Bioinspired Head-to-Shoulder Reference Frame Transformation for Movement-Based Arm Prosthesis Control

Bianca Lento, Vincent Leconte, Lucas Bardisbanian, Emilie Doat, Effie Segas, and Aymar de Rugy

***Abstract*—Movement-based strategies are being explored as alternatives to unsatisfactory myoelectric controls for transhumeral prostheses. We recently showed that adding movement goals to shoulder information enabled Artificial Neural Networks (ANNs), trained on natural arm movements, to predict distal joints so well that transhumeral amputees could reach as with their valid arm in Virtual Reality (VR). This control relies on the object's pose in a shoulder-centered reference frame, whereas it might only be available in a head-centered reference frame through gaze-guided computer vision. Here, we designed two methods to perform the required head-to-shoulder transformation from orientation-only data, possibly available in real-life settings. The first involved an ANN trained offline to do this transformation, while the second was based on a bioinspired space map with online adaptation. Experimental results on twelve participants controlling a prosthesis avatar in VR demonstrated persistent errors with the first method, while the second method effectively encoded the transition between the two frames. The effectiveness of this second method was also tested on six transhumeral amputees in VR, and a physical proof of concept was implemented on a teleoperated robotic platform with computer vision. Those advances represent necessary steps toward the deployment of movement-based control in real-life scenarios.**

***Index Terms*—Prosthetics and exoskeletons, robust/adaptive control, reference frame transformation, human factors and human-in-the-loop, artificial neural network.**

I. Introduction

DESPITE the remarkable advancements in upper limb prosthetic devices, simultaneous myoelectric control of multiple degrees of freedom (DOFs) continues to pose challenges [1], [2]. In cases of amputation at the humeral level, research focuses on restoring elbow flexion-extension, wrist supination-pronation, and hand opening-closing [3], [4], commonly available in commercial prostheses. Yet, no myoelectric solution exists to efficiently control those 3 DOFs, to which 2 DOFs (wrist flexion-extension and radial-ulnar deviation) would still need to be added to correctly position and orient the hand so as to grasp objects in a large reachable space. This lack of efficient control over sufficient DOFs contributes to the high rate of abandonment of the prosthetic device [5], [6].

Movement-based approaches are promising alternatives whereby the control of the prosthetic limb is based on the motion of the residual limb and its natural coordination with the missing joints [7], [8], [9], [10], [11]. Grounded on these advances, we previously presented a movement-based control system which also takes the movement goal as an input [12], [13]. An Artificial Neural Network (ANN), called Proximo Contextual (PC) was designed to forecast the configuration of distal joints (red block in Fig. 1), missing to people with transhumeral limb loss, based on proximal joint motion (green block) and movement goals (blue block). Predictions of PC ANN led to remarkable results, enabling individuals with transhumeral limb loss to control a prosthesis avatar and perform pick-and-place tasks in a wide range of locations in a virtual reality (VR) environment as well as with their valid arm [13].

Deploying this control system in real-world settings requires the integration of a camera and computer vision algorithms for real-time object estimation, along with Inertial Measurement Units (IMUs) to gather data on the head, the trunk, and the arm. In this setup, information regarding the object to reach (target) might only be available in a head-centered reference frame through gaze-guided computer vision [14], [15], [16], [17] (purple block in Fig. 1). Yet, PC relies on the target expressed in a reference frame called armroot, which has its origin at the shoulder articulation and has the orientation of a sensor attached to the trunk of the operator. Furthermore, positional data needed to transform the target information from the head to the armroot reference frame may not be available as we might only have access to orientation data from IMUs (*cf* question mark in Fig. 1).

Given the complex biomechanics of the human cervical spine and shoulder girdle, and the difficulties associated with scaling a kinematic model of this region to each individual participant, relying on a traditional kinematic model to perform the required

Manuscript received 20 March 2024; accepted 11 July 2024. Date of publication 25 July 2024; date of current version 1 August 2024. This article was recommended for publication by Associate Editor V. Vashista and Editor P. Valdastri upon evaluation of the reviewers' comments. The work of Bianca Lento was supported by French Direction Général de l'Armement (DGA) under Ph.D. Grant. The work of Aymar de Rugy was supported by ANR-PRCE Grant I-Wrist under Grant ANR-23-CE19-0031-01. *(Corresponding authors: Bianca Lento; Aymar de Rugy.)*

This work involved human participants in its research. All participants provided informed written consent, and the study presented here was conducted in accordance with the principles of the Declaration of Helsinki and received approval from a local ethics committee (CPP Est II: no. 2019-A02890-57).

The authors are with University of Bordeaux, CNRS, INCIA, UMR 5287, F-33000 Bordeaux, France (e-mail: bianca.lento@u-bordeaux.fr; aymar.derugy@u-bordeaux.fr).

Digital Object Identifier 10.1109/LRA.2024.3433751

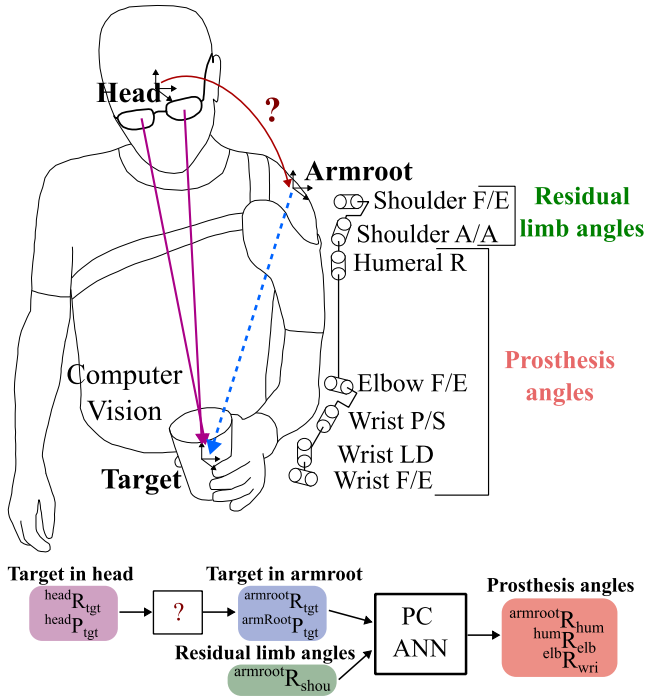

Fig. 1. Problem definition. Computer vision algorithms retrieve information about the target in the head reference frame (purple block). This information requires transformation into the armroot reference frame (blue block) to be employed, along with the residual limb angles (green block), by the PC ANN to predict the angles to apply to the prosthesis (red block). ${}^{B}R_{A}$ and ${}^{B}P_{A}$ indicate the rotation and position of reference frame A expressed in the reference frame B.

transformation from orientation-only data would merely yield a rough approximation (for a review, see [18]) which would not suit our purposes.

