radial movements. Thus, the potentials were not elicited at the beginning or the end of the radial turn, but rather depended on when the users became aware of the error.

The performance rates achieved for the entire test set are shown in Table II. It can be observed that the number of false positives was even lower than for the previous case, which may be caused by the extrapolation between training and testing data. Surprisingly, the number of erroneous trials detected as error (true positives), reached 67.33%, which is a slightly higher that in the previous case.

Finally, Figure 7 displays the trajectories executed by the device according to their classification. Once again, it can be seen that correct trials have a similar behavior to the previous condition, and comparable detection rates are achieved. Once more, it can be observed that many of the erroneous trials detected as correct have their trajectories very close to the goal, which could be due to the subjective user's interpretation.

D. *Post-electrophysiology* event-less *analysis*

Figure 8 shows the grand average of error (as detected by the sliding window classifier) and correct trials plus the difference (error minus correct) for channel FCz averaged for the two subjects, both in temporal and frequency domain. Notice that the signal is referenced to the instant of time where the ErrP has been detected. Thus, the peaks analysis leads to a first positive peak that appears -750 ms before the detection of the ErrP. This positive peak is analog to the positive peak that appeared at 150 ms after the onset in the previous condition. The following observed positive peaks are found at -570 and -300 ms, while negative peaks are seen at -640 and -500 ms before the ErrP detection. All these peaks correspond to the positive peaks located at 280 and 600 ms; and to the negative peaks found at 210 and 400 ms respectively, observed in Figure 4. Furthermore, the scalp topographies of the described peaks had a similar morphology to the ones observed in the previous condition. Nonetheless, despite the averages and topographies resembled a similar morphology to the one obtained in the standard electrophysiology analysis, the ErrP amplitudes in this case were lower, most likely due to an inaccurate signal averaging. Regarding the frequency information of the averaged signals, a slight power increment in the theta band was present for the erroneous trials. Finally, Figure 8(Bottom), displays the source localization results at the most prominent negative peak (-500 ms). The brain activity corresponding to this peak was located at the Brodmann area 6 and 32 (premotor cortex and dorsal anterior cingulate cortex). In sum, these results suggest that the error potentials are elicited within continuous motions performed by a device, even when there is not a clear event that trigger them.

IV. Conclusions and future work

This paper studies the online detection of error potentials when the onset of the event that elicits them is unknown. The results obtained for the proposed experimental protocol show that the error potentials appear when the user monitors a target reaching task and that they can be detected in single trial. Furthermore, the electrophysiology analysis of the recorded signals indicated that the main components of the error-related potentials were similar to previous studies and they had their origin in the anterior cingulate cortex. These promising results are a first step towards the use of this type of cognitive information to control or teach robotic devices in realistic and complex tasks. We have included a preliminary video of a mobile robot controlled using these signals. There are plenty of opportunities for future work. First, we are currently extending the study to more users and more error conditions. Second, further studies are required to understand whether the error potential appears every time an error is detected or the frequency features allow detecting errors even when no potential is present. Finally, the long term goal is to understand what cognitive information related to error perception can be decoded and incorporated in a BMI to control a robotic device in realistic scenarios.

References

[1] J. Millán *et al.*, "Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges," *Frontiers in neuroscience*, vol. 4, 2010.

[2] Y. Gu, D. Farina, A. R. Murguialday, K. Dremstrup, and N. Birbaumer, "Comparison of movement related cortical potential in healthy people and amyotrophic lateral sclerosis patients," *Frontiers in neuroscience*, vol. 7, 2013.

[3] Z. Khaliliardali, R. Chavarriaga, L. A. Gheorghe, and J. d. R. Milln, "Detection of Anticipatory Brain Potentials during Car Driving," in *Proceeding of the 34th Annual International IEEE EMBS Conference*, ser. IEEE Engineering in Medicine and Biology Society Conference Proceedings. New York: Ieee, 2012.

[4] B. Blankertz, G. Dornhege, C. Schafer, R. Krepki, J. Kohlmorgen, K. R. Muller, V. Kunzmann, F. Losch, and G. Curio, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans Neural Syst and Rehab Eng*, vol. 11, no. 2, pp. 127–131, 2003.

[5] I. Iturrate, L. Montesano, and J. Minguez, "Shared-control brain-computer interface for a two dimensional reaching task using eeg error-related potentials," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013.

