

Künstl Intell  
DOI 10.1007/s13218-010-0064-9

FACHBEITRAG

# Approaching Manual Intelligence

Jonathan Maycock · Daniel Dornbusch ·  
Christof Elbrechter · Robert Haschke ·  
Thomas Schack · Helge Ritter

Received: 22 April 2010 / Accepted: 28 July 2010

© The Author(s) 2010. This article is published with open access at [Springerlink.com](http://Springerlink.com)

**Abstract** Grasping and manual interaction for robots so far has largely been approached with an emphasis on physics and control aspects. Given the richness of human manual interaction, we argue for the consideration of the wider field of “manual intelligence” as a perspective for manual action research that brings the cognitive nature of human manual skills to the foreground. We briefly sketch part of a research agenda along these lines, argue for the creation of a manual interaction database as an important cornerstone of such an agenda, and describe the manual interaction lab recently set up at CITEC to realize this goal and to connect the efforts of robotics and cognitive science researchers towards making progress for a more integrated understanding of manual intelligence.

## 1 From Robots to Manual Intelligence

Progress in mechatronics, sensing and control has made sophisticated robot hands possible whose potential for dexterous operation is at least beginning to approach the superb performance of human hands [1–3]. The increasing availability of these hands, together with sophisticated, physics-based simulation software, has spurred a revival of the field of anthropomorphic hand control in robotics, whose ultimate goal is to replicate the abilities of human hands to handle everyday objects in flexible ways and in unprepared environments.

---

The authors are cooperating within the Bielefeld Excellence Cluster Cognitive Interaction Technology (CITEC) and the Bielefeld Institute for Cognition and Robotics (CoR-Lab).

---

J. Maycock · D. Dornbusch · C. Elbrechter · R. Haschke ·  
T. Schack · H. Ritter (✉)  
Bielefeld University, Bielefeld, Germany  
e-mail: [helge@techfak.uni-bielefeld.de](mailto:helge@techfak.uni-bielefeld.de)

A major focus of these works is a deeper understanding of manual interaction at the level of geometry, contact and force physics [4, 5]. However, similar to language, whose essence it not well captured by analyzing the interaction with sound pressure waves in the air, a deeper understanding of manual interaction may require us to go significantly beyond the analysis of control issues found at the level of geometry and physical contacts alone.

The richness of manual interactions involves numerous higher levels, including object recognition, exploration and shaping of articulated or deformable items, performing complex assembly tasks, tool use, manual gesture, and ranges even into artistic and emotional expression. Indeed if the need arises, its scope can even stretch to encompass the capacity of full linguistic expression.

This motivates us to conceive a research field of *manual intelligence* as a more modest, but still rich concretization and focusing field of the more elusive topic of intelligence and cognition in general. And one may hope that a deep understanding of manual intelligence will constitute a major step towards our understanding of general cognition.

## 2 A Sketch of a Research Agenda

While at the physics level almost all laws that are involved in manual interaction situations are known, this knowledge alone is only sufficient to derive meaningful grasping and interaction strategies in highly simplified and constrained situations. Taking again the analogy with language: the physics of the vocal tract and its interaction with the surrounding air permits the implementation of a huge variety of languages. However, the commonalities of their structure

cannot be derived from the physics alone, but requires a deeper understanding of linguistic phenomena that become observable and meaningful only at higher levels of abstraction.

A research agenda for studying manual intelligence therefore has to observe and analyze manual actions at a hierarchy of levels and analyze this data in order to arrive at a comprehensive picture of the cognition enabling manual interaction and manual intelligence.

An important entry point is the construction of a comprehensive database of manual interaction patterns in a variety of situations. While today linguistic databases [6] exist in a great variety, the construction of databases for manual interaction patterns is still largely in an infancy stage. In Sect. 3 we report on our current efforts towards developing a comprehensive and versatile database as a major cornerstone for manual interaction research.

