



Design, Requirements, and Challenges of a Human-Robot Imitation System

Darja Stoeva , Margrit Gelautz 

Abstract

Body motion is an important aspect in human-robot interactions. Giving robots the ability to imitate human motion can be beneficial for research on robot motion in a variety of applications. Furthermore, such a human-robot imitation system has the potential to provide a platform to investigate the facilitation of different types of trust in human-robot interactions. The goal of this paper is to describe the framework of a human-robot imitation system and investigate the system requirements imposed by different interaction settings. Several applications of imitation systems are discussed, along with their important characteristics and required features. Furthermore, open challenges for designing and developing human-robot imitation systems are discussed.

Keywords

human-robot interaction, body motion, imitation

1 Introduction

Because humans tend to assign social meaning to movements, human body movement is frequently perceived as expressive, making body motion an important component in social interactions. This tendency has been demonstrated in human interactions [Argyle 1975] as well as in situations where humans observed interactions between inanimate objects (moving geometrical shapes) [Heider and Simmel 1944]. Within the field of human-robot interaction, robot body movement and nonverbal behavior have been shown to influence how the robot is perceived in terms of animacy [Fukuda and Ueda 2010; Rosenthal-von der Pütten et al. 2018], anthropomorphism [Salem et al. 2013], feeling of co-presence [Krämer et al. 2016], and children's perceptions of a robot's warmth and competence [Peters et al. 2017].

Body movements designed for robots are often inspired by human body movements. It has been demonstrated that people prefer to interact with robots that exhibit human-like behavior over robots that exhibit machine-like behavior [Park et al. 2011]. Furthermore, when it comes to the dynamics of human-robot interactions, research has shown that people are more likely to coordinate their movement when interacting with a humanoid robot rather than a mechanical one [Chaminade et al. 2005] and when the motion is human-like rather than machine-like [Chaminade et al. 2008]. The dynamical feature of movement coordination is an important aspect of social interactions because it influences whether the interaction is perceived negatively or positively, which has a direct impact on the efficiency and stability of the interaction [Burgoon et al. 1995; Schmidt et al. 2012].



Human body movements are categorized with respect to their expressiveness by [Karg et al. 2013] as *communicative*, when they express or convey a message, *functional*, when they are used to accomplish a particular task, artistic, when they express a message in an exaggerated manner or when they are described as unfamiliar when compared to daily movements, or abstract, when they neither express a message nor serve a functional purpose. In the field of human-robot interaction, imitation, which is described as the ability of a robot to replicate human movement [Schaal 1999], serves as a promising tool to generate human-like movements. In principle, imitation systems provide a method to design and develop robotic body movements in any of the aforementioned categories of human movement. As a result, human-robot imitation systems can be used as an interactional framework to study body motion in human-robot interactions.

Furthermore, depending on the interaction setting for the system's intended application, human-robot imitation systems have the potential to provide a platform for studying different types of trust. There are two types of trust which would be applicable in this context, (1) interpersonal trust, which describes trust in social interactions based on the relationship that develops among the interactants [Ogawa et al. 2019], and (2) reliance trust, which is based on the belief that the robot will function as expected [Coeckelbergh 2012]. As a result, interpersonal trust can be studied in systems designed for social interactions, and reliance trust can be studied in systems designed for cooperative interactions. Using the system in various interaction settings may also allow for a comparison between these two different types of trust that can be facilitated in human-robot interactions.

The work presented here aims to propose an approach for the design and development of a human-robot imitation system with an intended application in mind. The main contribution is describing a framework for the design, development and evaluation of a human-robot imitation system. A second contribution is extending several existing applications of imitation systems, such as teleoperation and imitation learning, to also account for aspects such as interpersonal coordination, movement data collection, and exploration of body movements. In this context, we also include applications in the performing arts, which are not a very common point of interest in the field of robotics research. Finally, as a third contribution, a link between the envisioned applications and the system requirements of the proposed framework is established, which could aid the development process of future imitation systems. The paper is structured as follows. First, we describe the framework of a human-robot imitation system (Section 2), then we identify the application-dependent requirements of such a system for several potential applications (Section 3), followed by a discussion of open challenges (Section 4), and finally we provide a general conclusion (Section 5).

2 Framework of a Human-Robot Imitation System

Alissandrakis et al. [Alissandrakis et al. 2002] describe an agent-based perspective on the design of an imitation system that addresses five central questions: who, when, what, how to imitate, and how to evaluate the quality of imitation. The question of who to imitate refers to figuring out how to allow the robot to choose which interactant to imitate, especially in the case of multiple interactants. When to imitate refers to the times the robot needs to imitate and which movements within a behavior need to be imitated by the robot. Next, the system should consider what to imitate as part of an observed behavior, such as states, actions, and so on. How to imitate addresses the issue of mapping behavior from human embodiment to robotic embodiment. Finally, the question of how to evaluate the imitation is about finding a suitable metric to evaluate the similarity between the demonstrated and the resulting imitated behavior. Each of these questions has challenges and specific requirements depending on how the imitation system is intended to be used.

Such an agent-based approach is typically considered in the case of autonomous robots and aims to provide an approach independent of the robotic platform and the imitation task. In contrast, we argue in our work that the robotic platform, imitation task and system application all play an important role in the design and development process of a human-robot imitation system. Moreover, the majority of imitation systems considered in the literature are primarily aimed at applications of imitation learning or teleoperation. As opposed to that, in our research we extend the possible applications of an imitation system for humanoid robots and their usage scenarios while we lay out the framework of an imitation system from a developmental perspective.

The proposed framework for a human-robot imitation system is divided into three main components: (1) an *intended application* of the system, (2) a *technical implementation* with considerations based on the intended application, and (3) a suitable evaluation method based on the distinctive features of the imitation type. A flowchart of the suggested components for a human-robot imitation system is shown in Figure 1.

Because different interaction settings and tasks will have different system requirements, the intended application is one of the system's key components, making the technical implementation and method of evaluation application-dependent. Within the technical implementation, there are two important sub-components: (2.1) a means to sense human motion, and (2.2) a method that *translates the observed human motion into robot motion* (also shown in Figure 1).

