Transferring Human Impedance Behavior to Heterogeneous Variable Impedance Actuators

Matthew Howard, David J. Braun, and Sethu Vijayakumar

***Abstract*—This paper presents a comparative study of approaches to control robots with variable impedance actuators (VIAs) in ways that imitate the behavior of humans. We focus on problems where impedance modulation strategies are recorded from human demonstrators for transfer to robotic systems with differing levels of heterogeneity, both in terms of the dynamics and actuation. We categorize three classes of approach that may be applied to this problem, namely, 1) direct, 2) feature-based, and 3) inverse optimal approaches to transfer. While the first is restricted to highly biomorphic plants, the latter two are shown to be sufficiently general to be applied to various VIAs in a way that is independent of the mechanical design. As instantiations of such transfer schemes, 1) a constraint-based method and 2) an apprenticeship learning framework are proposed, and their suitability to different problems in robotic imitation, in terms of efficiency, ease of use, and task performance, is characterized. The approaches are compared in simulation on systems of varying complexity, and robotic experiments are reported for transfer of behavior from human electromyographic data to two different variable passive compliance robotic devices.**

***Index Terms*—Behavior transfer, imitation learning, passive impedance control, variable stiffness actuation.**

I. Introduction

In RECENT years, variable impedance actuation has become increasingly popular in the design and control of novel robotic mechanisms [1], [2]. Variable impedance actuators (VIAs) promise many benefits for the next generation of robots, including 1) increased safety in settings where there is human–robot interaction [3], 2) increased dynamic range (e.g., when throwing, energy may be stored in spring-like VIAs, before being released explosively for the throw [4]), and 3) increased robustness when interacting with the environment [5]. Despite these benefits, however, a number of challenges remain associated with the deployment of such actuators to the current generation of robots. One major issue is that of how to control such mechanisms, and in particular, how to best utilize variable impedance so that the benefits (such as compliance) are exploited, while compromise on other aspects of performance (such as precision) is avoided.

A promising approach to finding appropriate impedance control strategies on robots is to take examples from human behavior and attempt to mimic it. The human musculoskeletal system, which is actuated by antagonistic muscles with inherent viscoelastic properties [6], represents one of the best examples of a system controlled with variable impedance actuation. A large body of research studying human impedance modulation exists in the biological literature and, as such, may be a rich source of inspiration for designing controllers for robots [7].

In order to exploit these biological insights for control of robotic VIAs, a number of technical problems must be addressed. A problem of primary concern is the heterogeneity in the kinematics, dynamics, and actuation between the human musculoskeletal system and robotic VIAs. This affects the transfer of impedance behavior, both in terms of the *control* of the variable impedance device and *strategy* employed in achieving task goals.

More concretely, *control* of robotic VIAs to mimic human impedance behavior remains a challenging problem. The control of impedance in the human musculoskeletal system can be achieved by cocontraction of groups of antagonistic muscles. By building robotic actuators with a similar antagonistic layout [8], one can simplify the imitation task [e.g., by drawing a direct correspondence between human electromyography (EMG) signals and actuator commands]. However, often such designs are unfavorable since they tend to have rather complex, coupled dynamics and can be hard to build into multijoint devices. Other proposed designs have focused on simplifying the dynamics (and thereby the control) [2] or improving scalability [9], [10]. These often have several benefits, such as compactness, but the difficulty then lies in finding appropriate controllers, especially when trying to mimic the capabilities of humans [11] and exploit the benefits of variable impedance.

On the other hand, some impedance *strategies*, which are employed by humans, are highly adapted to certain specific properties of the human body and may not transfer directly to those of robotic plants. For example, it is well known that the human musculoskeletal system suffers from signal-dependent noise, that is, noise in the kinematics of movement in direct proportion to the control signal [12]. To counter the effects of signal-dependent noise, humans adapt their impedance in different ways, depending on the task, e.g., in tasks requiring high precision, humans tend to increase impedance by cocontracting [13]. However, most robotic systems do not suffer from such

Manuscript received April 13, 2012; revised October 25, 2012; accepted March 26, 2013. Date of publication April 30, 2013; date of current version August 2, 2013. This paper was recommended for publication by Associate Editor A. Albu-Schäffer and Editor G. Oriolo upon evaluation of the reviewers' comments. This work was supported by the European Union Seventh Framework Programme as part of the STIFF and TOMSY projects.

M. Howard is with the Nakamura Lab, Department of Mechano-Informatics, University of Tokyo, Tokyo 113-8654, Japan (e-mail: m_howard@ynl.t.u-tokyo.ac.jp).

D. J. Braun and S. Vijayakumar are with the Institute of Perception Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, EH1 2QL, U.K. (e-mail: david.braun@ed.ac.uk; sethu.vijayakumar@ed.ac.uk).

This paper has downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2013.2256311

noise characteristics (e.g., noise is more commonly constant, additive, and much smaller in magnitude); therefore, direct transfer of the human impedance strategy may not be reasonable: maintaining the same level of stiffness on a less noisy robot would waste energy and reduce compliance without significantly improving accuracy.

To overcome such problems, in this paper, we suggest two approaches to the problem of transferring impedance control strategies across plants with *heterogeneous dynamics and actuation*.

The first is a scheme in which human impedance characteristics can be directly tracked on a robotic VIA. We employ a closed-loop tracking scheme (first proposed in [14]) and illustrate its use in the context of imitation. In particular, we focus on the issue of transfer of impedance (i.e., "impedance matching") between different systems with this approach, and demonstrate its use in the context of online teleoperation of robotic VIAs. This can be considered as imitation at the *control* level, i.e., building controllers that achieve the same impedance behavior across heterogeneous systems.

The second approach is to employ inverse optimal control (OC) to seek the objective of demonstrated behavior in the form of a performance measure (cost function) [15]. We use recordings of task-oriented human behavior in which the impedance strategy employed is (assumed to be) optimized with respect to his or her dynamics. By applying apprenticeship learning (AL) [16], [17], we show how the underlying optimization criteria, which are used by the human to shape their impedance strategy, can be extracted and transferred to design impedance strategies that are suitable to different (heterogeneous) variable impedance robots. We demonstrate and compare these approaches in simulation and through human/robot experiments.

II. Problem Definition

Our aim is to transfer behavior of an expert demonstrator (e) to an apprentice learner (l) given that the expert and learner have a very different embodiment,[1] both in terms of their dynamics and actuation. Specifically, we assume the expert has state ${}^e\mathbf{x} \in \mathbb{R}^m$, controls movement with commands ${}^e\mathbf{u} \in \mathbb{R}^n$, and has dynamics

$$ {}^e\dot{\mathbf{x}} = {}^e\mathbf{f}({}^e\mathbf{x}, {}^e\mathbf{u}) \quad \in \mathbb{R}^m. \tag{1} $$

Note that the effect of the commands ${}^e\mathbf{u}$ on the dynamics (i.e., the form of ${}^e\mathbf{f}(\cdot)$) depends on the actuation mechanism of the expert. In particular, we can rewrite (1) as

$$ {}^e\dot{\mathbf{x}} = {}^e\mathbf{g}({}^e\mathbf{x}, {}^e\boldsymbol{\tau}) \quad \in \mathbb{R}^m $$

where ${}^e\boldsymbol{\tau} = {}^e\boldsymbol{\tau}({}^e\mathbf{x}, {}^e\mathbf{u})$ is the (in general, state-dependent) relationship between the expert's command signal ${}^e\mathbf{u}$ and the torques/forces applied by the expert's actuators.

Our goal is to transfer behavior to a learner with a different embodiment, both in terms of the dynamics and actuation. For example, we may wish to take impedance behavior measured from the human arm (actuated by antagonistic muscles) and transfer it to a robotic manipulator (actuated by VIAs). We denote the learner's state as ${}^l\mathbf{x} \in \mathbb{R}^r$, command signal ${}^l\mathbf{u} \in \mathbb{R}^s$, and dynamics

$$ {}^l\dot{\mathbf{x}} = {}^l\mathbf{f}({}^l\mathbf{x}, {}^l\mathbf{u}) = {}^l\mathbf{g}({}^l\mathbf{x}, {}^l\boldsymbol{\tau}) \quad \in \mathbb{R}^r \tag{2} $$

where ${}^l\boldsymbol{\tau} = {}^l\boldsymbol{\tau}({}^l\mathbf{x}, {}^l\mathbf{u})$ denotes the torques produced by the learner's actuators. Note that, in general, the state and action space (${}^e\mathbf{x}, {}^e\mathbf{u}$ and ${}^l\mathbf{x}, {}^l\mathbf{u}$) may differ significantly between the two plants: For example, for a human expert, ${}^e\mathbf{u}$ may correspond to *muscle activations*, whereas for a robot learner, ${}^l\mathbf{u}$ may correspond to *desired positions of a set of servomotors*. The state of the robot ${}^l\mathbf{x}$ may be sufficiently described by the joint angles and positions (${}^l\mathbf{x} = ({}^l\mathbf{q}^\top, {}^l\dot{\mathbf{q}}^\top)^\top$), while that of a human demonstrator may include additional biomechanical variables (e.g., tendon slack lengths, muscle pennation angles, etc. [18]). In addition, ${}^l\mathbf{f}(\cdot)$ and ${}^e\mathbf{f}(\cdot)$ may also differ, both in terms of the parameter values (e.g., inertia, link lengths, joint axis positions, and orientations), as well as in their parametric form.

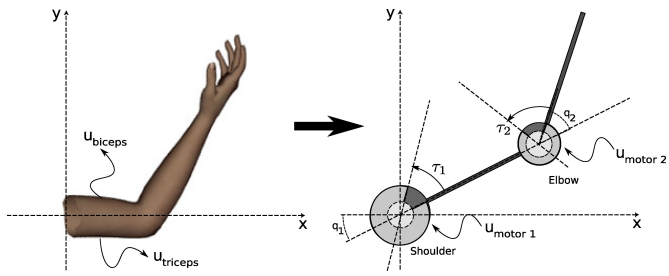

Fig. 1. Correspondence problem between human and robotic actuation systems. (Left) Humans use muscle activations (e.g., u_{triceps} and u_{biceps}) to control movement. (Right) Robotic systems are controlled with command signals to the different motors (e.g., u_{motor1} and u_{motor2}). The torque generated by those motors depends on the actuators used.

Clearly, the differences in embodiment between demonstrator and learner cause numerous difficulties when attempting to transfer behavior. As an example, consider the problem of transferring the control strategy used by a human to perform some task to a robotic imitator, as illustrated in Fig. 1. Imagine that we are given a set of recordings of the behavior (e.g., in the form of muscle activation profiles), and we wish to use this data to reproduce the movement on a robotic system. Depending on the hardware, there are a number of approaches that we may take (see Fig. 2). In the following, we characterize these approaches and the domains to which they are applicable.

A. Direct Imitation for Biomorphic Systems

First, if there is a close correspondence between the robot and the human, the simplest approach is the *direct imitation of behavior*. In the case of open-loop imitation, one would define the *correspondence* ${}^e\mathbf{u} \equiv {}^l\mathbf{u}$ (and, therefore, $s = n$) and execute commands

$$ {}^l\mathbf{u}(t) = {}^e\mathbf{u}(t) \quad \in \mathbb{R}^n. \tag{3} $$

[1] In principle, we avoid making any assumption on the extent to which the expert and learner plants may differ. However, in order to make a meaningful comparison between their respective behaviors, we assume that there is a sufficient overlap in their capabilities that they may both achieve similar success at a given task.

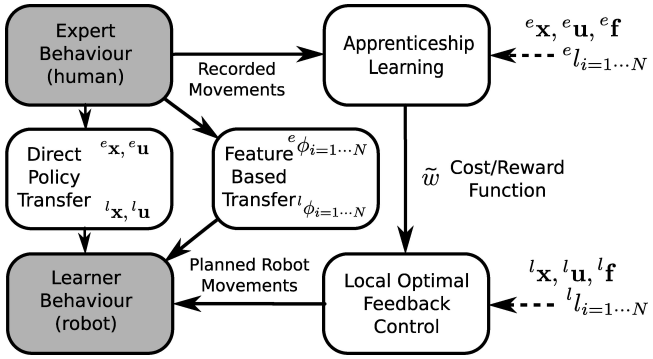

Fig. 2. Routes to behavior imitation. Starting with recordings of the expert (human) behavior, we can identify three ways in which behavior can be transferred. The first is to do a direct policy transfer, i.e., make a direct correspondence between the human state and commands ${}^e\mathbf{x}, {}^e\mathbf{u}$ and feed those directly as commands to the robot ${}^l\mathbf{x}, {}^l\mathbf{u}$. The second is to record features of the expert behavior ${}^e\boldsymbol{\phi}({}^e\mathbf{x}, {}^e\mathbf{u})$ (e.g., the stiffness profile during movement) and attempt to track these with corresponding features of the robot ${}^l\boldsymbol{\phi}({}^l\mathbf{x}, {}^l\mathbf{u})$. The third is to take an inverse optimal approach, in which recordings of task-oriented behavior are used to extract the underlying objective function ${}^eJ$ and then generate robot behavior by optimizing for a corresponding robot cost function ${}^lJ$.

For closed-loop control, the demonstrator's behavior can be described in terms of a *control policy*

$$ {}^e\mathbf{u} = {}^e\boldsymbol{\pi}({}^e\mathbf{x}, t) \quad \in \mathbb{R}^n \tag{4} $$

and imitation is achieved by drawing correspondence in both the state and action space (i.e., ${}^e\mathbf{x} \equiv {}^l\mathbf{x}$ and ${}^e\mathbf{u} \equiv {}^l\mathbf{u}$) and implementing a controller

$$ {}^l\mathbf{u} = \tilde{\boldsymbol{\pi}}({}^l\mathbf{x}, t) \quad \in \mathbb{R}^n \tag{5} $$

where $\tilde{\boldsymbol{\pi}}$ is an approximation of ${}^e\boldsymbol{\pi}$ (e.g., estimated through supervised learning on the demonstration data [19]–[22]).