The construction of such a database is intimately connected with the creation of sophisticated motion capturing facilities to observe manual interactions at a high level of spatio-temporal resolution. This involves numerous technical challenges with regard to data acquisition, integration of different input channels as well as the calibration and mutual registration of the involved modalities. Many of the associated questions turn out to be inseparable from key research questions connected with the observation and identification of highly articulated movements in the presence of occlusion and noise. Section 4 provides a concise overview of the issues involved and of our solutions developed for the realization of a flexible and performant manual interaction capture lab environment.

Motion capture data is represented at a very low level of abstraction. This level is very unlikely to be related in any simple and straightforward fashion with the cognitive representations utilized by the brain to control sensorimotor manual action; likewise, this level will only be of very limited use for shaping the sensorimotor interaction repertoire of a robot. One of the major challenges is to arrive at a systematic hierarchy of interconnected representations that are suitable to capture a sizeable body of manual interaction knowledge at different levels of abstraction. In Sect. 5 we sketch some of our current efforts towards such representations, combining computational methods from machine learning and robotics with approaches from cognitive science to infer the structure of cognitive representations of manual action in humans.

Finally, the resulting insights and computational hypotheses have to be explored and validated on actual robot platforms in order to judge their reach for real world situations. This aspect is taken up in Sect. 6, where the opening of a container is discussed as a basic manual interaction setting. Section 7 concludes with a discussion and some perspectives on future research.

### 3 A Manual Interaction Database

Linguistic databases, such as WordNet [6], have become an important tool for language research and for gaining insights into the structure of our mental concepts. This suggests the development of a similar research tool for manual action could perform the same kind of function.

Unlike language, much of whose essence can rather easily be directly encoded in symbol strings, the required representations to capture the essential parts of manual interaction situations are less obvious to delineate. However, there is little doubt that an important core part will be formed by motion data of the involved hand(s). There are some freely available motion capture databases such as the CMU Graphics Lab Motion Capture Database [7], the Karlsruhe Human Motion Library [8] or the recent TUM Kitchen Data Set [9] focusing on different aspects and settings of human motion data. However, as their focus is on full body movements, they only allocate a very small number of degrees of freedom to the hands, enabling only a very coarse representation of manual actions. Thus, to date there does not exist a database representing a larger variety of manual actions in greater detail.

Besides requiring the allocation of a larger number of degrees of freedom to the representation of the human hand, a manual interaction database should also include recordings of the position and orientation of involved objects, and can benefit from additional modalities recording interaction sounds, eye gaze data or information about tactile forces at the finger tips. This multimodality makes the design of a manual interaction database a highly non-trivial task that is intimately coupled with the design of suitable lab environment to actually capture the envisaged data. The Manual Intelligence Lab (MILAB) (see Fig. 1) currently supports a set of eight devices and these are listed, along with their specifications, in Table 1. This allows us to observe manual interactions at a high level of spatio-temporal resolution.



**Fig. 1** Manual Intelligence Lab (MILAB)

**Table 1** Overview of currently supported devices

Modality	Number	Frequency	Sensors/Pixels	Bits	Interface	Accuracy
Vicon MX3+ Cameras	14	240 Hz	659 × 493 px	8	Vicon-Special	< 1 mm
Hi-res Basler Cameras	2	35 Hz	1600 × 1200 px	8	GigE	–
Highspeed Basler Cameras	2	210 Hz	640 × 480 px	8	GigE	–
Swiss Ranger 4000	1	54 Hz	176 × 144 px	16	USB	10 mm
Cyber Glove II	2	80 Hz	22	10	Bluetooth	< 1°
Tactile Sensors	2	80 Hz	5 or 9	10	Bluetooth	–
Microphones	2	48 kHz	–	16	IEEE 1394	–
Eye Tracker	1	200 Hz	376 × 240 px (scene video)	8	USB	< 0.5°–1°