Here, we approached the problem from a black box perspective, our contribution being to propose and test alternative methods inspired by the brain's mechanisms for coordinate transformation and representation of peripersonal space [19], [20], [21], [22], [23], [24], [25], [26]. In a first approach, an ANN called Armroot2Head (A2H) was trained on a previously acquired database of natural head and arm movements [27], in order to predict the head position in the armroot referential from participants-morphology and ongoing head and shoulder orientation data. As errors persisted that affected the control, a second approach was designed to fine-tune in real-time a representation of the head position in the armroot reference frame that is specific to each participant, and that is encoded in a spatial map (SMAP) of the different movement-goal positions.

The present paper is organized as follows. In Section II, the proposed methods are detailed, and experimental protocols designed to test them in VR are presented. Experimental results from those protocols, conducted on participants with both valid arm (n=12) and amputated arm (n=6), are provided in Section III. Section IV briefly presents a physical proof of concept conducted on the robotic platform Reachy [28], and Section V discussed and concluded the work.

II. Materials and Methods

A. Adaptation Algorithms and Associated Control Systems

We implemented and tested two methods to derive the missing positional data of the head needed to transform target information from the head to the armroot reference frame (*cf* question mark in Fig. 1 and ${}^{armroot}\widehat{P}_{head}$ in red in Fig. 2): A2H and SMAP. The A2H is an ANN trained offline on a dataset of natural head and arm movements collected previously [27] to predict the head position in the armroot referential given the head orientation, the participant height, and the residual limb joints angles (i.e., shoulder F/E and A/A, as depicted Fig. 1). The A2H architecture consists of seven layers: an input linear layer, two hidden layers, each with 128 units and rectified linear unit activation functions, three dropout layers with a dropout rate of 0.25 to prevent overfitting (one positioned after each layer), and an output linear layer. Fig. 2(a) shows the proposed control system combining the PC ANN (red block) with A2H ANN (yellow block): the head position in the armroot referential predicted by the A2H ANN is used to transform target information from the head to the armroot reference frame, before being sent as an input to the PC ANN prosthesis control of [13].

Our second method, SMAP, draws inspiration from the brain's multisensory integration processes [19], [20], [21] and computes the head position in the armroot referential by combining information from multiple sensory modalities (i.e., computer vision from a head-mounted camera, combined with artificial neck and arm proprioception from sensors on the head, trunk and arm). Specifically, the system can compute the prosthetic hand position in the armroot referential through forward kinematics while concurrently detecting the prosthetic hand position in the head referential using computer vision. This hand position in the head referential can be aligned with the orientation of the armroot (${}^{P_{head}R_{armroot}}P_{hand}$ in Fig. 2(b)) and compared to the hand position in the armroot referential obtained through forward kinematics. The difference between the two represents the position of the head in the armroot referential. As this difference depends on both the user's morphology and the particular posture of their arm, head, and upper body while reaching targets in space, we chose to encode it in a spatial map (SMAP) accessed based on the target the user is reaching. Inspired by the brain's encoding of peripersonal space [22], [23], [24], [25], [26], this spatial encoding, illustrated in Fig. 3, was achieved by adapting the weights of a single-layer network of spatially tuned neurons online, according to the following formula:

$$w_i(t+1) = w_i(t) + \alpha \cdot g_i \cdot error$$

$$error = \left({}^{armroot}P_{hand} - {}^{P_{head}R_{armroot}}P_{hand}\right) - \sum_{i}^{N} g_i \cdot w_i$$

in which w_i is the weight of the *i-th* neuron, α is the learning rate, g_i is the firing rate of the *i-th* neuron (Gaussian activation function), N is the number of neurons, and $\sum_i^N g_i \cdot w_i(t)$ represents what the algorithm already learned.