[6] P. W. Ferrez and J. del R Millan, "Error-related eeg potentials generated during simulated brain-computer interaction," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 3, pp. 923–929, 2008.

[7] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "Erp components on reaction errors and their functional significance: a tutorial," *Biological psychology*, vol. 51, no. 2, pp. 87–107, 2000.

[8] H. van Schie, R. Mars, M. Coles, and H. Bekkering, "Modulation of activity in medial frontal and motor cortices during error observation," *Nature Neuroscience*, vol. 7, no. 5, pp. 549–54, May 2004.

[9] R. Chavarriaga and J. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces," *IEEE Trans Neural Syst Rehabil Eng*, vol. 18, no. 4, pp. 381–388, 2010.

[10] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, "Adaptive classification for brain computer interfaces," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2007, no. C, Jan. 2007, pp. 2536–9. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18002511

[11] C. B. Holroyd and M. G. H. Coles, "The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity," *Psychological Review*, vol. 109, pp. 679–709, 2002.

[12] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," in *Robotics and Automation (ICRA), IEEE International Conference on*. IEEE, 2010, pp. 4822–4829.

[13] I. Iturrate, J. Omedes, and L. Montesano, "Shared control of a robot using eeg-based feedback signals," in *Proceedings of the 2nd Workshop on Machine Learning for Interactive Systems (MLIS13) at IJCAI13*, 2013.

[14] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. del R Millan, "Latency correction of error potentials between different experiments reduces calibration time for single-trial classification," in *Int Conf of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012.

[15] R. Croft and R. Barry, "EOG correction of blinks with saccade coefficients: a test and revision of the aligned-artefact average solution," *Clinical neurophysiology*, vol. 111, no. 3, pp. 444–51, Mar. 2000.

[16] A. Schlögl *et al.*, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical neurophysiology*, vol. 118, no. 1, pp. 98–104, Jan. 2007.

[17] S. Nieuwenhuis, C. Holroyd, N. Mol, and M. Coles, "Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance," *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 4, pp. 441 – 448, 2004.

[18] P.-M. RD, "Standardized low resolution brain electromagnetic tomography (sLORETA): Technical details," *Methods Find Exp Clin Pharmacol*, pp. 5–12, 2002.

[19] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Trans Neural Syst and Rehab Eng*, vol. 8, no. 4, pp. 441–446, 2000.

[20] G. Dornhege, "Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials," *IEEE EMBS Conference on Neural Engineering*, pp. 595–598, 2003. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1196898

[21] M. X. Cohen, "Error-related medial frontal theta activity predicts cingulate-related structural connectivity." *NeuroImage*, vol. 55, no. 3, pp. 1373–1383, 2011.

[22] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J Neural Eng*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.

[23] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Proceedings of the 15th European Conference on Machine Learning (ECML)*, 2004, pp. 39–50.

[24] T. U. Hauser, R. Iannaccone, P. Stämpfli, R. Drechsler, D. Brandeis, S. Walitza, and S. Brem, "The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization," *NeuroImage (to be published)*, Aug. 2013.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk I

Is artificial emotion really emotional?

Prof. Minoru Asada

Osaka University, Osaka, Japan

Abstract: Emotion, a driving force to generate different behaviors, is one of the most fundamental but difficult structures/functions to design for robots. Starting from primitive emotions, the secondary emotions may be differentiated from them. During this developmental process, sociality has an important role to derive the differentiated emotions. In this talk, I argue how artificial emotion can be more realistic in the social context by showing some attempts, and discuss the future stories in SFs and comics.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk II

Shared body for self and others in the brain

Prof. Akira Murata

Kinki University, Osaka, Japan

Abstract: It has been known that dorsal visual stream of the two visual pathways, directing to the parietal cortex, is related to visual spatial perception. However, the parietal cortex is not the terminal station of the dorsal visual stream, but has strong mutual anatomical connection with the premotor cortex. The spatial information in the parietal cortex is sent to the premotor cortex, and then the final goal is visuo-motor control. On the other hand, on line representation of one's own body (body schema) is formed in the sensory-motor process, and this map can dynamically change, depending on sensory-motor experiences and learning. This network is considered integration of efference copy/corollary discharge and sensory feedback that is an essential factor both for sensory motor control and body schema. In our recent findings, it is suggested that one's own body schema also provides basic reference frame for mapping of other's body. This means that neuronal substrates for monitoring one's own action is shared with the system for recognize and understand other's action. In this lecture, we would discuss about body schema that is a key to link between sensory-motor control and high-order cognitive functions; body recognition.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk III