Fourteen *Vicon*<sup>1</sup> MX3+ cameras track reflective markers attached to both subjects and objects. Having so many cameras in such a small space allows us to track the human hand, which due to self occlusions and occlusions caused by objects is a difficult task. Vicon has a marker tracking accuracy of below 1 mm and this provides us with good ground truth data for our other vision modalities. It also provides us with 6D model information for non-deformable objects and higher dimensional models for complicated structures such as the human hand. Another important data input stream we utilize are the *stereo-vision cameras*. The captured image/video sequences provide an intuitive description of what is going on in the scene. Video data is indispensable for manual annotation and navigation within the time line of a trial. Furthermore, our stereo-camera vision setup is close to the camera setup of common robotics systems. We use different versions of Basler-pilot cameras<sup>2</sup> (high resolution or high speed), as these can be directly triggered and calibrated by the Vicon system. Two cameras can be used to compute 3D scene information using stereo-vision algorithms. For more robust 3D information, we added a *Swiss Ranger 4000* time-of-flight camera.<sup>3</sup> In certain circumstances where the Vicon system's output suffers from too many occlusions, the human hand's joint-angles can be captured using the *Cyber Glove II*.<sup>4</sup> By this means, an alternative input channel for creation and tracking of a hand model is provided. Furthermore, we attached a set of *tactile sensors* (5 for the finger tips and another 4 that can optionally be placed somewhere in the palm), to the hand. These provide information about contact with objects and the forces required to lift and move objects. In order to get insights into where humans look as they perform everyday tasks with their hands, we use a SMI IViewX (monocular) mobile *eye tracking system*<sup>5</sup> for our ex-

periments. The eye tracker captures a low-resolution scene video as well as eye fixation points and saccade events. Finally, we capture sound information using table top *microphones*, which initially we are using as an aid for the temporal segmentation of trials, but in future we envision robots using sound in a similar way that humans and animals do to enhance awareness of their surroundings.

A central part of the overall challenge is the development of a framework that allows us to connect the captured data with representations at higher levels of abstraction, such as the task level of an operation. Taking the relatively simple task of placing a lid back on a jar, what are the common building blocks of such a task? What is the variability in terms of the trajectory of hand and arm movements among different subjects and indeed across trials carried out by the same subject? Questions such as these are important if we are to build stable models of manual interactions that can be ported to robotic systems. Event information such as when and where contact occurs during a trial is also important to be able to imitate what has taken place. With this in mind our database has been designed to not only store raw data from various devices, but to also hold higher level information such as segmentation points, description labels of what is occurring and force contact information.

Work in MILAB has already begun in earnest. We recently recognized sequences in multidimensional time-series by first learning a smooth quantization of the raw data, and then using a variant of dynamic time warping to recognize short sequences of prototypical motions in a long unknown sequence. Short manual actions were successfully recognized and the approach was shown to be spatially invariant [10].

#### 4 The Challenges of Manual Action Capture

The challenges of manual action capture are many and varied, but in most cases reduce to the problem of ensuring that there is a high degree of both spatial and temporal coherence

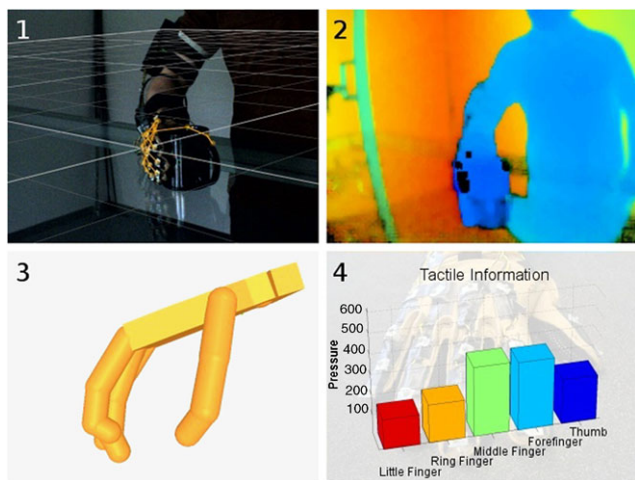
<sup>1</sup>Vicon motion capture system. <http://www.vicon.com>.