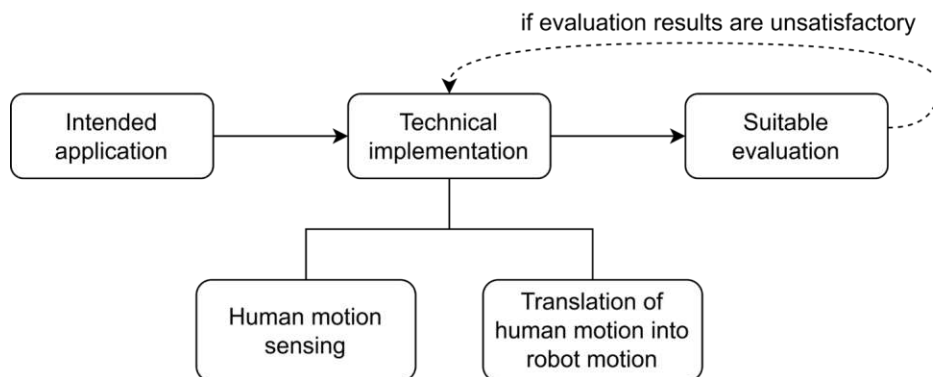


Figure 1 Flowchart showing the components of a human-robot imitation system

These two components are necessary in order to allow the robot to imitate human motion, and the choice of each of the components will affect the overall system performance. After the system has been implemented, a suitable evaluation should be carried out, which would be dependent on the specified system requirements and the distinctive features of the imitation type. Once the evaluation has been performed, depending on the results, certain improvements to the imitation system might be necessary, leading the development of such a system to undergo another cycle of (refined) technical implementation and evaluation. The subsections that follow go into greater detail about each of the components.

2.1. Intended Application

The first thing to consider when designing and developing a human-robot imitation system is the intended application. The importance of the application in the development of an imitation system has less to do with the end-goal and more to do with the interaction setting, which can be more interpersonal (e.g., mirroring) or more cooperative (e.g., teleoperation). The interaction setting is important because of the specific requirements that are required in various application contexts, which would define the necessary distinctive features of the imitation system. Some applications have stricter requirements while others provide more room for exploration. Additionally, the application also determines which methods are suitable candidates for system evaluation.

The potential applications of human-robot imitation systems vary depending on the interaction setting and the goal to be achieved with the imitation. Examples of such applications include *imitation learning* [Calinon and Billard 2007], *teleoperation* [Zuher and Romero 2012], *interpersonal coordination* (mirroring and synchrony) in social interactions [Hasumoto et al. 2020; Alibeigi et al. 2017],

movement data collection for interactive scenarios and expressive behavior (e.g., building datasets for nonverbal behavior), *exploration of body movements*, and the *performing arts* [Nakazawa et al. 2002]. Section 3 delves deeper into each of these applications in terms of technical requirements and important system evaluation features.

2.2. Technical Implementation

Following the selection of an application, the next step is to define the interaction settings, which will influence the method of human motion sensing in terms of joint positions, as well as the translation of human motion into robot motion in terms of converting joint positions to joint angles. The latter is required because the motor commands for robots are usually specified in terms of joint angles. In addition, to make the conversion from joint positions to joint angles feasible, the human joint positions need to be derived and processed in 3D space. The two most common methods for detecting 3D human joint positions are the use of motion capture systems such as Vicon¹ and the use computer vision algorithms for human pose estimation. Motion capture systems include camera based systems which comprise markers, attached to specific body parts (e.g., joints), and multiple cameras to track the markers and provide their positions in 3D space, or systems based on inertia sensors positioned on body parts without the need for external cameras. Computer vision algorithms are markerless pose estimators and provide joint positions directly in 3D space, such as the Kinect skeleton tracking module [Shotton et al. 2011], or estimate the 3D joint position from 2D body pose estimation [Mehta et al. 2017], or provide joint positions in the 2D camera space [Cao et al. 2019], which can then be used in combination with a depth-sensing camera to get the joints in 3D space (e.g., [Zabala et al. 2020]).

When choosing a method for sensing human motion it is important to consider the application scenario, which may impose different system requirements for obtaining the 3D human joint position data (depth-sensing camera, motion capture setups, or other means of sensing). The choice of method for motion sensing also includes the choice between usually more accurate sensing of 3D human joint positions in the case of motion capture systems, or allowing the human to move more freely in the case of using computer vision algorithms for human pose estimation.

Next, a model for converting joint positions to joint angles should be selected according to the system requirements imposed by the imitation goal of the application context. The imitation goal affects on the choice of the imitation type, which is closely connected to the method of system evaluation. The imitation

¹ <https://www.vicon.com/>

type can aim to preserve the position of the end-effector with respect to the body [Zuher and Romero 2012], the pose for achieving body pose matching [Stoeva et al. 2021], or both [Alibeigi et al. 2017]. The choice of a model will depend on the imitation type but also the system's efficiency and allowed time for a delay in the imitated movement often need to be considered. The temporal aspect varies from aiming for real-time, to a specific tolerable time delay which can be further relaxed for offline applications.

Different approaches can be found for the model used to translate 3D human joint positions into robot joint angles. In the fields of robotics and mechanics, there are two main kinematic equations used for translating between 3D positions and angles: forward kinematics, which is the calculation of the 3D (end-effector) position given the joint angles, and inverse kinematics, which is the calculation of the joint angles given the 3D position. Methods for calculating the robot's joint angles based on the inverse kinematics include analytic and numeric solutions [Lynch and Park 2017; Craig 2005]. Numerical approaches are usually based on iterative algorithms that try to solve the inverse kinematics as an optimization problem (e.g., using the Jacobian). Analytical solutions, on the other hand, are usually approached in two ways, using geometry to find the angle between the links connecting two joints, or using algebra to express the angles in equations derived from forward kinematics. Both analytical and numerical approaches have advantages and disadvantages. For instance, numerical solutions are oftentimes much slower due to their inherent iterative nature and are highly dependent on the initial guess of joint angles. In contrast, even though analytical approaches provide closed-form solutions, it could be that they are too complex to manipulate into solvable equations. After choosing a suitable mode, the technical system can be considered complete and consists of two modules (2.1) a method for *sensing human motion* and (2.2) a model for *translating this motion to a robotic platform*, as depicted in Figure 1.