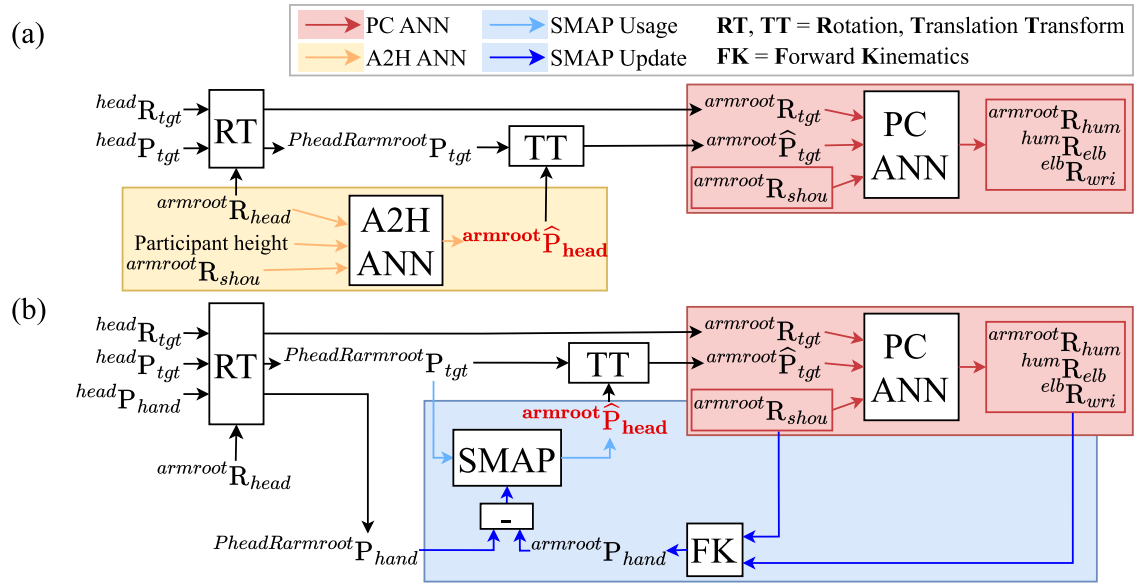

Fig. 2. Adaptation algorithms and associated control systems. **a.** Control system combining the PC ANN (red block) with the A2H ANN (yellow block). Computer vision algorithms provide information about the target in the head referential. Assuming that the head and armroot orientations are available from IMUs, we transform the target orientation from the head to the armroot referential (rotation transform (RT) block). The same rotation is applied to the target position. Then, the A2H ANN predicts the head position in the armroot referential, which is applied to the resulting target position to complete the reference frame transformation (translation transform (TT) block). The target information in the armroot referential, combined with the residual limb angles, is employed by PC ANN to predict the angles to apply to the prosthesis. **b.** Control system combining the PC ANN (red block) with the SMAP (blue block). Computer vision algorithms supply information about the target and the hand in the head referential. As in **a.**, we convert the target orientation from the head to the armroot referential and apply the same rotation to both the target and hand positions (RT block). The SMAP takes as input the resulting target position and computes the head position in the armroot reference frame. The predicted head position is then applied to the input to complete the reference frame transformation (TT block) for the PC ANN to be functional. Simultaneously, real-time weight updates occur based on the difference between the hand position detected by computer vision algorithms and aligned with the orientation of the armroot and the hand position obtained through forward kinematics. The SMAP leverages the difference between the two data to rectify its weights and correct errors.

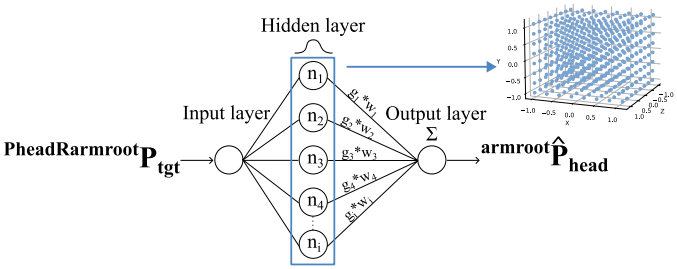

Fig. 3. SMAP algorithm structure. Each neuron is characterized by a 3D position in the input space and a Gaussian activation function centered on its position. When the target position is sent as input to the SMAP, each neuron fires according to its distance from this input, and the output is computed as the weighted sum of all firing rates.

TABLE I
EXP2 PARTICIPANTS' AMPUTATION DESCRIPTION

Id	Time since amputation	EHI pre-amputation	Amputated arm side
1	12 years	Right-handed	Right
2	14 years	Ambidextrous	Right
3	2 years	Right-handed	Right
4	37 years	Right-handed	Right
5	2 years	Left-handed	Right
6	25 years	Right-handed	Left

A tuning process was carried out to choose the values for the learning rate, the width of the Gaussian, and the distance between neurons, all of which significantly impact the algorithm's performance and efficiency. We explored different values for the three parameters by replaying the data previously collected [27] and updating the weights of the SMAP as if it were operating in real-time. The goal was to maintain spatial specificities and memory by preventing learning from spreading too widely, all while allowing sufficiently fast learning to ensure successful task completion within a single trial. This was achieved with values of 0.004 for the learning rate, and 3 cm for the width of the Gaussian and the distance between neurons.

Fig. 2(b) illustrates the proposed control system, integrating the SMAP with PC ANN alongside the online adaptation process of the SMAP.

Offline assessments of the A2H ANN revealed a mean prediction error of 2 cm. Capitalizing on the SMAP's capacity to learn from errors, we implemented and tested a third solution which combines the A2H ANN with the SMAP, employing the latter to correct errors made by the former.

B. Participants

Exp1 was conducted on twelve intact-limbs participants (four males) aged 20-28 (mean 22.83; SD 1.92), whereas Exp2 was conducted on six participants presenting a transhumeral amputation (five males) aged 34-64 (mean 49; SD 11.5). Participants' handedness was assessed using the Edinburgh Handedness Inventory (EHI, Exp 1 mean 82; SD 17) [29]. Table I provides all the information about Exp2 participants' amputation. None of

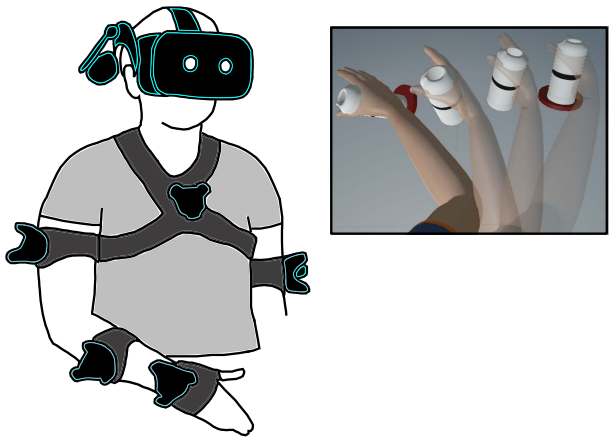

Fig. 4. Participant wearing headset and trackers, attached to the main segments of the dominant arm and the upper arm of the non-dominant arm. The task is illustrated at the top right from an egocentric point of view. It involved picking a bottle from a platform and releasing it onto another.

the participants suffered from any mental or motor disorder that could interfere with the task performance (except limb difference in Exp2).