Clearly, direct imitation is only possible in a few special cases where the dynamics and actuation of the robot are especially similar to that of the human. For instance, if the robot is actuated with artificial muscles (e.g., McKibben muscles [23]), it may be possible to directly feed the muscle activations recorded from a human as a command signal to the robot actuators. Evidently, this approach has the benefit of simplicity, but its applicability is limited since such direct correspondence between demonstrator and imitator is rare.

B. Feature Tracking for Abstracting Hardware Differences

A second approach is *feature-based imitation* of the observed behavior. The basis of this approach is to select a set of salient features of the demonstrated behavior ${}^e\boldsymbol{\phi}({}^e\mathbf{x}, {}^e\mathbf{u}, t)$, find the "equivalent" features of the robot's behavior ${}^l\boldsymbol{\phi}({}^l\mathbf{x}, {}^l\mathbf{u}, t)$, and draw correspondence between the two (i.e., ${}^l\boldsymbol{\phi} \equiv {}^e\boldsymbol{\phi}$) [24]. For example, the features might include the joint stiffness and damping profiles of the human arm that occur during movement. By drawing an equivalence between these and the joint stiffness and damping of the robot, the feature-based approach imitates behavior by matching those features as closely as possible during the movement.

1) Benefits of Feature-Based Imitation: One of the benefits of this approach is that it allows one to focus only on the key features of the demonstrated behavior, while ignoring those that are irrelevant and emerge solely as a consequence of the demonstrator's specific embodiment. For example, it is known that there is a coupling between the damping and stiffness of the human musculoskeletal system [25] so that any human demonstrated behavior inherently contains a nonnegligible damping profile, in addition to stiffness and position modulation. Since this damping is inherent to the dynamics of the human, it cannot be avoided whether or not it is beneficial for a given task. In throwing, for instance, damping may be detrimental to performance as it dissipates energy that could be used to throw greater distances. A robotic imitator with decoupled control of stiffness and damping (see, e.g., [26]) is not subject to such restrictions, and, therefore, may profit from imitating the stiffness only (to exploit energy storage, similar to the human), while avoiding energy dissipation by minimizing the damping during the throw [4].

With a feature-based approach, we would seek to match only the *key beneficial features*, and ignore extraneous properties of the demonstrations. In other words, we seek to *abstract* the behavior from the specific embodiment of the demonstrator and seek ways to imitate these features *independent of the specific embodiment* of the imitator system (i.e., design of the robotic device). We clarify the issues involved in this with an example.

2) Example: Ideal VSA, MACCEPA, and Edinburgh SEA: To illustrate the influence that different mechanical designs have on the control of impedance features, such as stiffness and equilibrium position, we consider three possible designs for a single-joint VSA (see Fig. 3).

The first and simplest of the three is the idealized single-joint VSA [see Fig. 3(a)], in which we assume that the stiffness and equilibrium position are directly controllable (i.e., $\mathbf{u} = (q_0, k)^\top$) and that the torque around the joint is given by

$$ \tau(q, \mathbf{u}) = -k(q_0 - q) \quad \in \mathbb{R} \tag{6} $$

where $q \in \mathbb{R}$ is the joint angle. In this case, the control of equilibrium position and stiffness is independent, enabling any combination of position and stiffness to be selected. This is illustrated in Fig. 3(a), right, where, for instance, moving along the y-axis (corresponding to u_2) adjusts the stiffness, but has no effect on the equilibrium position, and vice versa. Unfortunately, in real mechanisms, it is rarely possible to achieve such ideal behavior.

In contrast, consider the MACCEPA [2] and the Edinburgh SEA [27] as examples of actuators of differing designs that have both been realized in hardware. For the MACCEPA, the joint torque is given by

$$ \tau(q, \mathbf{u}) = \kappa BC \sin(u_1 - q) \left(1 + \frac{ru_2 - (C - B)}{A(q, u_1)}\right) \in \mathbb{R} \tag{7} $$

where $\mathbf{u} = (u_1, u_2)^\top$ are the commanded positions of the two servomotors [see Fig. 3(b)], $q \in \mathbb{R}$ is the joint angle, κ is the spring constant, r is the radius of the winding drum (mounted on the servo that extends the spring). $A(q, u_1), B$, and C are the distances illustrated in Fig. 3(b), with $A(q, u_1) = \sqrt{B^2 + C^2 - 2BC\cos(u_1 - q)}$. Note that, due to the multiplication of terms dependent on u_1 and u_2, there exists a coupling

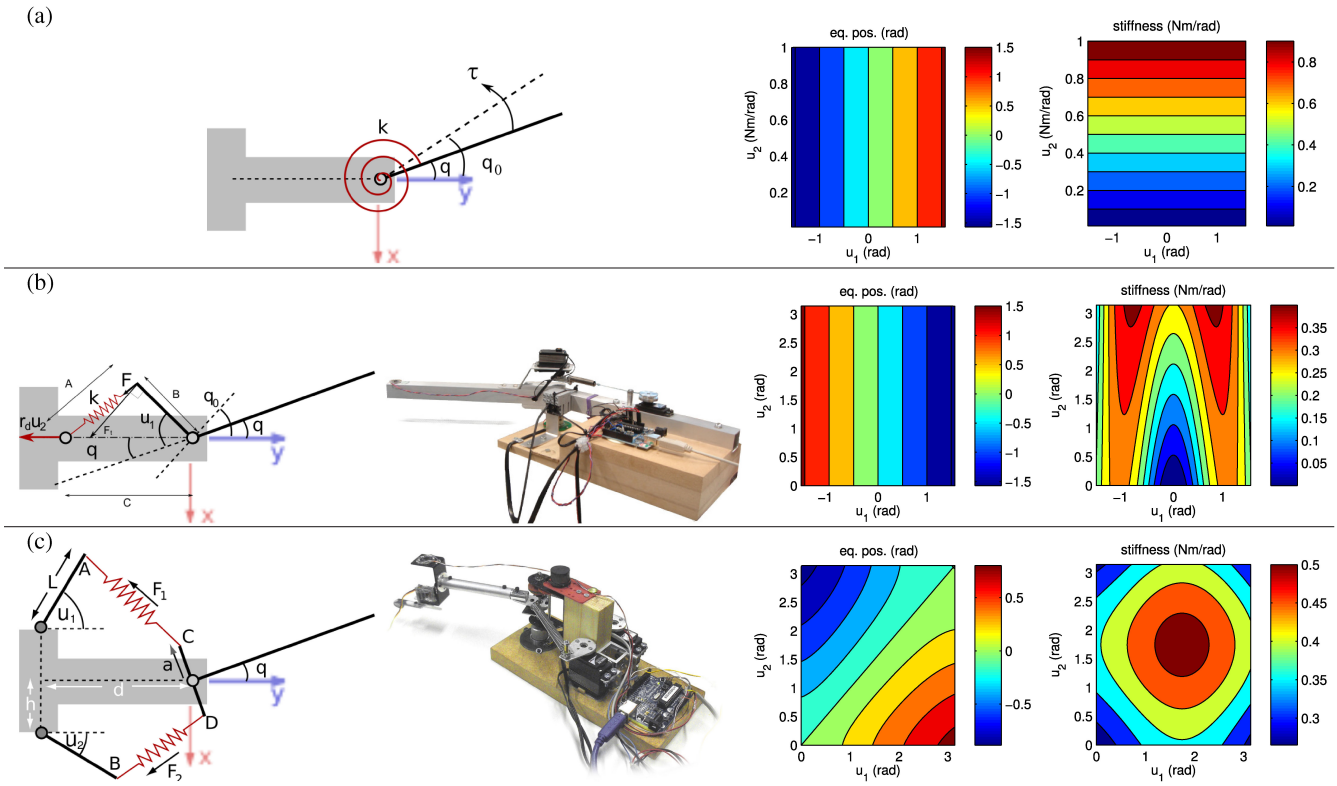

Fig. 3. (Left) Geometry, dynamics, and hardware implementation of the 1-link variable stiffness actuators used in the numerical simulations and experiments. (Right) Equilibrium position and stiffness as a function of commands $\mathbf{u}$ (evaluated at $q = 0\,\text{rad}, \dot{q} = 0\,\text{rad}\cdot\text{s}^{-1}$). (a) Ideal VSA. (b) MACCEPA. (c) Edinburgh SEA.

between equilibrium position and stiffness, making independent control of the two difficult. To illustrate this, we can make a similar plot of the equilibrium position and stiffness as a function of motor commands for this plant [see Fig. 3(b), right]. Here, we can see that, although the equilibrium position is only influenced by the first motor (u_1), there is a rather complex, nonlinear relationship between u_1 and u_2 and the stiffness.

For the Edinburgh SEA, an antagonistic arrangement is used in which the motors adjust the position of two levers connected through springs to the free link [see Fig. 3(c)]. In this case, the torque around the joint is given by

$$\tau(q, \mathbf{u}) = \hat{\mathbf{z}}^\top((\mathbf{F}_2(q, u_2) - \mathbf{F}_1(q, u_1)) \times \mathbf{a}(q)) \quad \in \mathbb{R} \quad (8)$$

where $\mathbf{u} = (u_1, u_2)^\top$ are the commanded positions of the two servomotors, $q \in \mathbb{R}$ is the joint angle, $\hat{\mathbf{z}}$ is the unit vector along the joint rotation axis, $\mathbf{a}(q) = (a\cos q, a \sin q, 0)^\top$, $\mathbf{F}_i(q, u_i) = \kappa(\sigma_i - \sigma_0)\frac{\sigma_i}{\sigma_i}, i \in \{1, 2\}$ are the forces due to the two springs (both with spring constant κ), $\boldsymbol{\sigma}_1 = (-h - L\sin u_1, -d + L\cos u_1, 0)^\top + \mathbf{a}(q)$ and $\boldsymbol{\sigma}_2 = (h + L\sin u_2, -d + L\cos u_2, 0)^\top - \mathbf{a}(q)$ are the extensions of the two springs (i.e., the vectors CA, and DB, respectively), and all other quantities are illustrated in Fig. 3(c). In this case, due to the antagonistic actuation, there is a strongly coupled, nonlinear relationship between the motor commands and the joint equilibrium position and stiffness [as illustrated in Fig. 3(c)], making it difficult to control these quantities directly.

These examples illustrate the fact that, even for relatively simple VSA designs, there is considerable difficulty in directly regulating the position and stiffness. At first glance, it would seem necessary to develop specialized controllers for each design, in order to exploit their physical properties. However, in Section III-A, we will outline a general method to control arbitrary VSAs with a constraint-based framework.

3) Disadvantages of Feature-Based Imitation: While the feature-based approach can be highly effective for behavior transfer in certain scenarios, it also has some difficulties in its application. One of the primary problems is in identifying which features of the demonstrator's behavior are key to achieving good task performance. In particular, effective use of feature-based imitation requires an appropriate understanding of the way in which different features affect *task performance* under the dynamics of both the demonstrator and the robotic imitator. Often, humans' strategies for employing variable impedance are highly adapted to certain specific properties of the musculoskeletal system. The consequence of this is that care must be taken when attempting to imitate that behavior to ensure that it is appropriate for the robotic plant.

As an example, consider the task of point-to-point reaching in free space (i.e., in the absence of external loads or other perturbations). Commonly, in such tasks, humans tend to increase their impedance toward the end of the movement to ensure that the target is hit accurately [13], [28] (i.e., to counter the effects of signal-dependent noise [12]). This comes at the cost of increased energy expenditure, since the human must cocontract muscles to achieve this. However, for a *less noisy* robotic imitator, this may be unnecessary, since relatively high accuracy (compared with the human) may be achieved even at relatively low impedance. As such, a better strategy for the robot would

be to keep impedance low throughout the movement, thereby avoiding unnecessary energy consumption.

While the feature-based approach may suffer from such issues, this does not mean that it should be ruled out for all applications. For example, in the context of online, teleoperated control of a robot, the feature-based approach can be exploited to provide an intuitive way for a human operator to control a slave robot. Feature-based imitation is particularly suitable in such online, interactive control scenarios, since its speed and efficiency makes the control responsive, and the proficiency of the operator at controlling the robot for a given task can help to overcome errors due to the mismatch in the dynamics. It is, therefore, worthwhile to consider feature-based imitation in the light of the requirements of applications. In Section III-A, we outline a constraint-based approach to feature tracking for the control of VIAs and illustrate its use experimentally in Section IV.

C. Inverse Optimal Control for Task-Based Imitation

The third behavior transfer approach, which is considered in this paper, is that of *task-based imitation* through inverse OC. The idea behind this approach is to seek the *objectives* (i.e., task goals) of the demonstrated behavior, and then present *a corresponding set of objectives* to the imitator. Specifically, in this approach, it is assumed that the demonstrator's behavioral goals are encoded in the form of some objective function ${}^{e}J(\cdot)$ by which task performance is measured. Demonstrated behavior is assumed to optimize this function with respect to the demonstrator's dynamics, and therefore, similarly optimal behavior may be achieved by the imitator if a correspondent objective function ${}^{l}J(\cdot)$ can be defined and optimized.[2]

With this representation (i.e., drawing the correspondence ${}^{e}J \equiv {}^{l}J$), behavior is transferred at the level of *task goals* (i.e., via the objective function that *defines the task*), independent of the specific control strategy, or embodiment of the demonstrator. Furthermore, by optimizing the imitator (robot) behavior in such a way as to take into account the *imitator's dynamics, task-based imitation* allows different *strategies* to be planned that are tailored to the imitator's embodiment.