2.3. Suitable Evaluation

The next step in system design and development is to adequately choose a suitable method of evaluating the imitation system. Depending on the modules chosen during the implementation phase, it may be necessary to evaluate the accuracy of each module separately, before evaluating the full imitation system. For example, one approach is to evaluate the chosen model that translates the movements on how accurately it estimates joint angles from joint positions.

The first thing to consider for the full imitation system evaluation is the distinctive features of the imitation type, which are most commonly either the accuracy of the end-effector position with respect to the body, the body pose similarity

between the human pose and the imitated robotic pose, or both. The second factor to consider (if applicable) is the computational effort or time delay, which is the amount of time it takes for the imitation system to capture the human motion, translate it to robot motion, and send it to the robot as a motion command.

The evaluation methods can include quantitative, qualitative, a mix of both quantitative and qualitative, and subjective measurements. Quantitative measurements usually include the computation of the cosine similarity for the angular configuration of the pose [Guo et al. 2019; Zhang et al. 2018; Alibeigi et al. 2017], the mean squared error of the targeted versus actual joint positions [Guo et al. 2019; Zhang et al. 2018; Alibeigi et al. 2017], and the computation effort [Koenemann et al. 2014]. Qualitative measurements, on the other hand, usually include trajectory plotting of the X, Y and Z axis of the end-effector [Hirschmanner et al. 2019; Alibeigi et al. 2017; Mukherjee et al. 2015], plotting of the total error over time [Zhang et al. 2016; Koenemann et al. 2014], visual images, usually of motion sequences or specific postures, of the human posture and the robot exhibiting the imitated posture side by side [Guo et al. 2019; Zhang et al. 2018; Alibeigi et al. 2017; Zhang et al. 2016; Kim et al. 2016; Mukherjee et al. 2015; Ou et al. 2015]. Subjective measurements are typically based on user studies in which participants are asked to rate the quality of imitation by showing images or videos of the actual and imitated movement [Zuher and Romero 2012]. Depending on the intended application of the system and the imitation type, a suitable evaluation should be designed and performed. A good practice in evaluations of systems is to combine several methods of evaluation.

3 Application-dependent Requirements

As mentioned in the previous section, the development of a human-robot imitation system is highly dependent on the interaction settings of the system's application. Table 1 shows some of the potential applications for imitation systems with their system requirements, such as the methods of *human motion sensing*, the *imitation type*, the *time delay* between the performed and imitated movement, the evaluation features important for the evaluation process, and the *trust type* that can be facilitated and studied. In Table 1, the abbreviation "CV" stands for computer vision in the *human motion sensing column*, while in the *imitation type column*, "task-dependent" indicates that the choice of imitation depends on the targeted task, "both" stands for a compromise of preserving the end-effector position and body pose matching, and "any" stands for preserving any of the three imitation types explained in Subsection 2.2. The following subsections look into each of the suggested applications in relation to the aforementioned system requirements in the context of the interaction setting.

| Applications | Human motion sensing | Imitation type | Time delay | Evaluation features | Trust type |
|--------------------------------------|-------------------------------|--------------------|---------------------|---|-------------------------|
| <i>Teleoperation</i> | CV algorithms, motion capture | both | no delay | end-effector position, body pose similarity, time delay | reliance |
| <i>Imitation learning</i> | CV algorithms, motion capture | task-dependent | no delay | imitation feature, time delay | reliance |
| <i>Interpersonal coordination</i> | CV algorithms | body pose matching | from no to 5s delay | body pose similarity, time delay | interpersonal |
| <i>Movement data collection</i> | CV algorithms, motion capture | body pose matching | flexible | body pose similarity, time delay | reliance |
| <i>Exploration of body movements</i> | CV algorithms | any | no delay | open | interpersonal |
| <i>Performing arts</i> | CV algorithms, motion capture | any | flexible | open | reliance, interpersonal |

Table 1 Potential applications of human-robot imitation systems with their system requirements and characteristics

3.1. Teleoperation

In situations in which the human operator cannot be physically present or in dangerous environments such as search and rescue, the method of robot teleoperation is envisioned as a possible approach [Penco et al. 2019; Koenemann et al. 2014; Stanton et al. 2012]. Due to the necessity of exact mapping of the human motion to the robot and the required high accuracy of the end-effector position with respect to the human body, an imitation system targeting teleoperation requires a high level of *human motion sensing accuracy*. For this application, a motion capture system usually provides more accurate readings than the use of available computer vision methods for the estimation of human pose. Motion capture often requires a specific interaction setting that typically includes several sensors or markers that need to be positioned on the human body, resulting in less spatial freedom and possibly discomfort for the interactant. This may not be an issue if the human and the robot are not interacting with each other face-to-face, which is usually the case for teleoperation. On the other hand, the accuracy of the involved human pose estimation algorithm has a significant impact on imitation performance when using computer vision methods. If computer vision is

the preferred method due to specific task requirements, the accuracy of the pose estimation algorithm can be evaluated using motion capture data as a reference. Since the idea behind robot teleoperation is for the human to be embodied in the robotic platform, the *imitation type* should preserve both end-effector position and body pose matching. The imitation should be performed with no *time delay* to allow for smooth control and quick feedback when controlling the robot. Thus, when evaluating an imitation system for teleoperation, the most important considerations for the *evaluation features* are end-effector position accuracy, body pose similarity metrics, and time delay. The *type of trust* that can typically be facilitated in this application is system reliance. For example, examining different types of teleoperation control and their influence on the trust of the system [Saeidi et al. 2017] or how different time delays affect the facilitated trust in the system [Rogers et al. 2017]. Ideally, for providing additional information to the human controller in order to ease the process of teleoperation, the imitation system should also include a virtual reality headset (e.g., [Hirschmanner et al. 2019]) and haptic force feedback (e.g., [Saeidi et al. 2017]).