<sup>2</sup>Basler. <http://www.baslerweb.com>.

<sup>3</sup>Mesa Imaging. <http://www.mesa-imaging.ch>.

<sup>4</sup>CyberGlove Systems. <http://www.cyberglovesystems.com>.

<sup>5</sup>Sensomotoric instruments (smi). <http://www.smivision.com>.



**Fig. 2** Multimodal data captured in MILAB: (1) Overlay of Vicon hand model and Basler camera image, (2) Pseudo-colour visualization of depth image provided by Swiss Ranger Camera, (3) Vortex visualization of Cyber Glove II data, and (4) Bar plot of touch sensor information

amongst the different input channels (see Fig. 2 for a snapshot of a trial captured in MILAB). At the core of MILAB is a 14 camera Vicon system providing us with high precision 3D positional data to which the other vision modalities need to be aligned. For modalities not directly supported by Vicon, i.e., the Swiss Ranger time-of-flight camera, it was necessary to develop custom made solutions in order to ensure spatial coherence. Equally important is the need to ensure temporal coherence across all modalities and for this dedicated software was developed for MILAB. Once coherence issues are solved, an interface allowing access to the captured trials is needed. We are developing an intuitive interface in which the captured data can be queried and indeed modified with annotations that add a higher level of abstraction to the trials.

#### 4.1 Calibration

Accurate calibration is essential if we are to have a high degree of spatial coherence. The Vicon system allows for high-precision camera calibration through its control application Nexus.<sup>6</sup> Our choice of using Basler-cameras was motivated by the fact that their calibration is directly supported by Nexus. However, this is not the case for the Swiss Ranger time-of-flight camera and therefore we implemented a custom calibration procedure for it. Along with the depth image, the Swiss Ranger camera provides a gray-scale intensity image that can be used for common camera calibration. As a first step we compensate for the lens distortion using the Image Component Library's [11] camera undistortion

feature, which uses the simple undistortion model of the ARToolkit [12]. Even though it has performed well in initial tests, we plan to implement a more general undistortion model going forward [13]. As the Swiss Ranger image is small, we can afford to undistort the image pixel-wise in order to work with undistorted images in the proceeding steps.

The second step involves using the direct linear transform algorithm [14], also implemented in Image Component Library, in order to obtain the remaining intrinsic and extrinsic camera parameters. The extrinsic parameters describe the 6D-pose of the camera with respect to the calibration object we use. Vicon markers were also attached to the calibration object, which enabled us to get the calibration object's 6D-pose with respect to the global Vicon coordinate system. Combining these two 6D-poses allows us to transform the 3D point cloud obtained by the Swiss Ranger camera into a point cloud in the global Vicon coordinate system.

#### 4.2 Synchronization

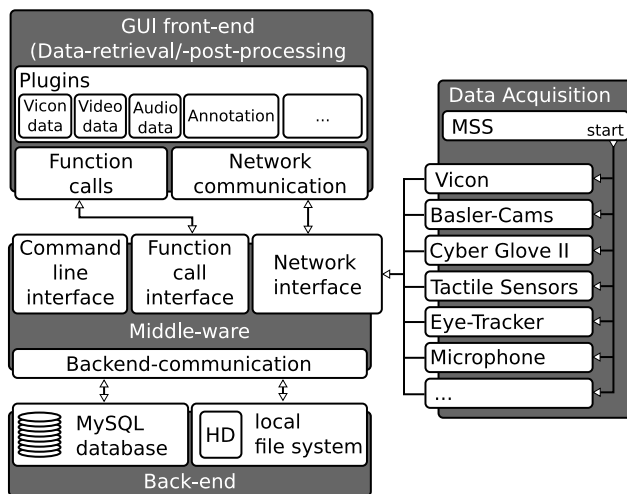
Synchronization is crucial if data is to be captured using several modalities over a distributed computer network. All devices must be managed so that they start and stop recording at the same time. Also, they should be able to grab data at defined points of time. This is important when, for example, the disparity from the stereo Basler cameras is calculated to estimate depth. If the images are not captured at the same time, objects in the field of view may have moved, which results in increased errors. Therefore, all cameras (Vicon, Basler, and Swiss Ranger) are triggered by the Vicon hardware using multiples of the slowest device's frequency. For example, the Swiss Ranger is triggered with 50 Hz and the others with 200 Hz.