1) Objective Functions for Demonstrated Trajectories: While inverse OC may be formulated in several different ways according to the setting [17], [29]–[31], in this paper, we primarily focus on discrete movements (i.e., with a finite duration). Specifically, we assume that each demonstration is given in the form of a trajectory through the state–action space of the demonstrator, ${}^{e}\mathbf{x}(t), {}^{e}\mathbf{u}(t)$, from start state ${}^{e}\mathbf{x}_0$, and with duration[3] T. The trajectory is assumed to be optimal with respect to some (unknown) objective function

$$ {}^{e}J = {}^{e}h({}^{e}\mathbf{x}(T)) + \int_0^T {}^{e}l({}^{e}\mathbf{x}, {}^{e}\mathbf{u}, t)\,\mathrm{d}t \tag{9} $$

where ${}^{e}h(\cdot), {}^{e}l(\cdot) \in \mathbb{R}$ are cost functions that are defined on the state–action space of the demonstrator. For instance, ${}^{e}l({}^{e}\mathbf{x}, {}^{e}\mathbf{u}, t)$ may describe the instantaneous power consumed by the demonstrator's actuators (e.g., the metabolic energy consumed by human muscles at a given activation). Note that here, since the optimality of the demonstrated trajectories depends on the demonstrator's dynamics ${}^{e}\mathbf{f}(\cdot)$, the recorded trajectories will not, in general, be optimal under the dynamics of a different (learner) system ${}^{l}\mathbf{f}(\cdot)$, i.e.,[4] $\{{}^{e}\bar{\mathbf{x}}, {}^{e}\bar{\mathbf{u}} \,|\, {}^{e}\mathbf{f}(\cdot)\} \neq \{{}^{l}\bar{\mathbf{x}}, {}^{l}\bar{\mathbf{u}} \,|\, {}^{l}\mathbf{f}(\cdot)\}$.

Accordingly, in order to seek appropriate strategies for the imitator, an equivalent objective function

$$ {}^{l}J = {}^{l}h({}^{l}\mathbf{x}(T)) + \int_0^T {}^{l}l({}^{l}\mathbf{x}, {}^{l}\mathbf{u}, t)\,\mathrm{d}t \tag{10} $$

must be defined on the learner's state–action space, where the terms ${}^{l}h(\cdot), {}^{l}l(\cdot) \in \mathbb{R}$ define cost terms with a meaningful correspondence to those of the expert ${}^{e}h(\cdot), {}^{e}l(\cdot)$. For example, if the term ${}^{e}l({}^{e}\mathbf{x}, {}^{e}\mathbf{u}, t)$ of a human demonstrator represents the energy consumption of the muscles, one might define ${}^{l}l({}^{l}\mathbf{x}, {}^{l}\mathbf{u}, t)$ as the electrical power consumed by the motors of a robotic manipulator. The goal of imitation then is to find the optimal behavior for the learner $\{{}^{l}\bar{\mathbf{x}}, {}^{l}\bar{\mathbf{u}}\}$ under the dynamics ${}^{l}\mathbf{f}(\cdot)$ with respect to the *equivalent objective function* (10).

2) Benefits of Task-Based Imitation: Similar to feature-based approaches to imitation (see Section II-B), the ease with which we can define correspondent cost functions [see (9) and (10)] will depend on the specific embodiments of the two plants. For example, cost terms that are dependent on features, such as end-effector position, may be defined as exactly correspondent, whereas terms that are dependent on other properties, such as the applied torque or impedance, may require more complex definitions. A major benefit of this approach, however, is that often it is *much easier to define correspondence at the level of the task*, rather than at the detailed control level of the plants. For instance, when imitating human behavior (see Fig. 1), the selection of which dynamics characteristics to match (e.g., stiffness, damping, etc.) in a feature-based imitation approach will depend critically on the effect those have on the dynamics of the two plants with respect to the task goals. In contrast, with task-based imitation, only the salient features (e.g., target accuracy, energy consumption) are specified, with the low-level details of the behavior automatically handled by the control optimization.

III. Method

In this section, we turn to the implementation details of behavior transfer under the different approaches. We first outline a method for feature-based transfer tailored to imitation of impedance using a closed-loop tracking framework. We then describe an approach to task-based imitation through AL for inverse OC.

[2]By convention, in this paper, it is assumed that ${}^{e}J$ and ${}^{l}J$ represent *cost* so that their *minimization* indicates better performance.

[3]For simplicity, through the paper, we assume finite length trajectories of equal length. However, as discussed in [17] and [31], inverse OC techniques are also readily extended to variable length, or even infinite horizon tasks.

[4]For compactness, here, we use the "bar" notation to denote optimality, i.e., $\bar{\mathbf{u}}$ denotes the optimal command sequence, and $\bar{\mathbf{x}}$ denotes the optimal trajectory in state space.

A. Imitation by Impedance Feature Matching

Here, we consider the problem of (feature-based) transfer of a demonstrator's impedance on a robotic system. In particular, we wish to imitate the stiffness and equilibrium position of the demonstrator, as features of the behavior, i.e., we draw the correspondence

$$ {}^{e}\boldsymbol{\phi} \equiv {}^{l}\boldsymbol{\phi} \;\Rightarrow\; ({}^{e}\mathbf{q}_0{}^{\top}, {}^{e}\mathbf{k}^{\top})^{\top} \equiv ({}^{l}\mathbf{q}_0{}^{\top}, {}^{l}\mathbf{k}^{\top})^{\top} \tag{11} $$

where ${}^{e}\mathbf{q}_0 \in \mathbb{R}^p$ is the joint equilibrium position vector[5] of the demonstrator, ${}^{e}\mathbf{k} = \mathrm{vec}({}^{e}\mathbf{K}) \in \mathbb{R}^{p^2}$ is the demonstrator's joint stiffness matrix ${}^{e}\mathbf{K} \in \mathbb{R}^{p\times p}$ in the vector form, and ${}^{l}\mathbf{q}_0$ and ${}^{l}\mathbf{k}$ are the corresponding equilibrium position and stiffness elements for the learner. Since we are considering impedance matching in joint space, we also assume that there is some meaningful correspondence between the joint space kinematics ($\mathbf{q}_0$) and impedance ($\mathbf{k}$) of the two systems, i.e., between ${}^{e}\mathbf{q}$ and ${}^{l}\mathbf{q}$, and in particular, that they are of equal dimension ${}^{e}\mathbf{q}, {}^{l}\mathbf{q} \in \mathbb{R}^p$ (and therefore ${}^{l}\mathbf{q}_0 \in \mathbb{R}^p$, ${}^{l}\mathbf{k} \in \mathbb{R}^{p^2}$, and ${}^{e}\boldsymbol{\phi}, {}^{l}\boldsymbol{\phi} \in \mathbb{R}^{\mu}$ with $\mu = p + p^2$).

1) Estimation of Impedance Features: In order to achieve imitation of the features defined in (11), some scheme for their estimation (or measurement) is required both for 1) the demonstrator and 2) the imitator. In the case of the former, feature estimation is required to *extract the desired impedance from the demonstrations*, i.e., to provide a reference to the imitator. In the latter case, estimation is desirable for feedback purposes (i.e., to evaluate the accuracy with which the demonstrator's impedance is reproduced).

To satisfy these requirements, we may appeal to several existing approaches for the estimation of the demonstrator's impedance. For example, one approach is to use surrogate measures of the impedance, based on measurable quantities such as muscle activations/cocontraction levels from EMG sensors [32]–[35]. An alternative approach (favored in this paper) is to use estimates of the demonstrator and imitator impedance that is derived from models of their respective dynamics.

Specifically, we assume that for both plants, the relationship between the state ${}^{e,l}\mathbf{x}$, the command vector ${}^{e,l}\mathbf{u}$, and the joint torque is given, i.e.,

$$ {}^{e}\boldsymbol{\tau} = {}^{e}\boldsymbol{\tau}({}^{e}\mathbf{x}, {}^{e}\mathbf{u}) \quad \in \mathbb{R}^p \tag{12} $$

for the demonstrator, and

$$ {}^{l}\boldsymbol{\tau} = {}^{l}\boldsymbol{\tau}({}^{l}\mathbf{x}, {}^{l}\mathbf{u}) \quad \in \mathbb{R}^p \tag{13} $$

for the imitator. These may be given in a closed form[6] or as a nonparametric model (e.g., from nonparametric regression).

The equilibrium position of the joints[7] as a function of state and command

$$ \mathbf{q}_0 = \mathbf{q}_0(\mathbf{x}, \mathbf{u}) \quad \in \mathbb{R}^p \tag{14} $$

is defined as the solution of $\boldsymbol{\tau}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$ for $\mathbf{q}$. This may be found either analytically, or numerically with a root-finding algorithm such as the Newton–Raphson method. The joint stiffness matrix is defined as

$$ \mathbf{K} = \mathbf{K}(\mathbf{x}, \mathbf{u}) = -\frac{\partial \boldsymbol{\tau}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{q}}\Big|_{\mathbf{q}} \quad \in \mathbb{R}^{p\times p}. \tag{15} $$

Again, this may be derived in a closed form, or numerically, e.g., through finite differences. Computing (14) and (15) from (12) therefore provides an estimate of the demonstrator's stiffness and equilibrium position, and computing the same from (13) yields similar estimates for the imitator.

Note that, for both plants, (14) and (15) may be nonlinear functions of the state and commands, and that, depending on the system, the number of independent elements of $\mathbf{K}$ may vary. For example, the stiffness of each joint may be coupled so that $\mathbf{K}$ is (nondiagonal) symmetric, as in the human musculoskeletal system where synergistic muscle groups, biarticular muscles, the arrangement of tendons, etc., can cause cross coupling of joints (see, e.g., [6]). Alternatively, the stiffness of individual joints may be independent (as would be the case, for example, in a chain of MACCEPAs [2]), in which case, $\mathbf{K}$ reduces to a diagonal matrix. In such cases, the dimensionality of the imitation problem (11) may be reduced (and computation efficiency gained) by omitting those elements that cannot be independently varied.

2) Resolved Equilibrium and Stiffness Tracking Control: Given estimates of the equilibrium position and stiffness for both the demonstrator and imitator, we are now in a position to design controllers that enable the robotic imitator to mimic the demonstrated impedance. While different approaches for stiffness modulation in VSAs have been proposed [36], [37], here, we briefly outline a recent model-based approach that is well suited for closed-loop stiffness tracking on a variety of different VSAs [14]. The proposed approach is based on closed-loop tracking, using the estimated stiffness ${}^{e}\mathbf{K}$ and equilibrium position ${}^{e}\mathbf{q}_0$ of the demonstrator as the reference target.

Noting that in general (13) may be a nonlinear function of the commands $\mathbf{u}$, a direct (linear, orthogonal) decomposition for control of (the imitator's) equilibrium position and stiffness is not feasible in general. Instead, we move to the command velocity space for control: Taking the time derivative of (14) and (15) for the imitator, the linearized, forward impedance dynamics are

$$ {}^{l}\dot{\mathbf{q}}_0 = \mathbf{J}_{\mathbf{q}_0}({}^{l}\mathbf{x}, {}^{l}\mathbf{u}){}^{l}\dot{\mathbf{u}} + \mathbf{P}_{\mathbf{q}_0}({}^{l}\mathbf{x}, {}^{l}\mathbf{u}){}^{l}\dot{\mathbf{x}} \tag{16} $$

$$ {}^{l}\dot{\mathbf{k}} = \mathbf{J}_{\mathbf{k}}\;({}^{l}\mathbf{x}, {}^{l}\mathbf{u}){}^{l}\dot{\mathbf{u}} + \mathbf{P}_{\mathbf{k}}\;({}^{l}\mathbf{x}, {}^{l}\mathbf{u}){}^{l}\dot{\mathbf{x}} \tag{17} $$

[5]For space reasons, here, we primarily consider impedance matching in the joint space; however, the methods that are presented can easily be extended to stiffness and equilibrium position matching in task space through the approach, which is described in [14].

[6]In our experiments, we employ a biomechanical model of the musculoskeletal system (including muscle dynamics) to predict human impedance features, and a rigid-body model of the actuators, validated by a system identification, for the robots (for details, see Section IV).

[7]For simplicity, we assume that the torque functions ${}^{e,l}\boldsymbol{\tau}(\cdot,\cdot)$ in (12) and (13) represent the torque around the joint in the absence of an external load, and therefore, (14) represents the *unloaded* equilibrium position. If, instead, we wish to estimate the (link side) equilibrium position under an external load, then $\mathbf{q}_0(\mathbf{x}, \mathbf{u})$ must instead be computed as the solution to $\boldsymbol{\tau}(\mathbf{x}, \mathbf{u}) + \boldsymbol{\tau}_e = \mathbf{0}$, where $\boldsymbol{\tau}_e$ represents the external torque due to the load.

where ${}^l\dot{\mathbf{q}}_0$, ${}^l\dot{\mathbf{k}}$ are the change in equilibrium position and stiffness with respect to time, ${}^l\dot{\mathbf{u}} \in \mathbb{R}^s$ is the rate of change of motor commands, $\mathbf{J}_{\mathbf{q}_0} \in \mathbb{R}^{p\times s}$ and $\mathbf{J}_{\mathbf{k}} \in \mathbb{R}^{p^2\times s}$ are the Jacobian of the equilibrium position and the stiffness with respect to motor commands, while $\mathbf{P}_{\mathbf{q}_0} \in \mathbb{R}^{p\times r}$ and $\mathbf{P}_{\mathbf{k}} \in \mathbb{R}^{p^2\times r}$ are the corresponding Jacobians with respect to the state.