C. Experimental Setting

The experimental set-up is similar to that of [27]: the participant sat on a stool wearing the headset (ViveTM Pro, HTC Corporation) and five (three in case of limb difference) motion trackers (ViveTM Tracker, HTC Corporation), as shown in Fig. 4.

The tracking setup received signals from the trackers and the headset at 90 Hz (sampling rate) using SteamVR (Valve Corporation) as middleware. Unity simulation engine (Unity Technologies) was used to manage the virtual scene's contents and interactions with the participant.

D. Task

The participants performed a pick-and-place task involving repeatedly picking a bottle from one platform and releasing it onto another (Fig. 4). A transparent bottle was added to the virtual hand or on the platform to facilitate the task. When the position and orientation of the virtual hand matched the bottle, within a positional tolerance of 3 cm and an angular tolerance of 10°, the bottle turned red. To grasp/release the bottle and complete the task, the participant had to press a key on a keyboard while the bottle was red. To avoid long and tedious experiments in case of control difficulties, a time restriction of 6 s per trial was applied, defined as the time from a successful pick to a successful place or vice versa.

E. Target Set Generation

The target set was generated using our previously acquired database [27] and adapting the data to the participant morphology. A generator algorithm drew at random 7-DOFs arm angular

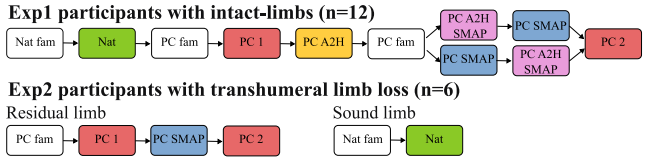

Fig. 5. Protocols of experiment 1 and 2.

configurations following a multivariate uniform probability distribution. The targets were then obtained through forward kinematics and filtered based on the criteria used in [27], excluding targets too close to the participant's trunk or legs and targets pointing downward. This process was iterated until a target set of 100 targets was obtained. For Exp2, a supplementary filter, which took into account the limits of the participant's residual limb, was added.

F. Virtual Arm Calibration

Once the participant was equipped with the trackers and the headset, following the procedure outlined in [27], we proceeded with the same calibration process to link the virtual arm with the participant's arm. To this end, we recorded data while the participant executed slow movements, using all DOFs of both arms for 15 s, followed by the head for 10 s. Based on this data, we employed the techniques detailed in [27] to estimate the joint centers' locations. The dimensions of the virtual arm were resized according to the participant's own dimensions, calculated as the difference between the estimated joint centers. To complete the calibration, the participant had to overlap an arm made of sticks and spheres, representing their arm, with the virtual arm locked in a reference posture. Additional details can be found in [13], [27].

G. Protocol

Experimental protocols are shown in Fig. 5. Familiarization phases ("fam") were designed to familiarize the participants with the environment and virtual arm controls. In these phases, the target set corresponded to the second half of the main target set. The other keywords represent the control used in that specific phase: "Nat" stands for natural control, whereby all joints of the virtual arm follow the user's actual arm movements. "PC" stands for Proximo Contexual, the prosthesis control implemented in [13]: the proximal joints of the virtual arm mimicked the participant's proximal joints, while PC ANN predictions were used to operate on the distal joints. "SMAP" and "A2H" stand for the use of the SMAP and A2H algorithms and their associated control systems (*cf* Fig. 2). In these phases, constraints of a real-world setting were simulated by computing target and hand information in the head reference frame "as if" obtained from computer vision, and then relying solely on orientation data "as if" obtained from IMUs.

Exp1 included nine phases and aimed to compare the performances of the proposed methods with that of an ideal case (PC), in which the change of reference frame from the head to the armroot was performed with high precision from all the information available in the VR system. A first block (PC 1)

was conducted early in both protocols for participants to get used to this ideal control before being tested in our alternative controls, and a second block (PC 2) was conducted at the end of both protocols to ensure comparison with the ideal control once participants were fully accustomed with the device and controls.

Offline tests of the A2H algorithm revealed a mean error in prediction of about 2 cm. To assess how much this error might affect the control, we tested the control condition illustrated in Fig. 2 a using A2H predictions (PC A2H). A "washout" phase, during which the participants reverted to using the ideal PC control again (PC fam), was then included to wash out possible compensations associated with this control before testing other controls.

A crucial aspect of the SMAP we aimed to assess was its memory. Recognizing that the learning process would take time and that the initial performances might not be optimal, we repeated the target set twice in PC SMAP and PC A2H SMAP. We anticipated that the second attempt would yield results closer to the ideal case, as the algorithm would have assimilated the necessary corrections. Moreover, the order between PC SMAP and PC A2H SMAP was counterbalanced among participants.

Exp2 included six phases: four involving the residual limb followed by two involving the sound limb. Building on the insights from Exp1, which demonstrated that the SMAP alone was sufficient for the task, we refined the protocol and compared only the SMAP with the ideal case PC.

III. Results

The proposed methods were assessed based on the following objective metrics: success rate, movement time, and shoulder spread volume. The first refers to the percentage of successful trials achieved; the second to the time taken to reach and validate each target; the latter refers to the volume of the ellipsoidal region containing 97% of the shoulder positions recorded within a phase, as a proxy of postural stability and body compensation [12]. Considering the high success rates obtained in all the phases (medians above 90%), movement times and shoulder spread volumes were calculated on successful trials only. Furthermore, after testing for the first time each control (excluding the familiarisation phases), participants provided qualitative feedback on their experiences with each control strategy through two surveys, the System Usability Scale (SUS) [30], [31] and the first part of the Prosthesis Task Load indeX (rawProsTLX) [32]. These surveys assess usability and cognitive load, respectively, providing a comprehensive understanding of user perceptions, mental effort, and overall satisfaction with the control system. A score for each survey and each control was obtained. A higher SUS score and a lower rawProsTLX score suggested that the control method is intuitive and demands minimal effort.