3.2. Imitation Learning

The concept of using imitation learning (also known as learning from demonstration or programming by demonstration) as a method of teaching a robot to perform certain actions or behaviors stems from social learning in human interactions [Nehaniv and Dautenhahn 2007]. Researchers believe that robots capable of reproducing human movement could have advantages not only in allowing experts and non-experts to program behaviors for robots, but also as a means to better understanding of the concept of social learning [Breazeal and Scassellati 2002]. In an interaction setting where a robot needs to observe a human and learn specific behaviors, the important challenges to consider are how to successfully transfer the movement from the human to the robotic platform and which parts of the movement need to be reproduced. For such interaction settings, the use of motion capture or computer vision algorithms for human pose estimation is a common choice [Argall et al. 2009; Lee 2017]. However, to ensure that the required accuracy for imitation learning is met, both methods of *human motion sensing* should be evaluated in terms of achievable joint position accuracy. Because the interactant usually teaches the robot how to interact with the environment, in the context of imitation learning, the *imitation type* should usually preserve the position of the end-effector. However, in certain situations, depending on the task or behavior that needs to be imitated, it could be that both the end-effector position and body pose need to be maintained. Consequently, the choice of imitation type will depend on the task that needs to be completed or learned by the robot. In addition, the imitation should not have a noticeable

time delay between the interactant's demonstrated behavior and the imitated behavior by the robot. Similarly to teleoperation, the delay between the original and imitated motion is important for synchronized robot control. When controlling the robot to perform a particular task, immediate visual feedback is required when the motion needs to be corrected in appropriate time. This is especially important for novice users, and perhaps less so for experienced interactants as they may be able to adjust to how the system works more easily. The *evaluation features* that need to be considered for this application context should include methods for evaluating the accuracy of the imitation type and measurement of the time delay. As for the concept of *trust*, an imitation system for imitation learning can provide a platform for studying reliance trust, where the interactant would evaluate whether the system works as expected in both short and long term interactions. Another approach to studying trust in such systems is to investigate different methods of providing explanation about robot behavior and its effect on the facilitated trust in the system [Edmonds et al. 2019].

3.3. Interpersonal Coordination

When interacting socially with a robot, it is important for the interaction to be intuitive and smooth, meaning that both the human and the robot mutually influence and adapt to each other's behaviors. Interpersonal coordination, which includes mirroring and synchrony, is a phenomenon observed in human interactions as patterns that contribute to movement coordination and adaptation among interactants [Burgoon et al. 1995]. For *human motion sensing*, given the spatial restriction imposed by motion capture systems and the use of wearable markers or sensors, it might be preferable for interpersonal communication involving face-to-face interaction to rely on computer vision methods. This way, the interactant does not have to pay attention to the sensors/markers and will feel more comfortable to move and interact freely. Ideally, an internal (built-in) camera would be used, as no additional external equipment would be required. However, depending on where it is placed on the robot, the use of an internal camera has the potential to introduce further restrictions. Often cameras are positioned on a movable robot body part, for instance, the robot Pepper² has a depth camera placed in its head at the location of its 'eyes'. This can cause instability of the camera stream and, as a result, interfere with the data when the robot moves its head and perceives at the same time. When mirroring human motion, the system's *imitation type* should preserve body pose matching with the least amount of *time delay*. Compared to imitation learning and teleoperation, where the control of the robot requires no delay, for interpersonal coordination, the requirements on the mirror-

² <https://www.softbankrobotics.com/emea/en/pepper>

ing behavior are more relaxed allowing for the time delay to range from no delay to 5 seconds. This time range comes from research in human interaction [Sato and Yoshikawa 2007; Louwerse et al. 2012], and it has also been investigated in human-robot interactions [Shimada et al. 2008]. Another significant difference from the applications of teleoperation and imitation learning is the complexity of interpersonal coordination within social interactions. In this case, the question of which body parts and when they should be imitated would need a greater consideration compared to imitation learning and teleoperation. The *evaluation features* for an interpersonal coordination system should use body pose similarity metrics and measurements of the time delay. Additionally, a user study can be designed to address the subjectivity of the perceived pose, which may include a collection of body pose similarity ratings as it was done in [V. Tuyen et al. 2018; Zuher and Romero 2012]. As interpersonal coordination usually manifests itself in social interactions, it provides a platform to study interpersonal *trust*, for instance how mirroring and synchrony behaviors affect the facilitated trust between the human and the robot. It is also important to note that privacy concerns arise in the context of social interactions. People who interact with the robot should be aware of any possible further usage of their data collected during the interaction.

3.4. Movement Data Collection

Translating human movement into robot movement is useful for designing and implementing body movements for interactive scenarios and expressive behavior for robots, especially nonverbal behavior. The ability to convert human motion into robot motion serves as a bridge and as a means for building datasets [Lee 2017] or potentially as a way to design expressive behavior for the targeted robotic platform [Fischer 2021; V. Tuyen et al. 2018; Liu et al. 2012; Häring et al. 2011]. The recorded and possibly annotated datasets can then be used to develop methods for recognizing and generating a nonverbal behavior of robots. Similar to teleoperation and imitation learning, the methods for *human motion sensing* can either rely on motion capture systems or computer vision algorithms for human pose estimation. In the best case scenario, for better recognition accuracy, the method of human motion sensing used to build the dataset should be the same as the one to be used in the application scenario. In order to generate human-like body movements, the *imitation type* in such systems should be body pose matching, so, as for interpersonal coordination, the *evaluation features* should include body pose similarity metrics and user studies. However, unlike the interaction setting for interpersonal coordination, in this case the interaction setting would not necessarily require a real-time interaction. Thus, there could be a more flexible requirement for the *time delay* between the human movement and the imitated movement by the robot. However, the *evaluation features* could also

include a measurement of the time delay. The observed delay could be a useful indicator of the overall system performance and allow for comparison with other imitation systems. The *type of trust*, in this case, would be system reliance, and a particularly interesting approach would be to study how the reliance on the system can have a feedback effect on the movement of the human being imitated.