To simultaneously control equilibrium position and stiffness (in joint space[8]), we can invert this relationship to yield[9]

$$ {}^l\dot{\mathbf{u}} = \mathbf{J}^\dagger\dot{\mathbf{r}} + (\mathbf{I} - \mathbf{J}^\dagger\mathbf{J})\mathbf{u}_0 \tag{18}$$

where $\dot{\mathbf{r}} = ({}^l\dot{\mathbf{q}}_0 - \mathbf{P}_{\mathbf{q}_0}{}^l\dot{\mathbf{x}}, {}^l\dot{\mathbf{k}} - \mathbf{P}_{\mathbf{k}}{}^l\dot{\mathbf{x}})^\top \in \mathbb{R}^{p+p^2}$, $\mathbf{J} = (\mathbf{J}_{\mathbf{q}_0}, \mathbf{J}_{\mathbf{k}})^\top$ is the combined Jacobian, $\mathbf{I} \in \mathbb{R}^{s\times s}$ is the identity matrix, $\mathbf{J}^\dagger$ denotes the Moore–Penrose pseudoinverse of $\mathbf{J}$, and $\mathbf{u}_0 \in \mathbb{R}^s$ is an arbitrary vector. The latter can be used to resolve any further redundancy in the actuation (such as additional actuators used for varying damping [38] or for mechanisms where multiple actuators are used to control variable stiffness elements, e.g., [39]).

Application of (18) requires state derivatives, provided by feedback, or calculated from the analytical model of the system dynamics. To avoid the requirement on analytical modeling, as well as to circumvent the noise and phase-lag issues that are related with the feedback on ${}^l\dot{\mathbf{x}}$, we employ online feedback about the current stiffness and equilibrium states, i.e., we choose $\dot{\mathbf{r}}$ according to the difference in the desired and actual equilibrium and stiffness values $\dot{\mathbf{r}} = \kappa_p({}^e\mathbf{q}_0 - {}^l\mathbf{q}_0, {}^e\mathbf{k} - {}^l\mathbf{k})^\top$, where κ_p is a gain parameter. This solution is similar to closed-loop inverse kinematic control [40], and mitigates instabilities due to drift [41].

3) Benefits of Resolved Impedance Tracking: Imitation of the demonstrated impedance through this approach has several benefits. The first is that it enables us to match these features of the demonstrator's behavior, with relative ease, in a *device-independent* way. For example, if we wish to track the stiffness of a single joint of a human demonstrator, then we are free to choose the robotic mechanism: in this case, any of the VSAs that are described in Section II-B could be used.

A second benefit is the flexibility that this approach gives in selecting the correspondence between demonstrator and imitator features. In particular, the explicit decomposition into task and nullspace parts in (18) means that features deemed to have lower importance in the imitated behavior can be either ignored (by eliminating rows of $\mathbf{J}$), or tracked with lower priority (by pushing these features into the nullspace).

For example, consider the case of behavior transfer from a demonstrator with stiffness matrix ${}^e\mathbf{K}$ that is constrained always to be symmetric (e.g., due to joint coupling arising from biarticular muscles) to a (fully actuated) robotic imitator (see, e.g., [10]) where the entire stiffness matrix ${}^l\mathbf{K}$ can be controlled. In this case, we can draw correspondence on a subset of the elements of ${}^e\mathbf{K}$ and ${}^l\mathbf{K}$ (e.g., define the feature vectors as ${}^e\boldsymbol{\phi} = ({}^e\mathbf{q}_0{}^\top, \mathrm{diag}({}^e\mathbf{K})) \in \mathbb{R}^{2p}$ and ${}^l\boldsymbol{\phi} = ({}^l\mathbf{q}_0{}^\top, \mathrm{diag}({}^l\mathbf{K})) \in \mathbb{R}^{2p}$) and then use the remaining degrees of freedom of the imitator for other objectives (e.g., joint stabilization through active damping [42]). The latter may be incorporated into the imitator's behavior through the nullspace term $\mathbf{u}_0$ in (18).

B. Imitation by Inverse Optimal Control

In this section, we consider the transfer of behavior through inverse OC. In particular, we wish to imitate the demonstrator on the level of *task goals* as encoded by the *objective function optimized*, i.e., drawing the correspondence

$$ {}^eJ \equiv {}^lJ \quad \in \mathbb{R} \tag{19}$$

where ${}^eJ$ and ${}^lJ$ are objective functions for the expert and learner, respectively.

In this paper, we pursue an approach based on AL [17], whereby the demonstrator's cost function is approximated by a parametric model $\tilde{J}(\mathbf{w})$ with the parameters $\mathbf{w}$ estimated from the demonstration data. A schematic overview is illustrated in Fig. 2 (outer path), which shows the processing steps, and the inputs required at each stage. Reading from the top left, we first collect demonstrations from an expert (e.g., a human) performing some task. This is fed into a module for AL (top right), along with information about the demonstrator's dynamics. Based on these, estimates $\tilde{\mathbf{w}}$ of the parameters of the expert's cost function are made that are then fed to the OC module (bottom right) along with a model of the imitator (robot) dynamics. The OC module finds the optimal strategy for the imitator, with respect to the learnt cost function and imitator dynamics, and this is, finally, sent to the robot for execution. In the following, we briefly describe the details of the AL and OC components.

1) Multiplicative Weights Apprenticeship Learning: In recent years, numerous approaches to inverse OC have been proposed [17], [29]–[31], [43]–[46]. The method chosen here is called multiplicative weights apprenticeship learning (MWAL), originally developed in [16]. The algorithm is based on principles of adversarial game theory and, as such, has been shown to be a robust method for AL. Furthermore, its efficiency makes it well suited for learning in the robotics domain, where state–action spaces are typically high-dimensional and continuous.

The method works on data given as a set of $\mathcal{J}$ trajectories $D = \{({}^e\mathbf{x}_0^j, {}^e\mathbf{u}_0^j), \ldots, ({}^e\mathbf{x}_T^j, {}^e\mathbf{u}_T^j)\}_{j=0}^{\mathcal{J}}$ of states ${}^e\mathbf{x}$ and actions ${}^e\mathbf{u}$ recorded from the demonstrator. In the model-based approach described here, the expert's dynamics (1) are assumed to be known, i.e., the function

$$ {}^e\dot{\mathbf{x}} = {}^e\mathbf{f}({}^e\mathbf{x}, {}^e\mathbf{u}) \quad \in \mathbb{R}^m$$

is given or may be approximated either through a system identification or dynamics learning[10] [8], [48].

[8]Note that the present approach can also be used for the tracking task (e.g., end-effector) space stiffness and equilibrium position. Denoting the task-space coordinates as $\mathbf{s} \in \mathbb{R}^p$, and the Jacobian from joint to task space as $\mathbf{W}({}^l\mathbf{q}) \in \mathbb{R}^{p\times p}$ (assumed to be square and full rank), the task-space stiffness is ${}^l\mathbf{K}_s = (\mathbf{W}^\top)^{-1}{}^l\mathbf{K}\mathbf{W}^{-1} \in \mathbb{R}^{p\times p}$, and the task-space equilibrium position ${}^l\mathbf{s}_0 \in \mathbb{R}^p$ is the solution of

$$\mathbf{F}_s = (\mathbf{W}^\top)^{-1}{}^l\boldsymbol{\tau} = \mathbf{0}$$

where $\mathbf{F}_s$ is the task-space restoring force. Given ${}^l\mathbf{s}_0$ and ${}^l\mathbf{K}_s$, we can then derive the Jacobians $\mathbf{J}_{\mathbf{s}_0} \in \mathbb{R}^{p\times s}$ and $\mathbf{J}_{\mathbf{k}_s} \in \mathbb{R}^{p^2\times s}$, with respect to the motor commands ${}^l\mathbf{u}$, and perform tracking in a similar way as in the joint space approach [14].

[9]We omit the dependence on ${}^l\mathbf{x}$ and ${}^l\mathbf{u}$ for readability.

[10]Note, however, that even in the absence of a model of ${}^e\mathbf{f}$, MWAL may also be applied using model-free approaches. See, e.g., [47].

The trajectories in D are assumed to be optimal under the dynamics (1), with respect to a cost function of the form

$$ {}^{e}J = \sum_{i=1}^{\eta} w_i {}^{e}h_i({}^{e}\mathbf{x}(T)) + \sum_{i=\eta+1}^{\mathcal{I}} w_i \int_0^T {}^{e}l_i({}^{e}\mathbf{x}, {}^{e}\mathbf{u}, t)\,\mathrm{d}t \tag{20} $$

or, more compactly

$$ {}^{e}J = \mathbf{w}^{\top}\boldsymbol{\psi}({}^{e}\mathbf{x}, {}^{e}\mathbf{u}). \tag{21} $$

Here, ${}^{e}h_i(\cdot), {}^{e}l_i(\cdot) \in \mathbb{R}$ are a set of basis functions that represent terminal and running costs, respectively, i.e., $\boldsymbol{\psi} = ({}^{e}h_1, \ldots, {}^{e}h_\eta, \int_0^T {}^{e}l_{\eta+1}\mathrm{d}t, \ldots, \int_0^T {}^{e}l_{\mathcal{I}}\mathrm{d}t)^{\top}$. These may be made up of a set of bases for a generic function approximator (e.g., Gaussian radial basis functions), or a set of salient features of the task (e.g., energy or accuracy costs). The weights $\mathbf{w} = (w_1, \ldots, w_{\mathcal{I}})^{\top}$ are the parameters to be estimated, and it is assumed (by renormalization, if necessary) that $w_i > 0\,\forall\, i$ and $\sum_i w_i = 1$.

The idea behind MWAL is that the weights w_i specifying the importance of the different components of the objective function (21) can be determined efficiently by comparing the expected value of the observed behavior D with that of a second set of trajectories ${}^{p}D$ that are optimal with respect to an estimate of (21) with weights $\tilde{w}_i$. Specifically, since the cost bases ${}^{e}h_i(\cdot), {}^{e}l_i(\cdot)$ are given (as part of our model), we can estimate the value of the trajectories in D and ${}^{p}D$, with respect to each of the bases separately. That is, for the ith basis function

$$ \tilde{v}_i = \frac{1}{\mathcal{J}} \sum_{j=0}^{\mathcal{J}} \int_0^T {}^{e}l_i({}^{e}\mathbf{x}_j(t), {}^{e}\mathbf{u}_j(t), t)\,\mathrm{d}t \tag{22} $$

if it is a running cost and

$$ \tilde{v}_i = \frac{1}{\mathcal{J}} \sum_{j=0}^{\mathcal{J}} {}^{e}h_i({}^{e}\mathbf{x}_j(T)) \tag{23} $$

if it is a terminal cost. We can then compare the difference in these value estimates to adjust the weights $\tilde{w}_i$, by scaling up those for which the value of the expert trajectories is lower (indicating a stronger preference to minimize these components of the cost), and scaling down those for which the values are higher (indicating the opposite). In successive iterations, MWAL alternates between solving the forward OC problem under the current estimate of $\tilde{\mathbf{w}}$ to find trajectories ${}^{p}D$, and then updating the estimate based on the difference in estimated values ${}^{e}\tilde{\mathbf{v}} = ({}^{e}\tilde{v}_1, \ldots, {}^{e}\tilde{v}_{\mathcal{I}})_D$, and ${}^{p}\tilde{\mathbf{v}} = ({}^{p}\tilde{v}_1, \ldots, {}^{p}\tilde{v}_{\mathcal{I}})_{{}^{p}D}$. This proceeds until convergence to a set of weights that, when optimized, reproduces the demonstrated behavior D. MWAL is summarized in Algorithm 1, and full details can be found in[11] [16]. For the forward optimization step (Step 6 of Algorithm 1), the iterative local quadratic Gaussian (ILQG) algorithm [49] is used, the details of which are described in the following.

[11]Note that, in Algorithm 1, we have made two adjustments to the formulation described in [16]. These are 1) introduction of a learning rate parameter α and 2) normalization of the vectors ${}^{e}\hat{\mathbf{v}} = {}^{e}\tilde{\mathbf{v}}/\|{}^{e}\tilde{\mathbf{v}}\|$ and ${}^{p}\hat{\mathbf{v}} = {}^{p}\tilde{\mathbf{v}}/\|{}^{e}\tilde{\mathbf{v}}\|$. While these adjustments do not affect the convergence properties of the algorithm (effectively, they correspond to a scaling of β), we found them convenient for adjusting the speed of learning, while maintaining robustness.

Algorithm 1 MWAL (modified from [16])

1: **Given**
- demonstration data $D\{({}^{e}\mathbf{x}_0^j, {}^{e}\mathbf{u}_0^j), \cdots, ({}^{e}\mathbf{x}_T^j, {}^{e}\mathbf{u}_T^j)\}_{j=0}^{\mathcal{J}}$,
- expert dynamics model ${}^{e}\mathbf{f}({}^{e}\mathbf{x}, {}^{e}\mathbf{u})$,
- cost bases $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_{\mathcal{I}})$.

2: **Initialise**
- 3: Let $\beta = \left(1 + \sqrt{\frac{2\log\mathcal{I}}{\mathcal{P}}}\right)^{-1}$ for some $\mathcal{P} \in \mathbb{Z}^{+}$.
- 4: Estimate ${}^{e}\tilde{\mathbf{v}} = ({}^{e}\tilde{v}_1, \ldots, {}^{e}\tilde{v}_{\mathcal{I}})$ from trajectories D according to (22) and (23). Normalise ${}^{e}\hat{\mathbf{v}} = {}^{e}\tilde{\mathbf{v}}/\|{}^{e}\tilde{\mathbf{v}}\|$.