The statistical analysis was conducted using the R software with a significance level set at 0.05 and applying a Bonferroni correction for multiple comparisons. Based on the results of the normality test (Shapiro-Wilk) and the test for homogeneity of variances (Maulchy's Test), a repeated measures ANOVA or a Friedman test was performed for movement time and shoulder spread volume. If a significant difference was detected at this stage, post-hoc tests were carried out (Tukey's or Conover's test, respectively). For success rate, a Cochran's Q test was performed, followed by a McNemar test if a significant difference was detected.

Fig. 6 illustrates the results of Exp1 (a-e) and Exp2 (f-j). For the objective metrics, in the phases PC SMAP and PC A2H SMAP, the two repetitions of the target set are analyzed separately. With PC A2H, the lower success rates, increased movement times, and increased shoulder spread volumes all indicate that the error in the A2H predictions does affect the control. Nevertheless, as the prediction error is small, the combination of A2H with SMAP results in a control system that is reliable and accurate in executing the reference frame transformation, obtaining performances similar to the ideal case PC. Furthermore, as the SMAP has to learn small corrections, there is mainly no difference between the repetitions of the target set (PC A2H SMAP B1 and PC A2H SMAP B2).

In PC SMAP B1, movement times and shoulder spread volumes are statistically higher than in PC due to the initial learning phase of the SMAP algorithm, during which it has to assimilate substantial errors. The increased movement times and shoulder spread volumes probably reflect the adaptation period required for the system to learn and memorize the necessary corrections. This is confirmed by the fact that the results of PC SMAP B2 (the second repetition of the same target set) are drastically improved. The learning process required time, but once the corrections were memorized, its performances closely resembled those of PC.

In Exp2, movement times tended to be longer, and shoulder spread volumes to be higher than in Exp1. Nevertheless, when comparing the different control strategies, PC SMAP B2 also exhibited performances similar to PC across all metrics. As in Exp1, the increased movement times, shoulder spread volumes, and lower success rates in PC SMAP B1 probably reflect the adaptation period required for the algorithm to learn the necessary corrections.

Video 1 illustrates a participant with transhumeral limb loss using PC ANN with SMAP. The orange and green virtual cylinders, not visible to the participant, represent the object transformed from the camera to the armroot reference frame without and with the application of the SMAP prediction. The green cylinder acts as a dynamic representation of the convergence status of the SMAP. As the algorithm learns the proper corrections, its position is progressively adjusted toward the actual object position. In the second part of the video, leveraging its past learning, the algorithm applies the corrections it has already acquired, resulting in the green cylinder being positioned closer to the actual object.

The results of the objective metrics are corroborated by the subjective assessments (SUS score and rawProsTLX score). In this context, the phase Nat serves as the baseline, as the virtual arm mimicked the participant's actual arm movements. In Exp1, the only control strategy exhibiting a significant difference from the baseline is PC A2H. As the error in prediction did affect the control, participants necessitated additional physical and mental effort to achieve the task. Both intact limb participants and participants with limb loss evaluated control strategies involving