3.5. Exploration of Body Movements

An imitation system could be useful for an overall exploration of the way the robot moves and getting a sense of its movement range, especially for novice users. Providing an interactive framework for movement exploration that relies on imitation could aid the interactant in understanding how the robot moves. This can support the creation of mental models of robotic behaviors and simulations of their movement capabilities. Furthermore, such a system could, under the supervision of a physical therapist, potentially be used in movement therapy, which usually consists of movement exercises (e.g., improvisation) designed to explore the physical capabilities of the human body [Halprin 2003]. Additionally, such a system can also be used as a way to promote social skills for individuals with autism spectrum disorder as it has been done in [Vallée et al. 2020; Boucenna et al. 2014]. For the application of body movement exploration, the person being imitated should be free to move around in space and interact in an unrestricted manner. Thus, similarly to interpersonal coordination, for *human motion sensing* the use of computer vision algorithms is preferable to motion capture setups. Because of the interaction setting, it is important that the imitation happens in real-time so that the observing-acting cycle is maintained. Accordingly, there should be no *time delay* in the movement imitation. Given the importance of how the body moves in this application, any of the three *imitation types* may apply, thus the *evaluation features* should be chosen accordingly. For body pose matching, the important feature for the evaluation would be the body pose similarity metrics, for preserving the position of the end-effector it would be the accuracy of the end-effector position. If the system is to be used in a therapeutic setting, it is also important to include experts (therapists) in the design and development process of the imitation system. For the *type of trust*, as the robot will play the role of an interactional partner with which an interpersonal trust can be facilitated, a possible investigation could be the link between trust and the success of movement therapy or improvement in social skills. Another approach could be to look into a possible relationship between the length of time spent interacting or moving with the robot and the facilitated interpersonal trust over time.

3.6. Performing Arts

A human-robot imitation system seems like an interactive platform that is likely to be an attractive tool for the performing arts. The reason for this is due to its ability to facilitate the processes of choreography development and performance preparation, among other things [Christiansen and Lindelof 2020]. Unlike teleoperation, imitation learning, and interpersonal coordination, which all have rather specific interaction setting and requirements, in the case of performing arts the approach is less restricted and allows for many different requirements to be considered. For instance, when the interaction setting is exploratory, the application of performing arts may have flexible requirements, but it can also have very strict requirements, as in choreographed dance. Therefore, the *imitation type* in a system with an envisioned application in performance could be approached in an experimental way, and the *evaluation features* would depend on the requirements of the artists interacting with the system, as well as the performance itself. The choice of a *human motion sensing* method would depend on the requirements and vision of the artist interacting with the system. The possibilities include a motion capture system or computer vision algorithm. However, it should be considered that performers often move their bodies in unpredictable and unconventional ways, for instance suddenly falling on the ground with full force. Thus, in those situations to avoid damaging wearable sensors or markers computer vision methods might be more favorable. In this case, the *time delay* between the performed and imitated movement is rather flexible, especially if the interaction setting is exploratory. When the imitation includes some delay, the artist may discover interesting movement responses by the robot. Similar to the exploration of body movements with an imitation system, in the case of performing arts, the *imitation type* can be preserving body pose matching, the end-effector position, or both. The imitation system should be evaluated using *evaluation features* chosen according to the imitation type and the artist's requirements. When developing an imitation system for a specific artist, or performance preparation, it is important to include the artist or art director in the design and development process of the system. Regarding trust, there is potential for both reliance and interpersonal trust to be facilitated in the case of performance, again depending on the interaction setting.

4 Discussion

Body motion is an important ability that allows for the fulfillment of different types of actions. Enabling robots to use body motion as a way to communicate and interact with humans is a promising behavior for a fluid and intuitive HRI. With the high relevance and increased research interest in nonverbal behavior for the

design of future robots, it is important to consider which human behaviors are appropriate to adopt to robot behaviors. To allow for such research possibilities, we propose a framework of a human-robot imitation system in the simplest form that can serve as a foundation on which more complex behaviors can be developed. This is partly inspired by the works of [Jordanous 2020] and [Brooks 1991], which argue for incremental development of robot behavior, where each behavioral layer adds more complexity to the robot's capabilities.

Human-robot imitation systems have a wide range of applications from which many different research paths emerge. The future technical development of imitation systems highly depends on the advances in motion capture systems and robotic body design. From a broader perspective, building upon an imitation system has the potential to provide platforms for a better understanding of how artificial agents (robots) and humans exchange movements, how they differ from human interactions and how they can contribute to our understanding of body motion. In this spirit, human-robot imitation systems could also provide further insights into the role body motion has in human interactions.

Even though imitation behavior is a promising skill for social robots, current and potential future challenges must be considered. For the design and development of a human-robot imitation system we have identified several open challenges. In the following, we look in more detail into these challenges, which include the accuracy of sensing human motion, the correspondence problem of mapping the behavior from one body to another morphologically different body, the characteristics of the imitated motion, the choice of suitable evaluation metrics, and some ethical considerations when imitation systems are used in social interaction settings.

- **Accuracy of Human Motion Sensing**

One of the challenges that arise when dealing with the requirement of sufficiently accurate imitation of human motion is choosing the appropriate method for human motion sensing. Due to the different characteristics of the currently available methods, there will be a trade-off between the availability, comfort handling and cost-efficiency of non-contact sensors and markerless methods (typically based on computer vision methods for human pose estimation) on the one hand, and high accuracy requirements (which are more easily met by motion capture devices) on the other. This compromise requires careful consideration of what is possible and what is necessary (in the case of special conditions) to meet the envisioned imitation goal. In addition, computer vision methods can introduce further challenges such as dealing with ambiguities, for example if there is more than one person in the camera view.