5: **For** $p = 1, \ldots, \mathcal{P}$ **do**
- 6: Find trajectories ${}^{p}D$ that optimise $\tilde{J} = \sum_i {}^{p}\tilde{w}_i\psi_i$ under dynamics ${}^{e}\dot{\mathbf{x}} = {}^{e}\mathbf{f}({}^{e}\mathbf{x}, {}^{e}\mathbf{u})$
- 7: Estimate ${}^{p}\hat{v}_i$ from trajectories ${}^{p}D$ for all i
- 8: Let ${}^{p+1}\tilde{w}_i = {}^{p}\tilde{w}_i\beta^{-\alpha({}^{e}\hat{v}_i - {}^{p}\hat{v}_i)}$ for $i = 1, \cdots, \mathcal{I}$
- 9: Re-normalise $\tilde{w}$

10: **end for**

11: **Return** $\tilde{w}$

2) Transferring the Learnt Objective to the Imitator: Having completed the AL stage to find a model of the demonstrator's objectives, our next task is to find an appropriate behavior for the imitator. For this, we use local OC to optimize an equivalent cost function to that used by the demonstrator. Specifically, we parameterize the learner's cost function as a similar weighted combination of terms

$$ {}^{l}J = \sum_{i=1}^{\eta} \tilde{w}_i {}^{l}h_i({}^{l}\mathbf{x}(T)) + \sum_{i=\eta+1}^{\mathcal{I}} \tilde{w}_i \int_0^T {}^{l}l_i({}^{l}\mathbf{x}, {}^{l}\mathbf{u}, t)\,\mathrm{d}t. \tag{24} $$

Here, ${}^{l}h_i(\cdot), {}^{l}l_i(\cdot) \in \mathbb{R}$ are a set of basis functions that correspond to those of the expert (21), and $\tilde{w}_i$ are the weights learnt by MWAL in the previous step. At this point, a design decision must be made as to the appropriate correspondence between the learner's cost bases ${}^{l}h_i(\cdot), {}^{l}l_i(\cdot)$ and those of the expert ${}^{e}h_i(\cdot), {}^{e}l_i(\cdot)$. In general, this will depend on the specific embodiments (dynamics and actuators) of the two plants. However, as noted in Section II-C, in practical settings, this is relatively easily resolved (and at worst, it is no more difficult than specifying the correspondence in features ${}^{e}\boldsymbol{\phi}(\cdot), {}^{l}\boldsymbol{\phi}(\cdot)$ for feature-based imitation). For example, different terms might include the total work done by the two plants, or accuracy (e.g., in terms of the end-effector positions of the two plants). Further examples are given in the experiments (see Section IV).

Having defined correspondence in terms of these bases, and given the learnt weights $\tilde{\mathbf{w}}$, all that remains is to solve the OC problem defined by (24) and (2). Here, since we are interested in high-dimensional, continuous robot control problems, we use an efficient local OC method. In the next section, we briefly describe the details.

3) Forward Optimal Control with ILQG: In our framework, solving the forward OC problem enters at two points. First, in MWAL, the optimal trajectories ${}^{p}D$, with respect to the estimated cost function, are sought at every iteration for updating the weights. Second, as discussed previously, given the learned cost function, we seek the optimal movement for the imitator

plant. In both cases, we need a technique that 1) can cope with continuous, nonlinear systems and that 2) is efficient (since it is called multiple times during MWAL). There are numerous recent forward OC algorithms available for this [50]–[52]. The algorithm employed here is ILQG [49] since we found it to be an efficient, approximate model-based solver of OC problems.

Briefly, ILQG works by making a local linear-quadratic-Gaussian (LQG) approximation to OC problems and iteratively improving its solution around a nominal trajectory. It starts with a time-discretized initial guess of a control sequence $\bar{\mathbf{u}}^j$ of length T. At each iteration j, this is used to find the corresponding state sequence $\bar{\mathbf{x}}^j$ under the deterministic forward dynamics $\mathbf{f}(\cdot)$ via numerical integration. Next, the dynamics are linearly approximated with a Taylor expansion, and similarly, a quadratic approximation of the cost function around $\bar{\mathbf{x}}_t^j$ and $\bar{\mathbf{u}}_t^j$ is made. Both approximations are formulated as deviations $\delta\mathbf{x}_t^j = \mathbf{x}_t^j - \bar{\mathbf{x}}_t^j$ and $\delta\mathbf{u}_t^j = \mathbf{u}_t^j - \bar{\mathbf{u}}_t^j$ from the current trajectory and, therefore, form the local LQG problem. The latter can be solved efficiently via a modified Ricatti-like set of equations.

With the solution to these equations, a correction to the control signal $\delta\bar{\mathbf{u}}^j$ is found, which is used to improve the control sequence for the next iteration: $\bar{\mathbf{u}}^{j+1}(t) = \bar{\mathbf{u}}^j(t) + \delta\bar{\mathbf{u}}^j$. Finally, $\bar{\mathbf{u}}^{j+1}(t)$ is applied to the system dynamics, and the new total cost along the trajectory is computed. The algorithm stops once the cost ceases to decrease significantly ($\Delta J \approx 0$). After convergence, ILQG returns a control sequence $\bar{\mathbf{u}}$, gains $\bar{\mathbf{L}}$, and a state sequence $\bar{\mathbf{x}}$, which represents the optimal trajectory. In our framework, these trajectories are then either collected as sample data for Step 6 of the MWAL algorithm, or used for OC of the imitator plant, using the gains to provide local optimal feedback.

IV. Evaluations

In this section, we evaluate the different approaches to imitation in several impedance control scenarios. In the first investigation, we conduct a simulation study into behavior transfer from a model of the human wrist to two robotic VIAs with heterogeneous dynamics. We then report an experiment in which feature-based imitation is used for online behavior transfer in the context of human teleoperation of a nonbiomorphic robotic device. Finally, we report experiments in which task-based imitation is used to learn from human demonstrations for behavior transfer to the Edinburgh SEA [53].

A. Transferring Impedance Behavior on a Single Joint

The goal of the first investigation is to compare the three approaches for transferring human impedance behavior with heterogeneous robotic systems. As a case study for this, we look at the problem of transferring a "hitting" behavior onto two different robotic VIAs (as illustrated in Fig. 4) in simulation.

As demonstrator, a biomechanical model of the human wrist is used. The wrist model consists of a single joint, actuated by two antagonistic muscles with Kelvin–Voight muscle dynamics [6] (see Fig. 4, left). Its equation of motion is

$$I\,{}^e\ddot{q} + b\,{}^e\dot{q} = {}^e\tau({}^eq, {}^e\dot{q}, {}^e\mathbf{u}) \tag{25}$$

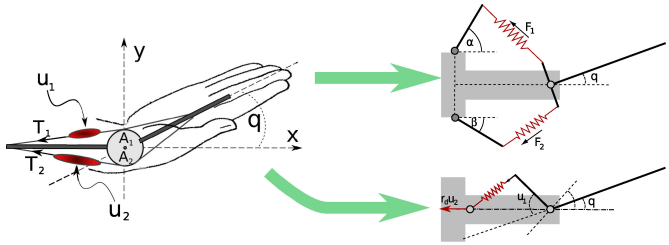

Fig. 4. Transfer from (left) human wrist model to (top right) the Edinburgh SEA and (bottom right) the MACCEPA.

where $I = 4.5 \times 10^{-3}$ kgm^2 is the moment of inertia, $b = 5 \times 10^{-3}$ N·msrad^{-1} is the damping, and the joint torque is given by

$${}^e\tau({}^eq, {}^e\dot{q}, {}^e\mathbf{u}) = -\mathbf{A}^\top\mathbf{T}({}^eq, {}^e\dot{q}, {}^e\mathbf{u}) \tag{26}$$

where the control inputs ${}^e\mathbf{u} \in \mathbb{R}^2$ represent muscle activations, $\mathbf{A} = (0.025, -0.025)^\top\, m$ are moment arms[12], and $\mathbf{T} \in \mathbb{R}^2$ are the muscle tensions

$$\mathbf{T}({}^eq, {}^e\dot{q}, {}^e\mathbf{u}) = \mathbf{K}_m({}^e\mathbf{u})\big(\mathbf{l}_r({}^e\mathbf{u}) - \mathbf{l}({}^eq)\big) - \mathbf{B}_m({}^e\mathbf{u})\dot{\mathbf{l}}({}^e\dot{q}). \tag{27}$$

Here, $\mathbf{l}({}^eq) = \mathbf{l}_{(q=0)} - \mathbf{A}\,{}^eq \in \mathbb{R}^2$ are muscle lengths, $\mathbf{l}_{(q=0)} \in \mathbb{R}^2$ is the muscle length at ${}^eq = 0$

$$\mathbf{K}_m({}^e\mathbf{u}) = \mathrm{diag}(k_0\mathbf{1} + g_k\,{}^e\mathbf{u}) \in \mathbb{R}^{2\times 2} \tag{28}$$

is the muscle stiffness

$$\mathbf{B}_m({}^e\mathbf{u}) = \mathrm{diag}(b_0\mathbf{1} + g_b\,{}^e\mathbf{u}) \in \mathbb{R}^{2\times 2} \tag{29}$$

the muscle damping, and $\mathbf{l}_r({}^e\mathbf{u}) = \mathbf{l}_0 + \mathrm{diag}(\mathbf{g}_r)\,{}^e\mathbf{u} \in \mathbb{R}^2$ is the muscle rest length. The elasticity coefficients $g_k = 1459.44\,\mathrm{Nm}^{-1}$ and $k_0 = 121.62\,\mathrm{Nm}^{-1}$ are given from the muscle model [6], $\mathbf{l}_0$ is set such that $\mathbf{l}_{(q=0)} - \mathbf{l}_0 = 0$, and $\mathbf{g}_r = (0.05, 0.05)^\top$ m. In this evaluation, the viscosity coefficients are also set to zero[13] (i.e., $g_b = 0$, $\mathrm{Nsm}^{-1} b_0 = 0\,\mathrm{Nsm}^{-1}$).

Combining all of the above, the expert's dynamics can be written as

$${}^e\mathbf{f}({}^e\mathbf{x}, {}^e\mathbf{u}) = ({}^e\dot{q}, ({}^e\tau - b)/I)^\top. \tag{30}$$

Using (26)–(29), the joint equilibrium position and stiffness can be computed through (14) and (15) as

$${}^eq_0({}^e\mathbf{u}) = (\mathbf{A}^\top\mathbf{K}_m\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{K}_m(\mathbf{l}_{(q=0)} - \mathbf{l}_0 + \mathrm{diag}(\mathbf{g}_r)\,{}^e\mathbf{u})) \tag{31}$$

and

$${}^ek({}^e\mathbf{u}) = \mathbf{A}^\top\mathbf{K}_m\mathbf{A}. \tag{32}$$

[12]In general, the moment arms around different joints (e.g., complex joints such as the shoulder) may depend on additional variables such as the joint angle q [18]. However, here, for simplicity, we assume the moment arms to be constant for this relatively simple joint [6].

[13]Note that to facilitate comparison of the imitated robot behavior with that of the demonstrator, in this experiment, we remove the command-dependent muscle damping from the wrist model; therefore, the only damping comes from the fixed joint damping term b. In Section IV-C, the full muscle model is used, including command-dependent damping.

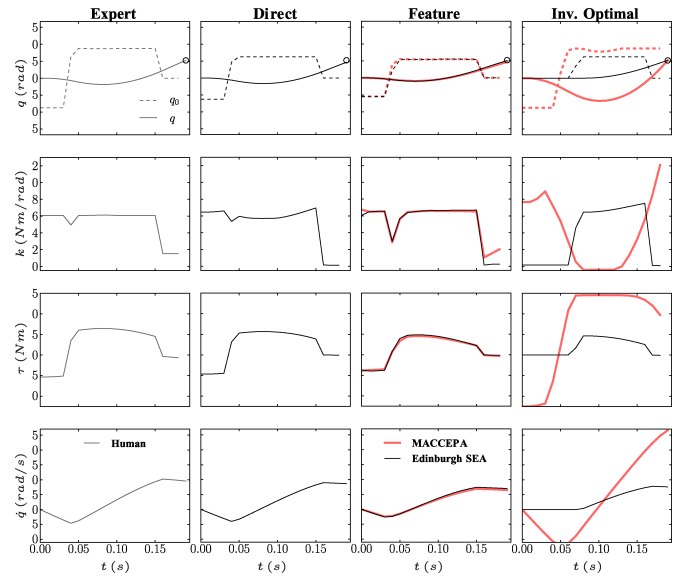

Fig. 5. Example joint position, stiffness, torque, and velocity profiles demonstrated from the simulated human wrist ("Expert") and transferred onto the two simulated robotic VSAs via the direct, feature-based, and inverse OC approaches.

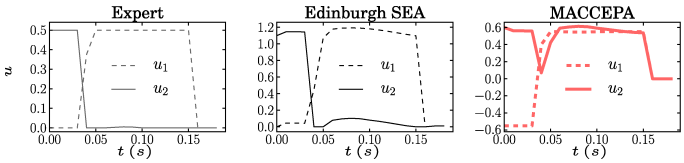

Fig. 6. Command profiles 1) demonstrated on the human wrist ("Expert") and 2) generated during feature-based imitation on the robotic plants.