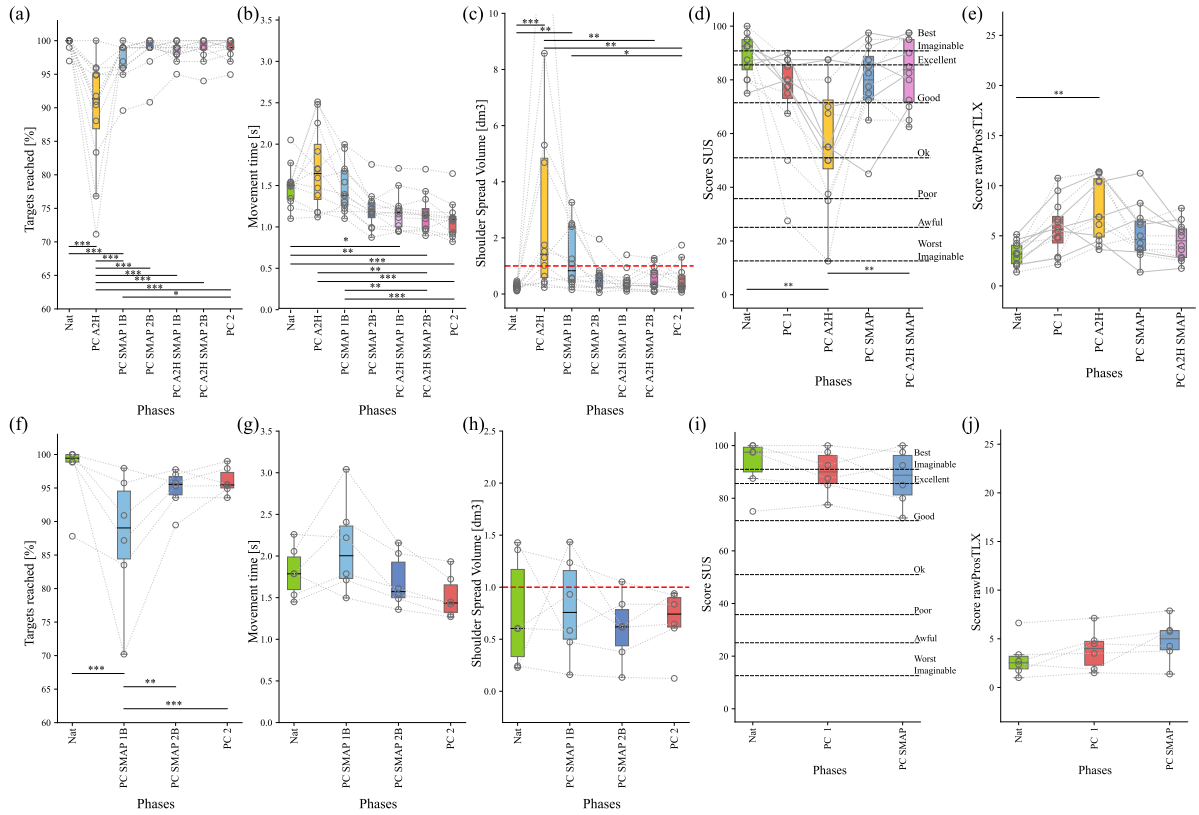

Fig. 6. Data analysis results. Individual data are represented by hollow dots and dotted lines. Stars represent significant differences between control strategies with * for $p < 0.05$, ** for $p < 0.01$ and *** for $p < 0.001$. Results of Exp1 are presented for each phase. **a**. Success rate (Cochran's Q test $Q = 304.45$, $p < 0.001$). **b**. Movement time (Friedman chi.sq $= 56.57$, $p < 0.001$). **c**. Shoulder spread volume (Friedman chi.sq $= 36.85$, $p < 0.001$). The dashed red line represents a volume of 1 dm^3(=1 L). **d**. SUS score (Friedman chi.sq $= 19.68$, $p < 0.001$). **e.** rawProsTLX score (Friedman chi.sq $= 18.02$, $p < 0.01$). Results of Exp2 are presented for each phase. **f**. Success rate (Cochran's Q test $Q = 45.57$, $p < 0.001$). **g**. Movement time (RM ANOVA DFn $= 3$, DFd $= 15$, $F = 6.15$, $p = 0.006$, ges $= 0.277$). **h**. Shoulder spread volume (Friedman chi.sq $= 1$, $p = 0.801$). The dashed red line represents a volume of 1 dm^3(=1 L). **i**. SUS score (Friedman chi.sq $= 4.66$, $p = 0.097$). **j**. rawProsTLX score (Friedman chi.sq $= 7$, $p = 0.03$).

SMAP with scores similar to the PC and Nat phases. This indicates that our control system, including the SMAP, exerts minimal influence on the virtual arm's behavior, thereby maintaining natural reaching movements. Furthermore, participants from Exp2 reported high satisfaction with the control system and expressed a keen interest in utilizing such natural control methods with real prostheses.

IV. Proof of Concept

Although Exp1 and Exp2 tested conditions with reduced information "as if" obtained from computer vision, the feasibility of employing SMAP in a real-life setting remains to be established. To do so, we used the robotic platform REACHY [28], which has been recently upgraded to include stronger arm and head actuators and enable teleoperation from an egocentric point of view (*cf* Fig. 7 and Video 2). In this setup, a Zed Mini camera, mounted on the head of the robot and connected to the VR headset, enabled the user wearing the headset to control the robot from an egocentric point of view.

Although determining the head position in the armroot referential is straightforward with a robot as its kinematic model is known, we aimed to verify that the SMAP is effective in this context whereby the hand position and orientation needed for the algorithm genuinely come in the head referential from computer vision.

One intact-limbs participant (male, aged 26, EHI 100), who had prior experience with the robotic platform, was instructed to control the robot to reach and grasp a block of five polyethylene foam cylinders (r = 1.8 cm, h = 11 cm) placed on a wooden board in various positions and orientations. The robot was controlled by applying the participant's arm joint configuration to its arm (teleoperation). ArUco markers, placed on both the robot's end effector and the cylinders, combined with user gaze data, enabled real-time and highly accurate object pose estimation.

The protocol included two phases: reaching at a natural pace (Nat Pace) and Grasping on Convergence (GoC), with the latter being recorded in Video 3. Visual feedback, presented as virtual cylinders, was provided to the participant: the red cylinder, representing the object detected by computer vision in the camera referential, was transformed in the armroot referential without and with the application of the SMAP prediction (the orange and green cylinder, respectively). As the algorithm learned the proper corrections, the green cylinder adapted its position toward the actual object position.

In Nat Pace, the participant performed the task at his natural pace whereas in the GoC, the participant executed the block twice (B1 and B2) and was instructed to grasp the object only

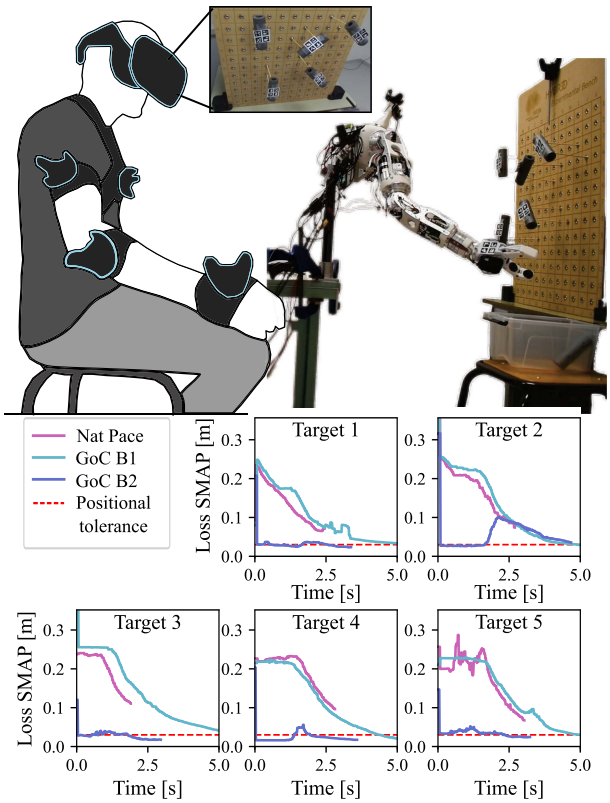

Fig. 7. Proof of Concept. Experimental setup (on the top) and algorithm's loss (on the bottom). Each subplot represents a target (a cylinder to grasp), while each color represents a different phase or block repetition.

once the algorithm had converged (i.e., when he saw the virtual green cylinder on the object). Indeed, we were not sure that convergence would be achieved at a natural pace.

Fig. 7 shows the loss of the algorithm, computed as the difference between the actual value of the head position in the armroot referential and the output of the SMAP. As shown, the SMAP algorithm effectively operates in real-world situations. It requires some time to learn and memorize the necessary corrections (GoC B1), as already established in the experiments in VR, but once it has learned and memorized them, its performances improve dramatically (GoC B2).