Cognitive Interaction Technology for Helpful Robots

Prof. Helge Ritter

Bielefeld University, Bielefeld, Germany

Abstract: The perceived helpfulness of a robot and, as a result, the well-being of a human in the presence of the robot, is only partly determined by the robot's function alone. An important co-determining factor is the robot's interface that determines appearance and social interaction with the human. This underscores the creation of flexible and human-adapting interfaces as a key task for the development of robots that are perceived as helpful. We present ongoing work on the creation of interfaces for supporting cognitive interaction between robot devices and humans, emphasising the role of touch and its interplay with vision. Examples include the development of flexible and 3D-shaped tactile sensors and their use in the context of the analysis and control of tactile-guided interaction, such as visuo-haptic servoing, as well as an outlook of how to augment such capabilities with further interfacing modalities towards systems that can interact with humans in natural and rich ways.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk IV

Human-derived sensor fusion principles used to control biped balancing of external disturbances in a humanoid robot

Prof. Thomas Mergner Neurologische Klinik, Freiburg, Germany

Background: Human sensorimotor control is very complex and current research still faces problems when it comes to re-embody the hypothesized control in the form of control models into robots for direct human-robot comparisons. We simplified the research task by studying reactive (sensor driven) postural reactions to exactly known external disturbances. Control of human biped stance during external disturbances lends itself to this research as a simple sensorimotor prototype.

Material and Methods: System analysis approaches with computer modeling in back and forth with human experiments was used in: (1) Investigations of human sensory systems, mainly vestibular and joint proprioceptive, using open loop psychophysics of self-motion perception. (2) Investigations of human postural responses to tilt and translation of the support surface and to pull stimuli having impact on the body. (3) Modeling of the human postural data using sensor fusion principle derived from the human perception and comparing model simulation data with the human postural data. (4) Re-embodiment of the model into a humanoid robot for direct human-robot comparisons in the human laboratory.

Results: (1) Psychophysics suggested that human self-motion perception (a) uses sensory transducer signals to reconstructs the kinematic and kinetic variables of the body-world interaction and (b) uses then these variables to reconstructs the external disturbances having impact on the body. (2) Using these sensor fusion algorithms allowed implementation of the human postural response findings into a simple sensory feedback model of human stance control. The model consists of (i) a servo loop for local joint control and, superimposed on the servo, of (ii) long-latency loops for disturbance estimation and compensation. (3) Model simulations delivered data that resembled the human data. (4) This similarity also applied when the

model was used to control the robot with its noisy and inaccurate sensors, etc., and when performing the simulations in the human test bed.

Discussion & Conclusion: The approach of deriving sensor fusion principles from human self-motion perception and of using these principles to model human sensor-based postural responses to external disturbances may help to better understand the human sensorimotor control. Its extension into a 'neurorobotics' approach provided a proof of principle of the sensorimotor control model and demonstrated certain advantages of this control such as versatility in face of changing disturbance scenarios, high robustness in terms of fail safe, low loop gain and low passive resistance. Currently the model and robot are extended to include voluntary movements, control policies (including fusion of predicted with sensor-derived disturbance estimates), and a modular architecture for multi-DOF systems.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk V

Motor primitives and central pattern generators: from biology to robotics

Prof. Auke Ijspeert

EPFL, Lausanne, Switzerland

Abstract: The ability to efficiently move in complex environments is a fundamental property both for animals and for robots, and the problem of locomotion and movement control is an area in which neuroscience and robotics can fruitfully interact. Animal locomotion control is in a large part based on central pattern generators (CPGs), which are neural networks capable of producing complex rhythmic or discrete patterns while being activated and modulated by relatively simple control signals. These networks are located in the spinal cord for vertebrate animals. In this talk, I will present how we model pattern generators of lower vertebrates (lamprey and salamander) using systems of coupled oscillators, and how we test the CPG models on board of amphibious robots, in particular a salamander-like robot capable of swimming and walking. The models and robots were instrumental in testing some novel hypotheses concerning the mechanisms of gait transition, sensory feedback integration, and generation of rich motor skills in vertebrate animals.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk VI

Mind-Controlled Humanoid Robots and Physical Embodiment

Prof. Abderrahmane Kheddar

National Institute of Advanced Industrial Science and Technology (AIST),
Ibaraki, Japan