We developed software called Multiple Start Synchronizer (MSS) to ensure that the recordings of all data streams start at the same time, and run for a specific duration or until a stop command is sent. MSS first checks if all computer clocks are synchronized correctly using the Network Time Protocol. Then, the user can enter all necessary pieces of information for the trial: experiment, subject and trial names, delay time after start button is pressed, duration, end time, computer names/IP addresses, and ports. Pressing start sends all relevant data via Open Sound Control protocol<sup>7</sup> to the listening clients. The client applications then wait for the starting time and begin capturing until the end time is reached or a stop signal from MSS is received. The Vicon system is controlled directly by MSS, which uses a telnet connection to the Vicon MX Control hardware to reset the hardware timer. An MSS remote control feature is used to

<sup>6</sup>Vicon motion capture system. <http://www.vicon.com>.

<sup>7</sup>Open sound control. <http://opensoundcontrol.org>.





**Fig. 3** Software layers and communication

control Vicon's Nexus software so that automatic capturing can be set up. Alternatively, an Arduino board<sup>8</sup> could be used to trigger capturing with a TTL signal.

After a trial is finished, all clients copy the locally stored data into the database or to a common hard disk using an interface provided by a software framework that is described in the next section. MSS is very flexible and can be easily extended to support additional modalities.

#### 4.3 Architecture

We are in the process of developing a software framework (see Fig. 3) in order to provide an intuitive interface for the experiments carried out in MILAB. A MySQL database is used to store most of the recorded data. To reduce the load on the database, large data blobs such as captured images and video files are stored directly on a local file system. The communication between the GUI front-end and the back-end is handled by an intermediate-layer that provides different interfaces for the storing and querying of data.

As different processes have to be able to access the data simultaneously (for writing and/or reading), a middle-ware process is instantiated for synchronizing database- and file-IO. In a single-user situation, the GUI-front-end can use a function call interface for faster communication with the data storage units. The GUI based front-end is inspired by common video editing software. It is used for visualization, querying, and exporting of data. Furthermore, it provides interfaces for data post-processing, e.g., manual setting of synchronization points or removal of outliers. We have also added a physics engine visualization plug-in to the GUI.

<sup>8</sup>Arduino. <http://www.arduino.cc>.

Currently we use Vortex,<sup>9</sup> but the meta data is xml-based and can be easily adapted to work with other engines. We are also in the process of developing a graphical annotation plug-in so that trials can be annotated with information such as segmentation points, force contact intervals and labels describing different aspects of trials.

## 5 Towards Cognitive and Computational Representations

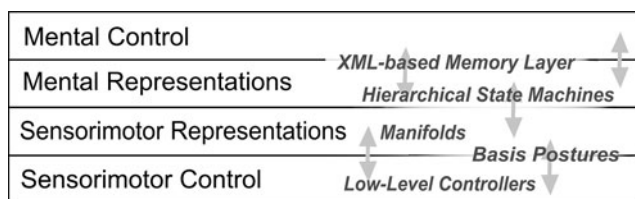
While there is a significant body of literature focusing on selected aspects of representations for manual actions, such as contact formation, grasp optimization and finger gaits (for an overview see [4, 5]), there is, thus far, relatively little known about the integration of these isolated aspects into an overarching and integrated architecture. A promising path towards the elucidation of such an architecture is the marrying of recent methods and concepts from cognitive psychology for studying mental representations of action, with ideas from cognitive robotics about how to integrate a rich set of skills in a systematic and manageable fashion [15].