V. Discussion

In this letter, we have presented and tested two methodologies to perform the head-to-shoulder transformation required for deploying in real-world settings our movement-based prosthesis control [13]. Rather than relying on a kinematic model of the biomechanically complex region encompassing the human spine and shoulder girdle, which would demand extensive individual calibration, we approached the problem as a black box by learning its input-output relationship.

The first method, A2H, proved ineffective in resolving this issue: the ANN, trained offline on a database of natural movements by multiple participants, failed to encapsulate sufficiently well the specificity of the user's morphology, resulting in a mean prediction error of 2 cm which was found to affect the prosthesis control. The second method, SMAP, elicited much better results, as it effectively encoded the transition from the head to the armroot associated with different targets in space. This bioinspired spatial coding [19], [20], [21], [22], [23], [24], [25], [26] enabled to capture online features specific to the user's morphology, as well as to the particular head, body, and arm postures employed when reaching various targets in space. Although the learning process required an adaptation period, subsequent performance on already explored targets was comparable to the ideal scenario, as observed from the twelve participants with intact limbs tested in a condition of simulated amputation in Exp1.

While the combination of both methods yields favorable results, the second approach alone suffices for the task, as shown on six participants with transhumeral limb loss in Exp2, eliminating the need for unnecessary complexity in the control system.

In addition to the effectiveness of SMAP in performing the required head-to-shoulder transformation in a VR environment, we also provided a physical proof of concept in a simple real-world scenario involving the teleoperation of the robotic platform REACHY. Instead of being known by design as in the VR system, the end effector information was gathered from computer vision in this proof of concept. Although this brings additional challenges associated with missing visual data, such as when outside the field of vision or obstructed, the memory associated with our spatial coding in SMAP proved robust to such situations, ensuring continued smooth operation.

Combining the hand pose from vision and forward kinematics to compute the head position in the armroot reference frame bears similarities to self-observation methods used for self-calibrating the parameters of the kinematic chain in robotics [22], [33], [34]. Yet, instead of learning constant offsets between consecutive frames as in the cited papers, SMAP learns a 3D position that varies according to the user's morphology and the specific posture of the user's arm, head, and upper body while reaching for targets in space. Therefore, this variable is encoded within a map of the peripersonal space, which enables accessing it through the target the user gazes at and endeavors to grasp.

While our proof of concept has shown that using the SMAP is feasible in a simplified real-world scenario, the lack of coupling between the robotic device and the user remains a limitation of this work. Overcoming this limitation necessitates a prosthetic device compatible with testing our control system, featuring five active joints–a capability currently absent in commercial options.

In future directions, we also aim to evaluate the algorithm's effectiveness in settings closer to real-life scenarios, using eye-tracking and robust computer vision to gather information about the prosthetic hand and the target object. In addition to advances in prosthesis control, the findings presented here also provide insights into broader applications of the SMAP. These include diverse scenarios involving reference frame transformations, online adaptation, or camera misplacement, enabling corrections in both position and orientation.

VI. Videos

Video 1: https://youtu.be/IsEasynQHkk
A transhumeral amputee using PC ANN with SMAP.

Video 2: https://youtu.be/7GQemjYP5rI
Reachy V2 teleoperated from an egocentric point of view.
Video 3: https://youtu.be/SQNKJrkQBXE
SMAP adaptation on Reachy.

REFERENCES

[1] J. M. Hahne, S. Dahne, H.-J. Hwang, K.-R. Muller, and L. C. Parra, "Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 618–627, Jul. 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7038151/

[2] M. Nowak, R. M. Bongers, C. K. V. D. Sluis, A. Albu-Schäffer, and C. Castellini, "Simultaneous assessment and training of an upper-limb amputee using incremental machine-learning-based myocontrol: A single-case experimental design," *J. NeuroEngineering Rehabil.*, vol. 20, no. 1, 2023, Art. no. 39. [Online]. Available: https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-023-01171-2

[3] M. Ortiz-Catalan, E. Mastinu, P. Sassu, O. Aszmann, and R. Brånemark, "Self-contained neuromusculoskeletal arm prostheses," *New England J. Med.*, vol. 382, no. 18, pp. 1732–1738, 2020, doi: 10.1056/NEJMoa1917537.

[4] L. J. Hargrove, L. A. Miller, K. Turner, and T. A. Kuiken, "Myoelectric pattern recognition outperforms direct control for transhumeral amputees with targeted muscle reinnervation: A randomized clinical trial," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 13840. [Online]. Available: https://www.nature.com/articles/s41598-017-14386-w

[5] S. Salminger et al., "Current rates of prosthetic usage in upper-limb amputees–have innovations had an impact on device acceptance," *Disabil. Rehabil.*, vol. 44, no. 14, pp. 3708–3713, 2022, doi: 10.1080/09638288.2020.1866684.

[6] E. A. Biddiss and T. T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," *Prosthetics Orthotics Int.*, vol. 31, no. 3, pp. 236–257, 2007. [Online]. Available: https://journals.lww.com/00006479-200731030-00003

[7] M. Desmurget et al., "Postural and synergic control for three-dimensional movements of reaching and grasping," *J. Neurophysiol.*, vol. 74, no. 2, pp. 905–910, 1995. [Online]. Available: https://www.physiology.org/doi/10.1152/jn.1995.74.2.905

[8] M. Popovic and D. Popovic, "Cloning biological synergies improves control of elbow neuroprostheses," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 1, pp. 74–81, Jan./Feb. 2001. [Online]. Available: https://ieeexplore.ieee.org/document/897830/

[9] M. Merad et al., "Assessment of an automatic prosthetic elbow control strategy using residual limb motion for transhumeral amputated individuals with socket or osseointegrated prostheses," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 1, pp. 38–49, Feb. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8972569/

[10] R. R. Kaliki, R. Davoodi, and G. E. Loeb, "Evaluation of a noninvasive command scheme for upper-limb prostheses in a virtual reality reach and grasp task," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 792–802, Mar. 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6138294/