The task is to hit a target as hard as possible. For this, the expert uses the cost function

$$
\begin{aligned}
{}^{e}J &= w_1({}^{e}q(T) - q^*)^2 - w_2\,{}^{e}\dot{q}(T) + w_3 \int_0^T {}^{e}\tau^2 \,\mathrm{d}t \\
&= w_1\,{}^{e}h_1 + w_2\,{}^{e}h_2 + w_3 \int_0^T {}^{e}l_3 \,\mathrm{d}t \qquad (33)
\end{aligned}
$$

where $q^* = 30^\circ$ is the target position in joint space, and ${}^{e}\tau$ is the torque applied around the joint [as given by (26)]. The three terms of (33), respectively, correspond to 1) minimizing the distance to the target (ball) at the time of impact ($T = 0.18\ s$), 2) maximizing the angular velocity at impact, and 3) minimizing effort during movement. The tradeoff between these objectives is determined by the weights $\mathbf{w} = (w_1, w_2, w_3)^\top = (0.9970, 0.0025, 0.0005)^\top$.

To generate demonstrations, ILQG is used to plan a set of trajectories optimizing (33) under the dynamics (30). Specifically, a set of $\mathcal{J} = 20$ such trajectories from a uniform-random distribution of start states ${}^{e}q(t=0) = U[-20, 0]^\circ$ are used. An example trajectory is illustrated in the leftmost column of Fig. 5, where we plot the joint position, stiffness, torque, and velocity over time.

As imitators, simulations of 1) the Edinburgh SEA [see Fig. 3(c)] and 2) the MACCEPA [see Fig. 3(b)] are used, with dynamics as described in Section II-B. The former is a biomimetic plant, with close (homomorphic) correspondence to the demonstrator (both have antagonistic actuation where coactivation of the commands $\mathbf{u}$ leads to increased stiffness). The MACCEPA is nonbiomorphic (i.e., geometrically dissimilar), but biomimetic in the sense that it also has variable stiffness (albeit controlled with a different mechanism). To enable fair comparison of demonstrated and imitated behavior, the parameters of the robotic plants are optimized as far as possible to have similar characteristics to those of the demonstrator: the dynamics parameters (e.g., inertia, damping, and friction constants) of the robots are made identical to those of the human, and the actuator parameters (e.g., spring constants, geometric parameters) are optimized so that the human and robots have similar capabilities in terms of the approximate torque, equilibrium position, and stiffness ranges. Note that, in reality, robotic actuators are often designed in a similar way, i.e., to try to match the capabilities and characteristics of humans. However, note also that the correspondence in these systems is not exact since the actuation relations [see (7), (8), and (26)] are different.

TABLE I
AVERAGE COST ${}^{l}J$ OF IMITATED TRAJECTORIES FROM DIFFERENT START STATES UNDER DIFFERENT IMITATION STRATEGIES

	Direct	Feature	Inv. Opt.
Swinger	-0.014 ± 0.005	-0.017 ± 0.007	-0.027 ± 0.004
MACCEPA	-	-0.017 ± 0.006	-0.066 ± 0.003

Shown are (mean ± s.d.) cost imitating trajectories from 50 random start states. Average cost incurred by expert during demonstrations: ${}^{e}J = -0.026 \pm 0.002$.

To compare the different methods, we apply

1) *direct imitation* by feeding the (renormalized) expert action sequence $\{{}^{e}\mathbf{u}_0, \ldots, {}^{e}\mathbf{u}_T\}$ as commands to the robot. Correspondence is defined as ${}^{e}\hat{\mathbf{u}} \equiv {}^{l}\hat{\mathbf{u}}$ (where ${}^{e}\hat{\mathbf{u}}, {}^{l}\hat{\mathbf{u}}$ are the commands normalized by their admissible range)[14];
2) *feature-based imitation*, tracking the computed equilibrium position and stiffness of the expert on the robots (as described in Section III-A). Here, correspondence is defined as $({}^{e}\hat{q}_0, {}^{e}\hat{k})^\top \equiv ({}^{l}\hat{q}_0, {}^{l}\hat{k})^\top$ (where, similarly, $\hat{\phi}$ denotes feature ϕ normalized by its admissible range);
3) *Apprenticeship learning*, as described in Section III-B (using the jth trajectory $\{({}^{e}\mathbf{x}_0, {}^{e}\mathbf{u}_0), \ldots, ({}^{e}\mathbf{x}_T, {}^{e}\mathbf{u}_T)\}_j$ as training data D, the expert's dynamics ${}^{e}\mathbf{f}(\cdot)$ as described by (30), with $\alpha = -300$ and $\mathcal{P} = 150$). Here, correspondence is drawn on the three terms of (33), i.e., $({}^{e}h_1, {}^{e}h_2, {}^{e}l_3)^\top \equiv ({}^{l}h_1, {}^{l}h_2, {}^{l}l_3)^\top$ where ${}^{l}h_1 = ({}^{l}q(T) - q^*)^2$, ${}^{l}h_2 = -{}^{l}\dot{q}(T)$, and ${}^{l}l_3 = \int_0^T {}^{l}\tau^2 \,\mathrm{d}t$.

The imitation approaches are applied to each of the $\mathcal{J}$ demonstrations, and evaluated by computing the average cost accumulated in each trajectory [according to the true expert cost (33)]. The results are summarized in Table I, and examples of the imitated behavior are plotted in Fig. 5.

Looking at the results for the Edinburgh SEA, we see that feature-based and direct imitation are both able to reproduce

[14]Note that since there is no direct correspondence between the human model and the MACCEPA, direct imitation is only performed with the Edinburgh SEA.

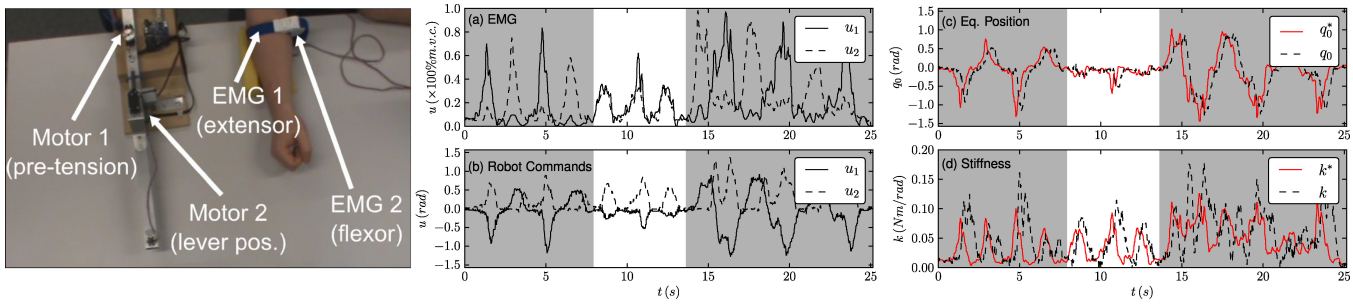

Fig. 7. Feature-based imitation of equilibrium position and stiffness on the MACCEPA. (a) Human EMG signals. (b) Robot motor commands. (c) Equilibrium position. (d) Stiffness. In (c) and (d), the black dashed line denotes the actual equilibrium position and stiffness realized by the actuator, and the red line indicates the desired equilibrium position/stiffness [predicted from the human data via (31) and (32)]. Note that to ensure that they remain within the admissible impedance range of the robot, the latter are normalized such that the theoretical maximum and minimum stiffness and equilibrium position achievable by the human (according to the wrist model) correspond to the maximum and minimum stiffness and equilibrium position achievable on the robot).

the hitting task. This can be confirmed by looking at the reproduced behavior (see Fig. 5), where we see 1) the hitting target (○) is reached accurately, and 2) peak velocity occurs at the end of the movement. The same is true for feature-based imitation on the MACCEPA, despite its totally nonbiomorphic design. Accurate tracking of the joint stiffness and equilibrium position is achieved, albeit with a very different command profile (cf., Fig. 6). These results indicate that, at minimum, 1) for a *biomimetic* hardware design, feature-based imitation is sufficient to reproduce the task, and 2) if the plant is additionally *biomorphic*, direct imitation is sufficient.

However, looking at the *task performance* (average cost), it is evident that neither of these approaches reach comparable performance with that of the demonstrator (see Table I). This is unsurprising since, due to differences in the dynamics, the optimal hitting strategy for the demonstrator is *suboptimal* for the imitators. If we take the inverse optimal imitation approach, on the other hand, the gap in performance is closed. For the Edinburgh SEA, inverse optimal imitation achieves similar performance to that of the demonstrator by adjusting the hitting strategy (e.g., compare differences in stiffness and equilibrium profiles in Fig. 5). For the MACCEPA, inverse optimal imitation significantly changes the hitting strategy so that performance even exceeds that of the demonstrator. This is possible since inverse optimal imitation *explicitly takes into account the imitator's dynamics* which, for the MACCEPA, are apparently better suited to the hitting task than the antagonistic plants.

B. Tracking Human Impedance Profiles

While feature-based approaches evidently do not always provide optimal behavior with respect to task goals, this does not rule out their use entirely. In particular, the simplicity and computational efficiency of feature-based imitation make it appealing for online, interactive transfer, where the demonstrator's expertise can help to compensate for errors (see Section II-B). In our next experiment, we investigate the transfer of impedance (equilibrium position and stiffness) in this setting, through the approach described in Section III-A. The experimental setup is as follows.

Data, in the form of muscle activations, are collected from a human operator demonstrating simple movements and variations of impedance. More specifically, a pair of EMG sensors (surface EMG electrodes, Otto Bock), which are affixed to the wrist extensor and flexor muscles of the forearm (see Fig. 7), measure muscle activation at a 500-Hz sampling rate. The raw signals are 1) filtered through a bandpass filter to remove the lowest and highest frequency components and smooth out noise and 2) normalized so that the activation at rest corresponded to ${}^{e}\mathbf{u} = \mathbf{0}$, and maximum voluntary contraction (m.v.c.) corresponds to ${}^{e}\mathbf{u} = \mathbf{1}$, respectively.

For simplicity, the same muscle model, as described in the preceding section, is used to predict the demonstrator's impedance. Note that this model provides a minimalistic model of the muscle dynamics in terms of the activations, and has been widely used in the literature to predict impedance behavior of humans [6], [27], [54].

The model is adjusted to the demonstrator through a combination of direct measurement, and/or estimation of parameters according to existing biomechanical models. In particular, the mass of the free-moving link (i.e., hand) is estimated as $m = 400$ g, i.e., the average adult male hand mass [55], [56], the muscle stiffness, and damping properties (k_0, g_k, b_0, g_b) are taken from [6] (which in turn are based on earlier measurements of joint stiffness in humans [57]), the muscle pretension (i.e., $\mathbf{l}_{(q=0)} - \mathbf{l}_0$) is assumed zero at the rest posture, and the moment arms [$\mathbf{A}$ in (27)] are measured directly from the demonstrator's wrist. The only remaining free parameter is the muscle extension coefficient $\mathbf{g}_r$, which is manually adjusted to ensure that the (kinematic) response of the model matched that of the demonstrator when presented with the same inputs (i.e., when simulating the wrist using the demonstrator's recorded muscle activations as control inputs to the model).

The feature-based approach (as detailed in Section III-A) is applied, using (31) and (32) to estimate the demonstrator's impedance from the recorded muscle activations. The (estimated) impedance of the demonstrator is then transferred onto the robotic imitator. For the latter, the MACCEPA [2] is used as an illustrative example of a nonbiomorphic robotic actuator.

To illustrate performance, imitation is performed for 25 s of operation in which the demonstration is broken into distinct

phases: 1) alternating left–right hand movement with muscles relaxed, 2) alternation between low and high stiffness at $q = 0$ (sequentially relaxing and cocontracting muscles), and 3) alternating left–right hand movement with muscles tensed (i.e., high activation/cocontraction). Representative results are reported in Fig. 7, where the first and last conditions are indicated by the shaded regions in the plots.

As can be seen, during phase 1), the EMG signals indicate alternating activation between the two muscles [see Fig. 7(a)], resulting in a left–right movement of the wrist equilibrium position [see Fig. 7(c)]. The robot tracks this movement with considerable accuracy, albeit with a slight time delay, which we attribute to the limited speed of the servos used in the device. During phase 2), the hand remains at the rest position $q = 0$, and the operator cocontracts three times. As can be seen, this causes three spikes in the stiffness profile [see Fig. 7(d)], which are also accurately tracked. It is interesting to note in the plot of the commands to the MACCEPA [see Fig. 7(b)], the controller primarily relies on the second (pretensioning) motor for this, since there is a linear dependence between u_2 and stiffness at equilibrium. Finally, during phase 3), we again see good tracking of the equilibrium position with increased overall stiffness, despite the relatively high noise in the recorded EMG.

Finally, in all three phases, we note that, for each spike in muscle activations, there is a corresponding spike in the stiffness [see Fig. 7(d)]. This is in accordance with the accepted view in biomechanics that stiffness increases with muscle activation, even outside isometric conditions [35]. Here, this characteristic of human impedance behavior is reproduced on the robotic actuator.

C. Inverse Optimal Control From Human Data

In this final experiment, we apply the inverse OC approach to learning from a set of human demonstrations with the goal of transferring behavior to the Edinburgh SEA [see Fig. 3(c)]. For ease of comparison with the simulation studies (see Section IV-A), we study the same hitting task, in which the demonstrator attempts to hit a target (ball) as hard as possible, while minimizing effort. The goal is to learn a model of the human's objective function in order to transfer it to the robotic hardware. The experimental setup is as follows.