In this way, we hope to arrive at an overall system that can organize different facets of manual interaction knowledge in a way that mimics cognitive representations in the brain and that allows the synthesis of complex manual operations in robots.

One of the first abstraction steps of captured motion data are *basis hand postures* for various grasping actions. Using a small set of prototypical hand postures as pre-grasps, we have developed a robust method for grasping a wide range of household objects [16]. This approach offers interesting cross connections with the concept of *basic action units (BACs)* proposed by one of the authors (TS) for the analysis of human motion [17]. These take the role of basic building blocks from which more complex motor skills can be formed. Moreover, they can be arranged into hierarchies, whose structure can be extracted with specialized interview techniques, allowing us to study how these representations change during learning. The facilities in MILAB will allow us to refine our current notions of basic action units, to connect computational concepts and the ideas from cognitive psychology in a more stringent way and devise novel experiments along with improved algorithms for object grasping.

A next abstraction step occurs when passing from prototypical postures to families of structurally related trajectories. A suitable computational concept that captures such structures are *configuration manifolds*. Using suitably refined variants of machine learning techniques, we have been

<sup>9</sup>CM-Labs Vortex 2.1. [www.cm-labs.com/products/vortex](http://www.cm-labs.com/products/vortex).



**Fig. 4** Tentative Architecture: left part indicates cognitively motivated levels. Right part indicates possible implementations in a technical system

able to extract such manifolds from noisy data sets of motion capture data in the context of medium-complex manipulation skills, such as the unscrewing of a bottle [18].

Such manifolds constitute a major part of the non-symbolic, continuous “control” knowledge within a particular manual interaction pattern or higher-level basic action unit. Most manual interaction skills require the organization of several such units into a network that is traversed in context-specific fashion (e.g., triggering compensatory actions in response to disturbances, or repeating a sub action until an intermediate goal has been reached). As a computational representation, we are using *hierarchical state machines* [19] to represent such networks. Currently, these networks are largely hand-crafted. An attractive perspective is to refine and ultimately synthesize such networks by utilizing MILAB, together with suitable machine learning approaches and structural background knowledge about the underlying cognitive representations.

These lines of development have exposed interesting parallels to a *cognitively motivated architecture of motor action* [15, 17]. This architecture postulates four levels: a sensorimotor level providing an interface to sensors and effectors, two intermediate levels of sensorimotor and mental representations, accommodating basic action units at different levels of abstraction and a topmost level of mental control shaping our purposeful behavior.

This overall picture is our rationale for establishing a comprehensive research lab environment for studying manual interactions in order to explore correspondences between the posture and manifold representations and their coordinating state machines in the computational manual action architecture on the one side, and the sensorimotor and mental representations levels of the cognitive model on the other side (see Fig. 4), and to bring to bear methods from both disciplines to mutually refine and crossconnect the two accounts towards a coherent and overarching picture of manual intelligence.

## 6 From Capture to Synthesis

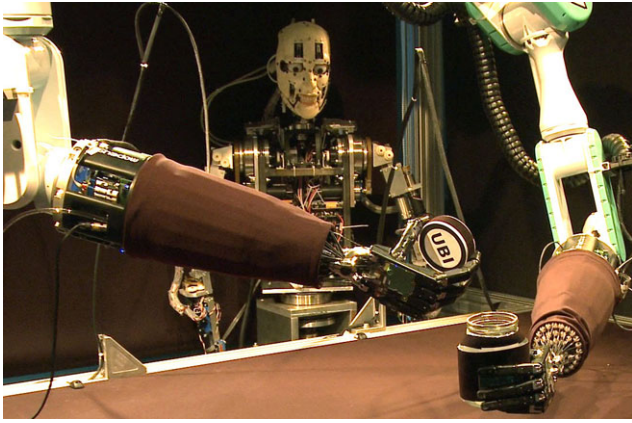
While being able to record, analyze and represent manual interaction patterns in a way that integrates cognitive and

computational aspects is no small feat, any deeper understanding of manual intelligence has to prove itself ultimately in the ability to *synthesize* complex manual actions within a considerable range of situations.