Abstract: This talk will address our ongoing research is robotic embodiment and thought-based control of a humanoid robot using brain computer interface. We efficiently integrate techniques from computer vision and the task-function based control together with the brain-computer interface into an immersive and intuitive control application despite the well-known shortcomings of BCI. Our approach is based only on steady state visual evoked potential patterns. The user is fed back on-line with video stream recorded from the humanoid embedded camera. Images are then segmented, clustered from which learned objects are recognized. 3D model of recognized objects are used to superpose their computer graphic representation using augmented reality techniques. The 3D models flicker at frequencies automatically assigned by our system. Once users attention is ported on a given object, SSVEP classifier reports it. Based on the affordance concept, the object of interests associated task is sent to the stack-of-task controller of the humanoid robot. This approach is assessed in a user experiment involving several subjects who successfully controlled the HRP-2 humanoid robot in a scenario involving both grasping tasks and steering. The user experiences and the interface performances are presented and give a rich insight into future research that can be made to improve and extend such interface.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk VII

Robots under Neural Control: How to create a neuron-based learning & memory system for behaving machines?

Prof. Florentin Worgotter

University of Goettingen, Goettingen, Germany

Abstract: Since several years we have tried to show the power of implicit, neural control for behaving artificial systems. We were able to demonstrate reactive as well as adaptive control in our 18DOG hexapode robot AMOS WD6, which leads to more than 10 different behavioral patterns in response to the robot's sensory input signal combination. This as such is a difficult problem as there are no rules or explicit control-laws present in this system. Rather, AMOS behaves like many insects by directly responding appropriately to the requisite variety of its world represented by its many sensors. This, however is not enough. Even simple insects can learn and memorize to some degree. Here we specifically show how a working memory can be implemented using pure neural mechanism directly linked to the behavior of the robot. Typical conditioning situation can thereby by learned and memorized for some time, very similar to, e.g., odor conditioning in insects.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk VIII

Adaptive robot skill synthesis through human sensorimotor learning

Dr. Jan Babic

Jozef Stefan Institute, Ljubljana, Slovenia

Abstract: In this talk, I will introduce a concept of obtaining complex robot motions based on the human sensorimotor learning capabilities. The idea is to include the human in the robot control loop and to consider the target robotic platform as a tool that can be iteratively controlled by a human. Provided with an intuitive interface between the human and robot, the human learns to perform a given task using the robot. The skilled control of the robot by the human provides data that are used for construction of autonomous controllers that control the robot independently of the human. To demonstrate the applicability of the concept, I will present several examples including statically stable reaching, cooperative dynamic manipulation skill and adaptive control of exoskelton robots. Besides, I will also explain how the interfaces built for the robot skill synthesis can be effectively used in the opposite direction to investigate human motor control mechanisms employed by the central nervous system during the full body motion.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk IX

Brain-Machine-Interface Improves Recovery Time from Perturbation in Flight Attitude on a Novel Complex Piloting Task

Dr. Daniel Callan

National Institute of Information and Communications (NICT), Kyoto, Japan

Abstract: The goal of this research is to develop adaptive automation that can improve response speed of a pilot's motor commands to an unexpected event by using a brain-machine-interface BMI to decode perceptual-motor intention. The experiment consisted first of a task in which subjects piloted an airplane from the first person perspective over the ocean. The object of the task was to allow the plane to fly straight without moving the joystick until at some point there may be a perturbation in flight attitude pushing the nose of the plane toward the water. The presence of a perturbation on a trial was randomly determined. Before each trial the subject decided whether they were going to respond to a possible perturbation by pulling back on the control stick or whether they would passively observe the trial and do nothing in the case of a perturbation. Brain activity during the task was recorded using magnetoencephalography MEG. Three 10-minute sessions of the perturbation task over the ocean were conducted. An additional session was conducted in which the task for the subject was to pilot an airplane through the Grand Canyon following closely the river below. In some cases there was a perturbation of the elevator forcing the nose down. Subjects were required to recover from the perturbation without crashing and attempting to maintain tracking along the river.

The challenge is to be able to decode motor intention to an unexpected perturbation while ignoring ongoing motor control related to the tracking task. Independent component analysis was conducted on trials from the first two sessions to separate environmental and physiological artifacts from task related brain activity. For each of the 7 subjects a single independent component was found that showed an averaged evoked response to the per-

turbation occurring prior to movement of the control stick. For each trial RMS amplitude was calculated within two consecutive 40ms windows prior to the time of the peak of the averaged evoked potential and one 40ms window after. The three amplitude values served as features to train a decoder (Least-Squares Probabilistic Classification) to classify between trials of the first 2 sessions in which the pilot intentionally pulled back on the stick in response to a perturbation versus passively watching the perturbation. The spatial filter of the task related independent component and the weights of the decoder were applied to sessions 3 and 4.