- **Correspondence Problem**

Dealing with the physical differences and constraints of robots is another chal-

lenge in designing and developing a human-robot imitation system. The task of properly mapping the human motion to the robotic platform has been defined as the *correspondence problem* [Nehaniv and Dautenhahn 1998]. A common approach to facilitate the mapping between dissimilar bodies is the use of humanoid robots due to their morphology being similar to that of humans (head, arms, etc.). However, this only partially solves the problem, because humanoid robot joint usually have different degrees-of-freedom than human joints [Yamane and Murai 2016]. The physical morphology differences between humans and robots also creates challenges in different interaction settings. One possible solution would be to allow the interactant to change the type of imitation within the interaction whenever the interactant finds it necessary. This, of course, could change as the interactant gains more experience with the robotic platform, but it would be a useful approach for novice users to try out different imitation types and explore the capabilities of the robot. In addition, this would allow for a better understanding of the robot's imitation capabilities for the human embodying the robot, as well as a reduction in the difficulty of properly mapping the human to robot behaviors for the targeted goal of imitation.

- **Imitated Motion Characteristics**

The following two challenges are identified when it comes to the characteristics of the motion reproduced by the robot, such as motion speed and smoothness. Many humanoid robots often move at a slower speed than humans, usually because of safety measures. Thus, if the human demonstrator moves faster than the robot's maximum speed there would be two possible options for approaching the speed of the imitated motion. The robot would either aim at imitating all poses within the motion sequence resulting in a delayed imitated movement, or skip some poses of the motion sequence to minimize the delay to near real-time imitation. Skipping some poses causes gaps in the imitated movement, implying that the robot will not reach all of the positions within the motion sequence as performed by the human. Second, smooth motion reproduction by minimizing motion jerkiness is still a feature that is being researched. To meet this challenge several methods have been proposed, such as pre-processing the data of the human motion [Luo et al. 2013], or post-processing the converted data to robot motion [Zhu et al. 2017]. Both pre-processing and post-processing the motion data usually includes filters (e.g. Kalman filter) that remove sensor noise and smooth the motion trajectory. However, finding a suitable method to smooth the motion trajectory remains an ongoing research topic.

- **Suitable Evaluation Metrics**

Another open challenge that goes hand in hand with the correspondence problem is how to suitably evaluate an imitation system in terms of the success of the imitation. So far, there are many inconsistencies in the literature regarding the methods used to evaluate human-robot imitation systems, making it difficult

to compare systems to each other. One solution would be to provide a comprehensive set of evaluation metrics that can be applied selectively based on the distinctive features of the system, which would include a combination of quantitative, qualitative, and possibly subjective observational evaluation methods. In this context, it would also be important to define the imitation type and identify the aim of the imitation. The imitation goal can be to focus on the motion itself (e.g., how human-like the motion is) or the accomplishment of a specific task (e.g., the success of grasping an object). This will also determine which distinctive features will be the focus of the evaluation process. The goal is to find a suitable method that measures how successful the imitation is based on the goal of the imitation and the system's key features (e.g. imitation type, time component, etc.).

- **Ethical Considerations in Social Interaction Settings**

When dealing with tracking of human data, it is important that privacy issues are taken into account and that people interacting with the technology are provided with transparent information on how their data is being used. Concerns have also been raised that imitation systems designed for social interactions, such as in the case of interpersonal coordination, may deceive interactants. This deception is described as deceiving interactants into thinking that the robot has more cognitive abilities than it does [Sharkey and Sharkey 2020]. However, the authors argue that not all deceptions are wrong as long as the deception does not cause any negative impact on the person or society in general. This distinction between wrong and not wrong deceptions is a topic of ongoing discussion in the fields of ethics and philosophy. On the other hand, findings in social psychology indicate that interpersonal coordination increases likability and rapport between interactants [Burgoon et al. 1995]. These findings may have an impact on how the way interpersonal coordination is transferred to be used in human-robot interactions. The ability of the robot to exhibit interpersonal coordination could be used to some advantage for the application or the stakeholders selling the robot, which could have a negative impact on the interactant. Thus, an open question from an ethical point of view is: How can we ensure that the imitation system is not used for the wrong deception of the interactants? And is it ethical (because of the possible deception) to allow robots to take part in social interactions and express interpersonal coordination with the interactants?

5 Conclusion

Imitation systems have a wide range of potential applications within the field of human-robot interaction. This paper proposes a method for designing and developing a human-robot imitation system in light of various application scenarios.

The following elementary system components are identified: intended application, technical implementation, and suitable evaluation. Each of these elements, as well as their interrelationships are described and discussed. Based on an examination of several potential applications, the interaction setting with its specific requirements is identified to be a key aspect to consider in system design. The interaction setting can range from having a higher interpersonal component (e.g., imitation for the purpose of interpersonal coordination) to having a higher cooperative component (e.g., imitation targeted for teleoperation) interaction settings. The system requirements that may emerge from the interaction setting have an important influence on decisions for the technical implementation, but also for choosing a suitable evaluation method. The interaction setting is also closely related to the possibility of facilitating different types of trust between the human and the robot. Finally, open challenges in developing human-robot imitation systems are discussed along with possible approaches as a way to tackle them. Further research should aim to better understand in what ways body motion contributes to the overall interaction between a human and a robot, and how it can be tested not only as a stand-alone capability but also in combination with other robotic social capabilities.

Bibliography

- Mina Alibeigi, Sadegh Rabiee, and Majid N. Ahmadabadi. 2017. Inverse kinematics based human mimicking system using skeletal tracking technology. *Journal of Intelligent and Robotic Systems* 85 (2017), 27–45.
- Aris Alissandrakis, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2002. Imitation with ALICE: learning to imitate corresponding actions across dissimilar embodiments. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 32, 4 (2002), 482–496.
- Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57, 5 (2009), 469–483.
- Michael Argyle. 1975. *Bodily Communication*. International Universities Press.
- Sofiane Boucenna, Salvatore Anzalone, Elodie Tilmont, David Cohen, and Mohamed Chetouani. 2014. Learning of social signatures through imitation game between a robot and a human partner. *IEEE Transactions on Autonomous Mental Development* 6, 3 (2014), 213–225.
- Cynthia Breazeal and Brian Scassellati. 2002. Robots that imitate humans. *Trends in Cognitive Sciences* 6, 11 (2002), 481–487.
- Rodney A. Brooks. 1991. Intelligence without representation. *Artificial Intelligence* 47, 1 (1991), 139–159.
- Judee K. Burgoon, Lesa A. Stern, and Leesa Dillman. 1995. *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press.