This part of the research agenda can only be realized with the aid of sufficiently advanced robot systems [20, 21]. Although current systems are far from offering an even remotely faithful approximation of what human hands are able to sense and to do (and, thereby, tend to make most tasks significantly harder), they can provide an already useful testing ground for an interesting variety of manual actions.

As an example of such a “developmental approach with the robot in the loop”, we briefly report on the task of unscrewing the cap of a jar. Early predecessors of this work focused on the process of grasping, using first simulations which then were implemented on two robot hand systems involving the TUM Hand and the Shadow Hand with 20 DOF [16]. This work could be seen as a (very coarse) sketch of parts of the first two levels in the cognitively motivated architecture of motor action discussed in the previous section. Additional flexibility was gained by adding the hierarchical state machine layer (loosely amounting to erecting the initial elements of a more abstract “mental representation” layer) and a (here not discussed, but see [22]) XML memory layer for structuring the overall system behavior at a very high level (which might be comparable to the “mental control” layer of the cognitive architecture). Extending the posture representation with a manifold-based representation and integrating machine learning methods for extracting configuration manifolds from motion capture data utilizing a dataglove, we were able to realize on a bi-manual system of anthropomorphic Shadow Hands mounted on a pair of PA-10 arms (totalling to 54 DOF; see Fig. 5) performing a medium-level skill based on captured data: the unscrewing of a jar passed to the robot by a human [23].

This medium-level skill is already integrating a considerable number of representations ranging from the visual perceptual front end to the low-level posture control of the hand-arm system, the manifold representation of the unscrewing operation, the state machine for the “basic action network” as well as a high level “mental control” layer diagnosing faults (such as when the target object is occluded or not visible) and triggering suitable speech output to inform the human partner. The ability to “play” the interacting of all these representations on a real robot system provides important insights into what contributes to robustness and what does not. It directs attention to aspects that may not yet be prominent in the conceptual picture, but become decisive in the real world, and allows us to explore generalizability and scalability to changed situations.



**Fig. 5** Autonomous opening of the lid of a jar by a bi-manual anthropomorphic robot hand system

## 7 Conclusions

Connecting current research in robotics and cognitive science on the control of manual actions exhibits mutually complementing ideas about the role of basic action units and their embedding into an overarching computational-cognitive architecture for synthesizing complex manual actions. To pursue this further will require us to complement the current, strongly control- and physics-based approach for the synthesis of robot manual actions with an observation-driven approach, combining modern capture technology with advanced analysis methods for enabling rich, multimodal recordings of human manual actions and to refine these into highly organized, multi-level representations of human manual actions. A database along these lines would be an important step towards mapping the large interaction knowledge underlying and enabling the “manual intelligence” exhibited in human manual actions and would constitute a valuable basis for shaping robot manual actions more closely according to our own abilities. We have illustrated some initial steps along such a path, touching on major aspects and some examples, together with perspectives for future research.