The decoder was able to significantly classify the perturbation trials for which there was a motor response versus those in which there was only passive viewing on test session 3 with an average accuracy of 70%. For the Grand Canyon session the 120ms window of the decoder was incremented in 8ms steps through each trial and the first occurrence of decoded motor intention was used as the point at which adaptive automation could be implemented. Average classification of trials for which there was a perturbation versus no perturbation was 73% with an improvement in response time by implementation of the adaptive automation of 72ms. This research demonstrates that a BMI can be used to generalize to more complex novel tasks and differentiate between motor intention to an unexpected perturbation from that used during normal maneuvering. Adaptive automation can be used to significantly enhance flight performance without taking control away from the pilot.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk X

Brain Exoskeleton-Robot Interface for Rehabilitation Assistance

Dr. Tomoyuki Noda

Advanced Telecommunications Research Institute International (ATR),
Kyoto, Japan

Abstract: I have been working on developing an assistive robot system with bio-signal interfaces such as Electroencephalogram (EEG) and surface Electromyogram (sEMG), which can contribute to Brain-Machine Interface (BMI) rehabilitation. For the BMI rehabilitation, we believe EEG-Exoskeleton robot system can enhance neuro-connectivity training, where the exoskeleton robot is connected to the EEG system so that the users can control the exoskeleton robot by using their brain activities. Our exoskeleton platform combines both of pneumatic and electric energy sources to provide powerful and compliant force-controlled actuation. We consider assisting the stand-up movement which is one of the most frequently appeared movements in daily life and also a standard movement as rehabilitation training. The results show that the exoskeleton robot successfully assisted user standup movements, where the assist system was activated only by user's motor imagery.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk XI

A Waypoint-based Framework and Data-driven Decoder for Brain-Machine Interface in Smart Home Environments

Dr. Motoaki Kawanabe

Advanced Telecommunications Research Institute International (ATR),
Kyoto, Japan

Abstract: The noninvasive brain-machine interface (BMI) is anticipated to be an effective tool of communication not only in laboratory settings but also in our daily livings. The direct communication channel created by BMI can assist aging societies, the handicapped and improve human welfare. In this talk we propose and experiment a BMI framework that combines BMI with a robotic house and autonomous robotic wheelchair. Autonomous navigation is achieved by placing waypoints within the house and, from the user side, the user performs BMI to give commands to the house and wheelchair. This waypoint framework can offer essential services to the user with an effectively improved information-transfer rate. Furthermore, a data-driven decoder utilizing large databases has been developed to deal with the complex and multi-modal data acquired in the house. Open issues of our system will also be discussed.

Ugur, E., Oztop, E., Morimoto, J., and Ishii, S. (Eds) Proceedings of
IROS 2013 Workshop on Neuroscience and Robotics
"Towards a robot-enabled, neuroscience-guided healthy society"
November 3rd, 2013, Tokyo, Japan

Invited Talk XII

Brain and body machine interfaces for assistive robot technology

Joern Vogel

Institute of Robotics and Mechatronics German Aerospace Center (DLR), Wessling, Germany

Abstract: This talk will address our ongoing research is robotic embodiment and thought-based control of a humanoid robot using brain computer interface. We efficiently integrate techniques from computer vision and the task-function based control together with the brain-computer interface into an immersive and intuitive control application despite the well-known shortcomings of BCI. Our approach is based only on steady state visual evoked potential patterns. The user is fed back on-line with video stream recorded from the humanoid embedded camera. Images are then segmented, clustered from which learned objects are recognized. 3D model of recognized objects are used to superpose their computer graphic representation using augmented reality techniques. The 3D models flicker at frequencies automatically assigned by our system. Once users attention is ported on a given object, SSVEP classifier reports it. Based on the affordance concept, the object of interests associated task is sent to the stack-of-task controller of the humanoid robot. This approach is assessed in a user experiment involving several subjects who successfully controlled the HRP-2 humanoid robot in a scenario involving both grasping tasks and steering. The user experiences and the interface performances are presented and give a rich insight into future research that can be made to improve and extend such interface.