[11] F. Montagnani, M. Controzzi, and C. Cipriani, "Exploiting arm posture synergies in activities of daily living to control the wrist rotation in upper limb prostheses: A feasibility study," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 2462–2465. [Online]. Available: https://ieeexplore.ieee.org/document/7318892/

[12] S. Mick et al., "Shoulder kinematics plus contextual target information enable control of multiple distal joints of a simulated prosthetic arm and hand," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, pp. 1–17, 2021. [Online]. Available: https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-020-00793-0

[13] E. Segas et al., "Intuitive movement-based prosthesis control enables arm amputees to reach naturally in virtual reality," *Elife*, vol. 12, 2023, Art. no. RP87317. [Online]. Available: https://elifesciences.org/articles/87317

[14] Y. Liu, "Gen6d: Generalizable model-free 6-DoF object pose estimation from RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 298–315.

[15] V. Nguyen, Y. Du, Y. Xiao, M. Ramamonjisoa, and V. Lepetit, "PIZZA: A powerful image-only zero-shot zero-CAD approach to 6 DoF tracking," in *Proc. Int. Conf. 3D Vis.*, 2022, pp. 515–525.

[16] P. P. D. S. Roman, J. Benois-Pineau, J.-P. Domenger, F. Paclet, D. Cataert, and A. D. Rugy, "Saliency driven object recognition in egocentric videos with deep CNN: Toward application in assistance to neuroprostheses," *Comput. Vis. Image Understanding*, vol. 164, pp. 82–91, 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1077314217300462

[17] I. González-Dí-az, J. Benois-Pineau, J.-P. Domenger, D. Cattaert, and A. D. Rugy, "Perceptually-guided deep neural networks for ego-action prediction: Object grasping," *Pattern Recognit.*, vol. 88, pp. 223–235, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0031320318304011

[18] R. Krishnan, N. Björsell, E. M. Gutierrez-Farewik, and C. Smith, "A survey of human shoulder functional kinematic representations," *Med. Biol. Eng. & Comput.*, vol. 57, no. 2, pp. 339–367, 2019. [Online]. Available: https://link.springer.com/10.1007/s11517-018-1903-3

[19] J. G. Makin, M. R. Fellows, and P. N. Sabes, "Learning multisensory integration and coordinate transformation via density estimation," *PLoS Comput. Biol.*, vol. 9, no. 4, 2013, Art. no. e1003035. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.1003035

[20] A. Pouget, S. Deneve, and J.-R. Duhamel, "A computational perspective on the neural basis of multisensory spatial representations," *Nature Rev. Neurosci.*, vol. 3, no. 9, pp. 741–747, 2002. [Online]. Available: https://www.nature.com/articles/nrn914

[21] G. Pugach, A. Pitti, O. Tolochko, and P. Gaussier, "Brain-inspired coding of robot body schema through visuo-motor integration of touched events," *Front. Neurorobot.*, vol. 13, 2019, Art. no. 5. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnbot.2019.00005/full

[22] M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer, "Body schema in robotics: A review," *IEEE Trans. Auton. Ment. Develop.*, vol. 2, no. 4, pp. 304–324, Dec. 2010. [Online]. Available: https://ieeexplore.ieee.org/document/5601749/

[23] E. Chinellato, M. Antonelli, B. J. Grzyb, and A. P. D. Pobil, "Implicit sensorimotor mapping of the peripersonal space by gazing and reaching," *IEEE Trans. Auton. Ment. Develop.*, vol. 3, no. 1, pp. 43–53, Mar. 2011. [Online]. Available: https://ieeexplore.ieee.org/document/5703113/

[24] M. Couraud, D. Cattaert, F. Paclet, P. Y. Oudeyer, and A. D. Rugy, "Model and experiments to optimize co-adaptation in a simplified myoelectric control system," *J. Neural Eng.*, vol. 15, no. 2, 2018, Art. no. 026006. [Online]. Available: https://iopscience.iop.org/article/10.1088/1741-2552/aa87cf

[25] G. D. Pellegrino and E. Làdavas, "Peripersonal space in the brain," *Neuropsychologia*, vol. 66, pp. 126–133, 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0028393214004230

[26] S. Lallee and P. F. Dominey, "Multi-modal convergence maps: From body schema and self-representation to mental imagery," *Adaptive Behav.*, vol. 21, no. 4, pp. 274–285, 2013. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1059712313488423

[27] B. Lento et al., "3d-ARM-gaze: A public dataset of 3 D arm reaching movements with gaze information in virtual reality," 2024, doi: 10.1101/2024.01.30.577386.

[28] S. Mick et al., "Reachy, a 3D-printed human-like robotic arm as a testbed for human-robot control strategies," *Front. Neurorobot.*, vol. 13, 2019, Art. no. 65[Online]. Available: https://www.frontiersin.org/article/10.3389/fnbot.2019.00065/full

[29] R. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971, doi: 1016/0028-3932(71)90067-4.

[30] G. Gronier and A. Baudet, "Psychometric evaluation of the f-SUS: Creation and validation of the french version of the system usability scale," *Int. J. Hum.–Comput. Interact.*, vol. 37, no. 16, pp. 1571–1582, 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10447318.2021.1898828

[31] A. Bangor, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.

[32] J. V. V. Parr et al., "A tool for measuring mental workload during prosthesis use: The prosthesis task load index (PROS-TLX)," *Plos One*, vol. 18, no. 5, 2023, Art. no. e0285382. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0285382

[33] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella, "Adaptive body schema for robotic tool-use," *Adv. Robot.*, vol. 20, no. 10, pp. 1105–1126, 2006. [Online]. Available: https://www.tandfonline.com/doi/full/10.1163/156855306778522550

[34] J. Sturm, C. Plagemann, and W. Burgard, "Body schema learning for robotic manipulators from visual self-perception," *J. Physiol.-Paris*, vol. 103, no. 3, pp. 220–231, 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0928425709000461