**Acknowledgements** This research was supported by the DFG CoE 277: Cognitive Interaction Technology (CITEC).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Shadow Robot Company, the shadow dexterous hand. <http://www.shadowrobot.com>
- Butterfass J, Fischer M, Grebenstein M, Haidacher S, Hirzinger G (2004) Design and experiences with the DLR Hand II. In: Proc world automation congress, pp 2464–2470
- Mouri T, Kawasaki H, Yoshikawa K, Takai J, Ito S (2002) Anthropomorphic robot hand: gifu hand III. In: Proc of int conf ICCAS, pp 1288–1293
- Bicchi A, Kumar V (2000) Robotic grasping and contact: a review. In: Int conf on robotics and automation. IEEE Press, New York, pp 348–353
- Okamura AM, Smaby N, Cutkosky MR (2000) An overview of dexterous manipulation. In: Proc int conf on intelligent robots and systems, vol 1. IEEE Press, New York, pp 255–262
- Fellbaum C (ed) (1998) WordNet: an electronic lexical database. MIT Press, Cambridge
- CMU Graphics Lab Motion Capture Database, note = <http://mocap.cs.cmu.edu>
- Azad P, Asfour T, Dillmann R (2007) Toward a unified representation for imitation of human motion on humanoids. In: Int conf on robotics and automation. IEEE Press, New York, pp 2558–2563
- Tenorth M, Bandouch J, Beetz M (2009) The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition. In: Workshop on tracking humans for the evaluation of their motion in image sequences (ICCV)
- Martin M, Maycock J, Schmidt FP, Kramer O (2010) Recognition of manual actions using vector quantization and dynamic time warping. In: 5th int conf on hybrid artificial intelligence systems (accepted)
- Elbrechter C, Götting M, Haschke R (2010) Image component library (ICL), Jan 2010. <http://iclv.org>
- Kato H, Billinghurst M (1999) Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proc of the 2nd int workshop on augmented reality. IEEE Press and ACM, New York, pp 85–94
- Zhang Z (1996) On the epipolar geometry between two images with lens distortion. In: Int conf on pattern recognition, vol 13, pp 407–411
- Wöhler C (2009) 3D computer vision: efficient methods and applications. Springer, Berlin
- Schack T, Ritter H (2009) The cognitive nature of action—functional links between cognitive psychology, movement science, and robotics. *Prog Brain Res* 174:231–250
- Röthling F, Steil JJ, Ritter H (2007) Platform portable anthropomorphic grasping with the Bielefeld 20-dof shadow and 9-dof tum hand. In: Proc int conf on intelligent robots and systems. IEEE Press, New York, pp 2951–2956
- Schack T, Mechsner F (2006) Representation of motor skills in human long-term memory. *Neurosci Lett* 391(3):77–81
- Steffen J, Haschke R, Ritter H (2008) Towards dexterous manipulation using manipulation manifolds. In: Proc int conf on intelligent robots and systems. IEEE Press, New York, pp 2738–2743
- Yannakakis M (2000) Hierarchical state machines. In: Proc IFIP TCS, vol 1872. Springer, Berlin, pp 315–330
- Ott C, Eiberger O, Friedl W, Bauml B, Hillenbrand U, Borst C, Albu-Schaffer A, Brunner B, Hirschmuller H, Kielhofer S, Konietzke R, Suppa M, Wimbock T, Zacharias F, Hirzinger G (2006) A humanoid two-arm system for dexterous manipulation. In: Conf on humanoid robots, pp 276–283
- Asfour T, Regenstein K, Azad P, Schröder J, Vahrenkamp N, Dillmann R (2006) ARMAR-III: an integrated humanoid platform for sensory-motor control. In: Int conf on humanoid robots (humanoids). IEEE Press/RAS, New York, pp 169–175
- Ritter H, Haschke R, Steil JJ (2007) A dual interaction perspective for robot cognition: grasping as a “rosetta stone”. In: Perspectives of neural-symbolic integration, pp 159–178
- Steffen J, Elbrechter C, Haschke R, Ritter H (2010) Bimanual opening of a screw cap glass in unconstrained settings. In: Conf on int robots and systems. IEEE Press/RSJ, New York (submitted)

**Jonathan Maycock** studied for his Ph.D. in Digital Holography at the National University of Ireland, Maynooth, and joined the CITEC team in 2008.

**Daniel Dornbusch** is Ph.D. student of Computer Science at CoR-Lab and at the Neuroinformatics group.

**Christof Elbrechter** is Ph.D. student of Computer Science at CoR-Lab and at the Neuroinformatics group.

**Robert Haschke** received his Ph.D. in Computer Science at the Faculty of Technology at Bielefeld University in 2004 and is leading the robotics team at the Bielefeld Faculty of Technology Neuroinformatics group headed by Helge Ritter.

**Thomas Schack** is Head of the Neurocognition and Action group at the Department of Sports Sciences, Bielefeld University.