For collecting demonstrations, the measurement rig, shown in Fig. 8, is used. The rig consists of a hinge joint with a paddle attached, which is aligned to a ball suspended from a string. The rig has a handle which the demonstrator grasps to rotate the joint and hit the ball with the paddle. A magnetic motion sensor (Flock of Birds, Ascension Technology Corporation, Milton, VT, USA) is used to measure the angle of the demonstrator's wrist (hinge angle) at a 500-Hz sampling rate. Simultaneously, surface EMG sensors (as described in Section IV-B), which are placed on the antagonistic muscles of the demonstrator's forearm, measure the muscle activations of the demonstrator at the same 500-Hz rate. With this setup, trajectories of the human through state–action space are recorded, where the state is modeled as ${}^e\mathbf{x} = ({}^eq, {}^e\dot{q}) \in \mathbb{R}^2$, i.e., the instantaneous wrist angle and velocity and actions are modeled as

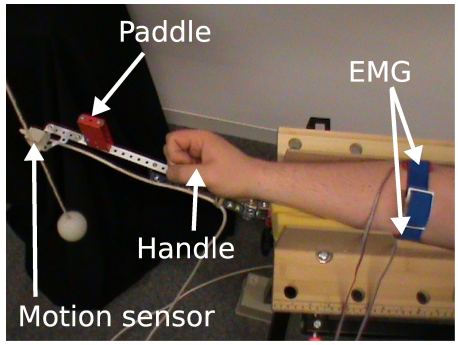

Fig. 8. Apparatus for recording human demonstrations of the hitting task.

the (feed-forward[15]) muscle activations ${}^e\mathbf{u} \in \mathbb{R}^2$, as measured by the normalized EMG.

Data are collected from a human attempting to hit the ball (suspended at a point corresponding to wrist angle $q^* = 34.0^\circ$) as hard as possible with the paddle, from a series of start positions, given a fixed time duration in which to complete the movement. Specifically, three trajectories are recorded from each of five start positions $q = \{10, 0, -10, -20, -30\}^\circ$, with a fixed duration of 0.2 s. To reduce the effects of noise and variability in the execution of the trajectories, the data are preprocessed by 1) smoothing the signals with a Butterworth filter and 2) temporal alignment of trajectories around the time of impact with the ball. The trajectories from each of the start states are then averaged, and the resultant $\mathcal{J} = 5$ mean trajectories used as training data for the learning.

Since our inverse OC approach requires a model of the expert's forward dynamics, the demonstrator's wrist dynamics are approximated using the same two-muscle wrist model, as described the preceding two sections, [i.e., with dynamics as computed from (25)–(29)], with the parameters optimized with respect to the normalized error between the recorded trajectories $D = \{({}^e\mathbf{x}_0^j, {}^e\mathbf{u}_0^j), \ldots, ({}^e\mathbf{x}_T^j, {}^e\mathbf{u}_T^j)\}_{j=0}^{\mathcal{J}}$ and those predicted by integrating the model under the same command sequence $\tilde{D} = \{({}^e\tilde{\mathbf{x}}_0{}^j, {}^e\mathbf{u}_0^j), \ldots, ({}^e\tilde{\mathbf{x}}_T{}^j, {}^e\mathbf{u}_T^j)\}_{j=0}^{\mathcal{J}}$.

For estimating the human objective, the cost function model (33) is used, with the best fit to the coefficients $\mathbf{w} = (w_1, w_2, w_3)^\top$ sought through MWAL. Note that, in this experiment, as ${}^e\tau$ cannot be directly measured during movement, we use the optimized parametric model to estimate the torques for the third term in (33) using (26). The model is trained on the demonstrated trajectories, with $\alpha = 300$ for 20 iterations. Note that, since the parameters of the true human cost function (i.e., $\mathbf{w}$) are unknown, we cannot explicitly calculate the error in the weight estimate. Instead, convergence can be measured by examining the magnitude of the weight update (i.e., $\Delta\tilde{w}$ in Algorithm 1). The results reported in the following are for the convergent estimate.

[15]Note that, while muscle activations recorded through EMG may also contain contributions from feedback controllers, in general, in the short duration, explosive movement considered here, such feedback contributions may be assumed to be negligible in the face of the inherent feedback delays of the human neuromuscular system [58].

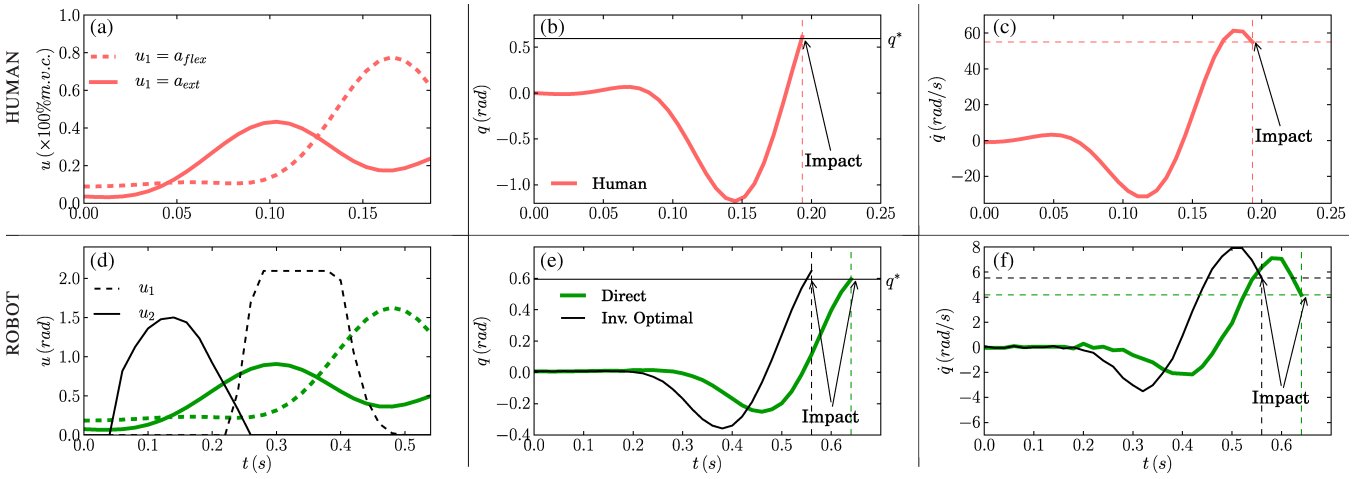

Fig. 9. Human demonstrated (a)–(c) ball-hitting behavior and (d)–(f) imitated robot behavior. In (d)–(f), the robot behavior that is generated through direct imitation is shown in thick green, and that generated with inverse optimal imitation is shown in thin black. (a) Filtered EMG readings of the human's forearm muscles during hitting. (b) Human wrist position until impact. (c) Wrist joint velocity. (d) Robot motor commands during imitation. (e) Robot joint positions. (f) Robot joint velocity. In (b) and (e), the thin black line marks the target position, and the dashed vertical lines mark the impact times. In (c) and (f), the dashed horizontal line marks the impact velocity, and the dashed vertical lines mark the impact times.

For evaluation, we compare the behavior of the robot when imitating behavior through 1) the inverse OC approach and, 2) the direct imitation approach. For the former, ILQG is used to find the optimal controller for the Edinburgh SEA with respect to the cost function [i.e., (33), using the learned weights]. For the latter, the human EMG signals are scaled according to the maximum admissible commanded angle of the robot motors, and then fed directly as commands to the robot, i.e., drawing the correspondence ${}^e\hat{\mathbf{u}} \equiv {}^l\hat{\mathbf{u}}$, where ${}^e\hat{\mathbf{u}}$ and ${}^l\hat{\mathbf{u}}$ are the commands normalized by the admissible ranges for the human and robot, respectively. Note also that, since the response of the robot's servomotors is significantly lower than that of the human (in terms of control frequency and other delays), control of the robot is scaled in time so that the command sequences have 0.5-s duration for both of the approaches compared.

The results are shown in Fig. 9 for an example trajectory starting at $q = 0°$. Looking at the joint angle and velocity profiles, we can see that the strategy used by the human is to first move the wrist away from the target before rapidly moving it in the positive direction toward the target. A similar movement occurs on the robot when using both the direct and the inverse-optimal approaches. However, comparing these, we see that for the direct approach, the amplitude of the movement is reduced, and the velocity at the time of impact is much smaller. In contrast, the inverse optimal approach optimizes the command sequence for the robot dynamics, resulting in earlier onset time for the movement, and a much larger movement of the motors (see Fig. 9). Consequently, it achieves a higher hitting velocity (with the ball traveling a greater distance) when executed on the robotic hardware. This can be verified in the accompanying video.

V. Discussion

In this paper, a study of competing methods to transfer behavior from humans to robots in the context of variable impedance control has been presented. We have illustrated the difficulties that this problem poses, given the inescapable heterogeneity between the human musculoskeletal system and robotic systems and analyzed the relative pros and cons of different approaches that may be employed to overcome these difficulties.

Approaches based on 1) *direct transfer*, 2) *feature-based tracking*, and 3) transfer based on *inverse OC* have been compared. The first, we may rule out in almost all cases unless the robotic system is highly biomimetic. The second two avoid this restriction but, as shown in our experiments, are best applied in different settings.

Our findings indicate that *feature-based tracking* can be effective in many settings where online, interactive control of the robotic device is required. This is the case, for example, in the teleoperation domain where behavior is transferred (in a supervised way) from a human operator. We have presented a model-based method for control of variable stiffness actuators using constraints on equilibrium positions and stiffness in task and joint space. The proposed approach is generic by its formulation, and can be applied to many different designs of variable stiffness devices for accurate tracking of desired stiffness and equilibrium position profiles. As shown in our experiments, it is fast to compute and can be used with ease for online behavior transfer, such as in the teleoperation setting explored here.

Outside such domains, however, our investigations show that transfer based on *inverse OC* can be more effective in dealing with heterogeneous dynamics and actuation between plants. Such an approach is effective for *task-oriented* behavior transfer, where we rather avoid prescribing specific features of behavior and instead require our system to derive its own *strategies* to meet task goals. We have presented a framework based on a two-step approach to learning, where in the first step, a parametric model of the objective function underlying observed behavior is learnt using AL. This enables us to find a task-based representation of the data in terms of the objectives (cost minimized), and then apply local OC techniques to find a similarly

optimal behavior for the imitator, taking into account the differences in dynamics and actuation. Our experiments show the effectiveness of this approach, where the proposed approach actually exploits the dynamics characteristics of the imitator in order to outperform the feature-based imitation approaches and, in some cases, even surpass the task performance of the expert.

A number of directions for future research remain. One issue to be investigated is that of scaling the different approaches to more complex tasks and plants. For example, with regard to feature-based imitation of impedance, it remains an open research issue as to the accuracy with which human impedance can be estimated, and thereby tracked, during complex, multijoint movements, e.g., during full body motion. Methods exploiting impedance observer techniques [59]–[61] and novel measurement devices [62], [63] may be exploited in future work. With regard to the inverse optimal approach, related issues of scalability exist in terms of the selection of cost basis functions. One such issue is the problem of finding appropriate cost bases to describe more *complex task objectives*. With this in mind, however, it should be noted that increased complexity of the *plant dynamics* does not necessarily equate to an increase in the complexity of the cost function. For example, for the task of hitting, even if the task is to be performed by a system with complex, nonlinear dynamics (such as full arm punching [15]), the indicators of task success (i.e., the cost bases) nominally remain the same (e.g., accuracy, impact velocity, and effort), albeit their functional form may be more complex to compute, and the (forward) optimization may become more difficult.

Another issue warranting further investigation is that of the selection and design of cost function models for the inverse optimal approach. At present, the selection of appropriate terms in the cost function is left open to the designer of the learning system (cf., Section II-C): She or he must make appropriate consideration of the important components of the task and the correspondence between demonstrator and imitator. While this was feasible in this study, it remains an open issue as to how to perform this selection in general.

In the hitting experiment, which is presented in Section IV-C, for instance, a cost function was chosen that is intuitively suitable for the task. It is not known whether this cost function can represent the demonstrator's true cost function without error. Crucially, however, the cost function chosen here is *flexible enough that imitation could have failed*: If the parameters had been incorrectly learned, the task would not have been reproduced. For example, if the learning outcome had been a high weight on the velocity term and a low weight on the accuracy term, then the resultant behavior would have been a "powerless" or "missed" hit (with high velocity toward the end of the movement, but poor accuracy, the robot would either hit the ball prematurely, or not at all[16]). This was not seen in the experiments reported here (the hitting task was correctly reproduced), lending support to the chosen cost model.

In general, however (and especially with more complex problems or greater heterogeneity in the dynamics), this issue of selection of the cost model will be less straightforward. An important direction of future work, therefore, is to look for robust ways of making this selection and to investigate the sensitivity of the choice of model with respect to 1) the differences in dynamics between the demonstrator and imitator and 2) the set of task outcomes afforded by optimizing behavior within the parameter space of that model. Nevertheless, a contribution of this paper is to illustrate that, under the right conditions, the inverse optimal approach can be a powerful alternative to the direct and feature-based approaches when dealing with behavior transfer across highly heterogeneous systems.

[16]This was seen in [47], where a poor dynamics model resulted in a poor estimate of the cost parameters.

ACKNOWLEDGMENT

The authors would like to thank T. Mori and J. Nakanishi for fruitful discussions regarding this study. They would also like to thank the anonymous reviewers for their insightful comments that helped to improve earlier versions of this study.

REFERENCES

[1] R. Schiavi, G. Grioli, S. Sen, and A. Bicchi, "VSA-II: A novel prototype of variable stiffness actuator for safe and performing robots interacting with humans," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 2171–2176.

[2] R. V. Ham, B. Vanderborght, M. V. Damme, B. Verrelst, and D. Lefeber, "MACCEPA, the mechanically adjustable compliance and controllable equilibrium position actuator: Design and implementation in a biped robot," *Robot. Auton. Syst.*, vol. 55, no. 10, pp. 761–768, 2007.