- Sylvain Calinon and Aude Billard. 2007. Active Teaching in Robot Programming by Demonstration. In *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 702–707.
- Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 1 (2019), 172–186.
- Thierry Chaminade, David W. Franklin, Erhan Oztop, and Gordon Cheng. 2005. Motor interference between Humans and Humanoid Robots: Effect of Biological and Artificial Motion. In *Proceedings of the International Conference on Development and Learning*. 96–101.
- Thierry Chaminade, Erhan Oztop, Gordon Cheng, and Mitsuo Kawato. 2008. From self-observation to imitation: Visuomotor association on a robotic hand. *Brain Research Bulletin* 75, 6 (2008), 775–784.
- Henning Christiansen and Anja Lindelof. 2020. Robots on stage. *EAI Endorsed Transactions on Creative Technologies* 7, 25 (2020), 1–13.
- Mark Coeckelbergh. 2012. Can we trust robots? *Ethics and Information Technology* 14 (2012), 53–60.
- John J. Craig. 2005. *Introduction to Robotics: Mechanics and Control* (3rd ed.). Pearson Education.
- Mark Edmonds, Feng Gao, Hangxin Liu, Xu Xie, Siyuan Qi, Brandon Rothrock, Yixin Zhu, Ying Nian Wu, Hongjing Lu, and Song-Chun Zhu. 2019. A tale of two explanations: Enhancing human trust by explaining robot behavior. *Science Robotics* 4, 37 (2019), eaay4663. doi.10.1126/scirobotics.aay4663
- Sarah Fischer. 2021. *Design and Evaluation of Non-Verbal Cues for the Robot Pepper*. Master's thesis. Technical University of Vienna (TU Wien), Vienna, AT.
- Haruaki Fukuda and Kazuhiro Ueda. 2010. Interaction with a moving object affects one's perception of its animacy. *International Journal of Social Robotics* 2 (2010), 187–193.
- Wei Guo, Jianxin Chen, Ming Zhang, and Zhaolai Pan. 2019. Geometry Based LM of Robot to Imitate Human Motion with Kinect. In *Proceedings of the International Conference on Image, Vision and Computing (ICIVC)*. 695–700.
- Daria Halprin. 2003. *The Expressive Body in Life, Art, and Therapy: Working with Movement, Metaphor and Meaning*. Jessica Kingsley Publishers.
- Markus Häring, Nikolaus Bee, and Elisabeth André. 2011. Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *Proceedings of the International Conference on Robot & Human Interactive Communication (RO-MAN)*. 204–209. 9
- Ryosuke Hasumoto, Kazuhiro Nakadai, and Michita Imai. 2020. Reactive Chameleon: A Method to Mimic Conversation Partner's Body Sway for a Robot. *International Journal of Social Robotics* 12 (2020), 239–258.
- Fritz Heider and Marianne Simmel. 1944. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259.
- Matthias Hirschmanner, Christiana Tsiourti, Timothy Patten, and Markus Vincze. 2019. Virtual Reality Teleoperation of a Humanoid Robot Using Markerless Human Upper Body Pose Imitation. In *Proceedings of the International Conference on Humanoid Robots (Humanoids)*. 259–265.

- Anna Jordanous. 2020. Intelligence without Representation: A Historical Perspective. *Systems* 8, 3 (2020), 1–18.
- Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. 2013. Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation. *IEEE Transactions on Affective Computing* 4, 4 (2013), 341–359.
- Mingon Kim, Sanghyun Kim, and Jaeheung Park. 2016. Human motion imitation for humanoid by recurrent neural network. In *Proceedings of the International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. 519–520.
- Jonas Koenemann, Felix Burget, and Maren Bennewitz. 2014. Real-time imitation of human wholebody motions by humanoids. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*. 2806–2812.
- Nicole C. Krämer, Carina Edinger, and Astrid M. Rosenthal-von der Pütten. 2016. The effects of a robot’s nonverbal behavior on users’ mimicry and evaluation. In *Proceedings of the International Conference on Intelligent Virtual Agents*. 442–446.
- Jangwon Lee. 2017. A survey of robot learning from demonstrations for human-robot collaboration. *arXiv preprint arXiv:1710.08789* (2017).
- Chaoran Liu, Carlos T. Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2012. Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*. 285–292.
- Max M Louwerse, Rick Dale, Ellen G Bard, and Patrick Jeuniaux. 2012. Behavior matching in multimodal communication is synchronized. *Cognitive Science* 36, 8 (2012), 1404–1426.
- Ren C. Luo, Bo-Han Shih, and Tsung-Wei Lin. 2013. Real time human motion imitation of anthropomorphic dual arm robot based on Cartesian impedance control. In *Proceedings of the International Symposium on Robotic and Sensors Environments (ROSE)*. 25–30.
- Kevin M. Lynch and Frank C. Park. 2017. *Modern Robotics*. Cambridge University Press.
- Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. 2017. VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera. *ACM Transactions on Graphics* 36, 4, 1–14.
- Shohin Mukherjee, Deepak Paramkusam, and Santosha K. Dwivedy. 2015. Inverse kinematics of a NAO humanoid robot using kinect to track and imitate human motion. In *Proceedings of the International Conference on Robotics, Automation, Control and Embedded Systems (RACE)*. 1–7.
- Atsushi Nakazawa, Shinichiro Nakaoka, Katsushi Ikeuchi, and Kazuhito Yokoi. 2002. Imitating human dance motions through motion structure analysis. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, Vol. 3. 2539–2544.
- Chrystopher L. Nehaniv and Kerstin Dautenhahn. 1998. Mapping between dissimilar bodies: Affordances and the algebraic foundations of imitation. In *Proceedings of the European Workshop on Learning Robots (EWLR-7)*. 64–72.
- Chrystopher L. Nehaniv and Kerstin Dautenhahn. 2007. *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge University Press.