[3] A. Bicchi and G. Tonietti, "Fast and soft arm tactics: Dealing with the safety-performance trade-off in robot arms design and control," *IEEE Robot. Autom. Mag.*, vol. 11, no. 2, pp. 22–33, Jun. 2004.

[4] D. Braun, M. Howard, and S. Vijayakumar, "Exploiting variable stiffness in explosive movement tasks," in *Robotics: Science and Systems VII*. Cambridge, MA, USA: MIT Press, 2011.

[5] S. Yun, "Compliant manipulation for peg-in-hole: Is passive compliance a key to learn contact motion?" in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 1647–1652.

[6] M. Katayama and M. Kawato, "Virtual trajectory and stiffness ellipse during multijoint arm movement predicted by neural inverse models," *Biol. Cybern.*, vol. 69, pp. 353–362, 1993.

[7] N. Hogan, "Impedance control—An approach to manipulation. Part III—Applications," *ASME Trans. J. Dyn. Syst., Meas., Control*, vol. 107, pp. 1–24, 1985.

[8] D. Mitrovic, S. Klanke, and S. Vijayakumar, "Adaptive optimal feedback control with learned internal dynamics models," in *From Motor Learning to Interaction Learning in Robots*. New York, NY, USA: Springer-Verlag, 2010.

[9] S. Wolf and G. Hirzinger, "A new variable stiffness design: Matching requirements of the next robot generation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 1741–1746.

[10] M. Grebenstein, A. Albu-Schäffer, T. Bahls, M. Chalon, O. Eiberger, W. Friedl, R. Gruber, S. Haddadin, U. Hagn, R. Haslinger, H. Höppner, S. Jörg, M. Nickl, A. Nothhelfer, F. Petit, J. Reill, N. Seitz, T. Wimbock, S. Wolf, T. Wusthoff, and G. Hirzinger, "The DLR hand arm system," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 3175–3182.

[11] G. Ganesh, A. Albu-Schäffer, M. Haruno, M. Kawato, and E. Burdet, "Biomimetic motor behavior for simultaneous adaptation of force, impedance and trajectory in interaction tasks," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 2705–2711.

[12] C. Harris and D. Wolpert, "Signal-dependent noise determines motor planning," *Nature*, vol. 394, pp. 780–784, 1998.

[13] D. Franklin, G. Liaw, T. Milner, R. Osu, E. Burdet, and M. Kawato, "Endpoint stiffness of the arm is directionally tuned to instability in the environment," *J. Neurosci.*, vol. 27, pp. 7705–7716, 2007.

[14] M. Howard, D. Braun, and S. Vijayakumar, "Constraint-based equilibrium and stiffness control of variable stiffness actuators," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5554–5560.

[15] M. Howard, D. Mitrovic, and S. Vijayakumar, "Transferring impedance control strategies between heterogeneous systems via apprenticeship

learning," in *Proc. IEEE Int. Conf. Humanoid Robots*, 2010, pp. 98–105.
[16] U. Syed, M. Bowling, and R. Schapire, "Apprenticeship learning using linear programming," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1032–1039.
[17] P. Abbeel and A. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 1–8.
[18] T. Buchanan, D. Lloyd, K. Manal, and T. Besier, "Neuromusculoskeletal modeling: Estimation of muscle forces and joint moments and movements from measurements of neural command," *J. Appl. Biomech.*, vol. 20, pp. 367–395, 2004.
[19] B. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robot. Auton. Syst.*, vol. 57, pp. 469–483, 2009.
[20] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Handbook of Robotics*. Cambridge, MA, USA: MIT Press, 2007, ch. 59.
[21] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *Int. J. Robot. Res.*, vol. 23, no. 4, pp. 363–377, 2004.
[22] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosoph. Trans.: Biol. Sci.*, vol. 358, no. 1431, pp. 537–547, 2003.
[23] G. Klute, J. Czerniecki, and B. Hannaford, "Mckibben artificial muscles: Pneumatic actuators with biomechanical intelligence," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron.*, 1999, pp. 221–226.
[24] A. Alissandrakis, C. Nehaniv, and K. Dautenhahn, "Correspondence mapping induced state and action metrics for robotic imitation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 299–307, Apr. 2007.
[25] R. Kirsch, D. Boskov, and W. Rymer, "Muscle stiffness during transient and continuous movements of cat muscle: Perturbation characteristics and physiological relevance," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 8, pp. 758–770, Aug. 1994.
[26] A. Radulescu, M. Howard, D. Braun, and S. Vijayakumar, "Exploiting variable physical damping in rapid movement tasks," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron.*, 2012, pp. 141–148.
[27] D. Mitrovic, S. Klanke, R. Osu, M. Kawato, and S. Vijayakumar, "A computational model of limb impedance control based on principles of internal model uncertainty," *PLoS ONE*, vol. 5, no. 10, p. e13601, 2010.
[28] E. Burdet, R. Osu, D. Franklin, T. Milner, and M. Kawato, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, pp. 446–449, 2001.
[29] K. Mombaur, A. Truong, and J. Laumond, "From human to humanoid locomotion: An inverse optimal control approach," *Auton. Robots*, vol. 28, pp. 369–383, 2010.
[30] B. Ziebart, A. Maas, J. Bagnell, and A. Dey, "Maximum entropy inverse reinforcement learning," in *Proc. 23rd Nat. Conf. Artif. Intell.*, 2008, pp. 1433–1438.
[31] A. Ng and S. Russell, "Algorithms for inverse reinforcement learning," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 663–670.
[32] J. Vogel, C. Castellini, and P. van der Smagt, "EMG-based teleoperation and manipulation with the DLR LWR-III," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2011, pp. 672–678.
[33] T. Tsuji, K. Shima, N. Bu, and O. Fukuda, "Biomimetic impedance control of an EMG-based robotic hand," in *Robot Manipulators Trends and Development*, A. Jimenez and B. A. Hadithi, Eds. Rijeka, Croatia: InTech, 2010.
[34] B. Kang, B. Kim, S. Park, and H. Kim, "Modeling of artificial neural network for the prediction of the multi-joint stiffness in dynamic condition," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2007, pp. 1840–1845.
[35] D. Franklin, F. Leung, M. Kawato, and T. Milner, "Estimation of multijoint limb stiffness from EMG during reaching movements," in *Proc. IEEE EMBS Asian-Pacific Conf. Biomed. Eng.*, 2003, pp. 224–225.
[36] A. Albu-Schäffer, C. Ott, and G. Hirzinger, "A unified passivity-based control framework for position, torque and impedance control of flexible joint robots," *Int. J. Robot. Res.*, vol. 26, no. 1, pp. 23–39, 2007.
[37] A. D. Luca, R. Farina, and P. Lucibello, "On the control of robots with visco-elastic joints," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2005, pp. 4297–4302.
[38] M. Laffranchi, N. Tsagarakis, and D. Caldwell, "A variable physical damping actuator (VPDA) for compliant robotic joints," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 1668–1674.
[39] L. Odhner and H. Asada, "Scaling up shape memory alloy actuators using a recruitment control architecture," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 1675–1680.
[40] P. Chiacchio, S. Chiaverini, L. Sciavicco, and B. Siciliano, "Closed-loop inverse kinematics schemes for constrained redundant manipulators with task space augmentation and task priority strategy," *Int. J. Robot. Res.*, vol. 10, no. 4, pp. 410–425, 1991.
[41] D. Braun and M. Goldfarb, "Eliminating constraint drift in the numerical simulation of constrained dynamical systems," *Comput. Methods Appl. Mech. Eng.*, vol. 198, no. 37–40, pp. 3151–3160, 2009.
[42] F. Petit and A. Albu-Schäffer, "State feedback damping control for a multi DOF variable stiffness robot arm," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5561–5567.
[43] A. Ng, "Reinforcement learning and apprenticeship learning for robotic control," in *Algorithmic Learning Theory*. New York, NY, USA: Springer-Verlag, 2006, pp. 29–31.
[44] D. Ramachandran and E. Amir, "Bayesian inverse reinforcement learning," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, 2006, pp. 2586–2591.
[45] G. Neu and C. Szepesvári, "Apprenticeship learning using inverse reinforcement learning and gradient methods," in *Proc. Uncertainty Artif. Intell.*, 2007, pp. 295–302.
[46] V. Freire da Silva, A. Reali Costa, and P. Lima, "Inverse reinforcement learning with evaluation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2006, pp. 4246–4251.
[47] T. Mori, M. Howard, and S. Vijayakumar, "Model-free apprenticeship learning for transfer of human impedance behaviour," in *Proc. IEEE Int. Conf. Humanoid Robots*, 2011, pp. 239–246.
[48] J. Peters and S. Schaal, "Learning operational space control," in *Robotics: Science & Systems.* Cambridge, MA: MIT Press, 2006.
[49] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *Proc. Amer. Control Conf.*, 2005, pp. 300–306.
[50] J. Kober and J. Peters, "Policy search for motor primitives in robotics," *Mach. Learn.*, vol. 84, pp. 171–203, 2011.
[51] E. Theodorou, J. Buchli, and S. Schaal, "Reinforcement learning of motor skills in high dimensions: A path integral approach," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 2397–2403.
[52] M. Lagoudakis and R. Parr, "Least-squares policy iteration," *J. Mach. Learn. Res.*, vol. 4, pp. 1107–1149, 2003.
[53] D. Mitrovic, S. Klanke, and S. Vijayakumar, "Learning impedance control of antagonistic systems based on stochastic optimisation principles," *Int. J. Robot. Res.*, vol. 30, pp. 556–573, 2011.
[54] E. Nakano, H. Imamizu, R. Osu, Y. Uno, H. Gomi, T. Yoshioka, and M. Kawato, "Quantitative examinations of internal representations for arm trajectory planning: Minimum commanded torque change model," *J. Neurophysiol.*, vol. 81, pp. 2140–2155, 1999.
[55] R. Chandler, C. Clauser, J. McConville, H. Reynolds, and J. Young, "Investigation of inertial properties of the human body," Aerosp. Med. Res. Lab, Wright-Patterson AFB, OH, USA, Tech. Rep. AD-A016485, 1975.
[56] C. Clauser, J. McConville, and J. Young, "Weight, volume, and center of mass of segments of the human body," Aerosp. Med. Res. Lab, Wright-Patterson AFB, OH, USA, Tech. Rep. AMRL-TR-69-70, 1969.
[57] D. Bennett, J. Hollerbach, Y. Xu, and I. Hunter, "Time-varying stiffness of human elbow joint during cyclic voluntary movement," *Exp. Brain Res.*, vol. 88, pp. 433–442, 1992.
[58] A. van Soest and M. Bobbert, "The contribution of muscle properties in the control of explosive movements," *Biol. Cybern.*, vol. 69, pp. 195–204, 1993.
[59] G. Grioli and A. Bicchi, "A real-time parametric stiffness observer for VSA devices," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5535–5540.
[60] S. van Eesbeek, E. de Vlugt, and M. Verhaegen, "Time variant subspace identification of joint impedance," in *Proc. 23rd Congr. ISB*, 2011.
[61] A. Serio, G. Grioli, I. Sardellitti, N. Tsagarakis, and A. Bicchi, "A decoupled impedance observer for a variable stiffness robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5548–5553.
[62] E. Rouse, L. Hargrove, E. Perreault, and T. Kuiken, "Estimation of human ankle impedance during walking using the perturbator robot," in *Proc. IEEE Int. Conf. Biomed. Robot. Biomechatron.*, 2012, pp. 373–378.
[63] H. Hoppner, D. Lakatos, H. Urbanek, C. Castellini, and P. van der Smagt, "The grasp perturbator: Calibrating human grasp stiffness during a graded force task," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 3312–3316.

Matthew Howard received the M.Sci. degree in physics from Imperial College London, London, U.K., in 2004 and the M.Sc. degree in artificial intelligence from Edinburgh University, Edinburgh, U.K., in 2005. He also received the Ph.D. degree in machine learning and robotics from Edinburgh University in 2009, in partnership with the Honda Research Institute, Germany.

He is currently a Japan Society for the Promotion of Science Postdoctoral Research Fellow with the Nakamura Lab, Deparment of Mechano-Informatics, University of Tokyo, Tokyo, Japan. Prior to this, he was with the Institute for Perception, Action and Behaviour, School of Informatics, Edinburgh University. His research interests span various topics in machine learning, robotics, and control.

David J. Braun received the M.Sc. degree in applied mechanics from the University of Novi Sad, Novi Sad, Serbia, in 2003 and the Ph.D. degree in mechanical engineering from Vanderbilt University, Nashville, TN, USA, in 2009.

He is currently a Postdoctoral Research Fellow with the Institute for Perception, Action, and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, U.K. Prior to this, he was with the Centre for Intelligent Mechatronics, Vanderbilt University. His research interests span issues in system dynamics and control, mechatronics, and robotics.

Sethu Vijayakumar received the Ph.D. degree in computer science and engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1998.

He is currently a Professor of robotics and the Director of the Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, U.K. Since 2007, he has been a Senior Research Fellow of the Royal Academy of Engineering in Learning Robotics, co-sponsored by Microsoft Research Cambridge. He is also an adjunct faculty member with the University of Southern California, Los Angeles, and a Visiting Research Scientist with the RIKEN Brain Science Institute, Tokyo. His research interests span a broad interdisciplinary curriculum ranging from statistical machine learning, robotics, planning, and optimization in autonomous systems to motor control and computational neuroscience.

Dr. Vijayakumar was elected as a Fellow of the Royal Society of Edinburgh in 2013.