- Rui Ogawa, Sung Park, and Hiroyuki Umemuro. 2019. How Humans Develop Trust in Communication Robots: A Phased Model Based on Interpersonal Trust. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*. 606–607.
- Yongsheng Ou, Jianbing Hu, Zhiyang Wang, Yiqun Fu, Xinyu Wu, and Xiaoyun Li. 2015. A real-time human imitation system using kinect. *International Journal of Social Robotics* 7 (2015), 587–600.
- Eunil Park, Hwayeon Kong, Hyeong-taek Lim, Jongsik Lee, Sangseok You, and Angel Pasqual del Pobil. 2011. The effect of robot's behavior vs. appearance on communication with humans. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*. 219–220.
- Luigi Penco, Nicola Scianca, Valerio Modugno, Leonardo Lanari, Giuseppe Oriolo, and Serena Ivaldi. 2019. A Multimode Teleoperation Framework for Humanoid Loco-Manipulation: An Application for the iCub Robot. *IEEE Robotics & Automation Magazine* 26, 4 (2019), 73–82.
- Rifca Peters, Joost Broekens, and Mark A. Neerinx. 2017. Robots educate in style: The effect of context and non-verbal behaviour on children's perceptions of warmth and competence. In *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 449–455.
- Hunter Rogers, Amro Khasawneh, Jeffery Bertrand, and Kapil Chalil Madathil. 2017. An investigation of the effect of latency on the operator's trust and performance for manual multi-robot teleoperated tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. 390–394.
- Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, and Jonathan Herrmann. 2018. The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics* 10 (2018), 569–582.
- Hamed Saeidi, John R Wagner, and Yue Wang. 2017. A mixed-initiative haptic teleoperation strategy for mobile robotic systems based on bidirectional computational trust analysis. *IEEE Transactions on Robotics* 33, 6 (2017), 1500–1507.
- Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5 (2013), 313–323. <https://doi.org/10.1007/s12369-013-0196-9>
- Wataru Sato and Sakiko Yoshikawa. 2007. Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1 (2007), 1–18.
- Stefan Schaal. 1999. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences* 3, 6 (1999), 233–242.
- Richard C. Schmidt, Samantha Morr, Paula Fitzpatrick, and Michael J. Richardson. 2012. Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior* 36 (2012), 263–279.
- Amanda Sharkey and Noel Sharkey. 2020. We need to talk about deception in social robotics! *Ethics and Information Technology* (2020), 1–8.
- Michihiro Shimada, Kazunori Yamauchi, Takashi Minato, Hiroshi Ishiguro, and Shoji Itakura. 2008. Studying the Influence of the Chameleon Effect on Humans using an Android. In *Proceedings of the International Conference on Intelligent Robots and Systems*. 767–772.

- Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. 2011. Real-time human pose recognition in parts from single depth images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. 1297–1304.
- Christopher Stanton, Anton Bogdanovych, and Edward Ratanasena. 2012. Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning. In *Proceedings of the Australasian Conference on Robotics and Automation (ACRA)*. 1–10.
- Darja Stoeva, Helena A. Frijns, Margrit Gelautz, and Oliver Schürer. 2021. Analytical Solution of Pepper’s Inverse Kinematics for a Pose Matching Imitation System. In *Proceedings of the International Conference on Robot & Human Interactive Communication (RO-MAN)*. 167–174.
- Nguyen T. V. Tuyen, Sungmoon Jeong, and Nak Y. Chong. 2018. Emotional Bodily Expressions for Culturally Competent Robots through Long Term Human-Robot Interaction. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*. 2008–2013.
- Linda N. Vallée, Sao M. Nguyen, Christophe Lohr, Ioannis Kanellos, and Olivier Asseu. 2020. How An Automated Gesture Imitation Game Can Improve Social Interactions With Teenagers With ASD. In *ICRA workshop on Social Robotics for Neurodevelopmental Disorders*.
- Katsu Yamane and Akihiko Murai. 2016. A Comparative Study Between Humans and Humanoid Robots. In *Humanoid Robotics: A Reference*, Ambarish Goswami and Prahlad Vadakkepat (Eds.). 1–20. https://doi.org/10.1007/978-94-007-6046-2_7
- Unai Zabala, Igor Rodriguez, José M. Martínez-Otzeta, and Elena Lazkano. 2020. Can a Social Robot Learn to Gesticulate Just by Observing Humans? In *Advances in Physical Agents II*, Luis M. Bergasa, Manuel Ocaña, Rafael Barea, Elena López-Guillén, and Pedro Revenga (Eds.). 137–150. https://doi.org/10.1007/978-3-030-62579-5_10
- Liang Zhang, Zhihao Cheng, Yixin Gan, Guangming Zhu, Peiyi Shen, and Juan Song. 2016. Fast human whole body motion imitation algorithm for humanoid robots. In *Proceedings of the International Conference on Robotics and Biomimetics (ROBIO)*. 1430–1435.
- Ming Zhang, Jianxin Chen, Xin Wei, and Dezhou Zhang. 2018. Work chain-based inverse kinematics of robot to imitate human motion with Kinect. *ETRI Journal* 40, 4 (2018), 511–521.
- Tehao Zhu, Qunfei Zhao, Weibing Wan, and Zeyang Xia. 2017. Robust regression-based motion perception for online imitation on humanoid robot. *International Journal of Social Robotics* 9 (2017), 705–725.
- Fernando Zuher and Roseli Romero. 2012. Recognition of human motions for imitation and control of a humanoid robot. In *Proceedings of the Brazilian Robotics Symposium and Latin American Robotics Symposium (SBR-LARS)*. 190–195.