A Gesture Based System for Context - Sensitive Interaction with Smart Homes

Robert Neßelrath, Chensheng Lu, Christian H. Schulz, Jochen Frey, Jan Alexandersson

Abstract This paper introduces a system for gesture based interaction with smart environments. The framework we present connects gesture recognition results with control commands for appliances in a smart home that are accessed through a middleware based on the ISO 24752 standard URC (Universal Remote Console). Gesture recognition is realized by applying three dimensional acceleration sensor information of the WiiMote from Nintendo. This information is trained to a toolkit for gesture recognition that implements machine learning algorithms well known from speech recognition. Our study focuses on two interaction concepts with the aim to exploit the context and special home scenarios. This serves to reduce the number of gestures while in parallel retaining the control complexity on a high level. A user test, also with older persons, compares both concepts and evaluates their efficiency by observing the response times and the subjective impressions of the test persons.

1 Introduction

Life expectation in Germany is rising by three months every year and researchers predict that in 60 years most people will be living to 100 years and longer. Con-

Robert Neßelrath
German Research Center for Artificial Intelligence, Saarbrücken e-mail: robert.nesselrath@dfki.de

Chensheng Lu
German Research Center for Artificial Intelligence, Saarbrücken e-mail: chensheng.lu@dfki.de

Jan Alexandersson
German Research Center for Artificial Intelligence, Saarbrücken e-mail: janal@dfki.de

Christian H. Schulz
German Research Center for Artificial Intelligence, Saarbrücken e-mail: chschulz@dfki.de

Jochen Frey
German Research Center for Artificial Intelligence, Saarbrücken e-mail: jochen.frey@dfki.de

versely the number of younger people is dramatically dropping compared to the increase of the older population. This demographic shift leads to an imbalance between elder and young people. This and the circumstance that older people need more support in their everyday life, invoked by physical and cognitive impairments, lead to the problem that not enough caregivers will be available to support the older generation. One solution to face this problem of the near future are age appropriated assistance systems for smart environments. A smart home that supports its aging residents in everyday life activities raises their life quality on the one hand. On the other hand, it extends the time elder people can live an autonomous life in their own four walls without being dependent on human support.

One important task is to design new interaction concepts based on different modalities like speech or gestures. We introduce a system for gesture based interaction with smart environments. The framework we present connects dynamic hand gestures with control commands for appliances in a smart home that are accessed by a middleware based on the ISO standard URC. The URC technology provides an approach called "pluggable user" interfaces allows for interfacing arbitrary networked appliances or services with personalized, accessible and adaptable user interfaces [1]. One challenge of gesture based interaction is to keep the gesture space as small and easy to remember as possible. Especially the elderly or people with cognitive impairments are overwhelmed by the task to remember a higher set of - in the worst case - complex gestures. This paper focuses on this problem by presenting two interaction concepts with the aim to exploit the context and special home scenarios in order to reduce the number of gestures while retaining the control complexity on a high level in parallel. For this, two gesture interaction concepts have been implemented and tested with a heterogeneous user-group focused on the elderly. The first one takes the interaction context into account by interpreting the performed gesture with the appliance in context as additional information. Thus we can deduce that one and the same gesture can trigger different events on different appliances although in the understanding of a person it has the same interpretation: E.g. lifting the hand stands for increasing a setting. However, on TV it has the effect to increase the volume and on the extraction hood, it increases the fan speed. The second concept combines multiple smart home functionalities of different appliances to several scenarios. Thus a complete room or even the complete house can be adapted to a special state that is achieved by performing only one gesture, e.g. we use a phone scenario that mutes the television or radio but also loud appliances such as the extraction hood.

The next section gives an overview on related work about gesture controls in smart environments and identifies some findings that influenced our study. After that we introduce our smart home infrastructure and our technology for gesture recognition. Subsequently the two examined interaction concepts are explained in detail. The final user test section deals with the test scenarios, test results and their analysis.

2 Related Work

Diverse researchers develop smart home environments with a special focus on personalized user interfaces for multimodal interaction. Burzagli discusses the challenges of designing multimodal interfaces in ambient intelligence scenarios giving main attention to speech interaction [2]. Machate focuses on multimodal interaction concepts for the elderly [3]. Similarly, Song [4] developed smart environments for the elderly and also disabled persons. In this approach voice and hand gestures are the modalities to control wheelchairs, transferring and bed robots. He installed cameras to recognize simple hand gesture patterns. In [5], hand movement paths recorded by cameras and image processing are applied to detect gestures in order to control a lamp and a movie player. For each functionality one gesture is used, four for the lamp (on, off, dim lighter, dim darker) and two for the movie player (play, pause). It is observed that users need some adjustment period with the effort to increase detection precision from 32% to 97%. Blumendorf [6] also addresses the fact that interaction commands for smart home environment must be learnable both easily and quickly. In user tests he found out that users are willing to learn easy commands for both speech and gesture input. Moreover the user appreciates multimodality and prefers short commands over more complex sentences in speech interaction. Furthermore, help for more complex commands is needed and users like to chain multiple commands to speed up interaction. In [7] they record gestures with the accelerometers of an iPhone and use them to control household appliances. The controllable appliances were TV, HIFI-system, lamp and a window shutter. In user studies with 27 persons the system was trained with 110 gestures, 42.31 per person. As expected, a correlation between recognition precision and quality rating was found.

3 Smart Home

3.1 Smart Kitchen

The smart kitchen (figure 1) at the German Research Center for Artifical Intelligence in Saarbrücken is a completely equipped kitchen with the feature that all appliances are accessbile via network. The type of connection and protocols are a various set of different technologies. Hood, oven, hob, fridge and freezer are connected via powerline technology and can be controlled by a WSDL interface. The light in the kitchen and additional sensor technology like movement, light, temperature or contact sensors for windows and doors are accessed by the battery and wireless radio technology EnOcean [1]. A television, in form of the Windows media center, runs on a desktop PC. The portfolio of appliances also includes technology for health care as bloodpressure and blood sugar measure meter.

[1] http://www.enocean.com/

The heterogenous number of different technologies complicates the rapid development of new user interfaces for an interaction with the smart home. This problem is faced by using the ISO 24752 standard URC (Universal Remote Console) realized with our own middleware framework UCH (universal control hub) that provides a consistent homogeneous solution to access the functionalities and states of all appliances. This installation serves as a test environment for different interaction concepts, e.g. a multimodal dialogue system [8] or a task based calendar that assists people with mild cognitive disabilities affecting their concentration and memory [9].

3.2 Gesture Recognition

3.2.1 The input device: WiiMote

Our input device is the WiiMote (figure 2) that was developed by Nintendo in 2006 as the main controller for their game console Wii. Besides normal controller attributes like buttons and a directional pad, it introduced the innovative concept of using three dimensional accelerometers for game control. Thus hand movements of the user are directly transfered to the game. With the Nunchuk controller, that contains three further accelerometers, this can be extended to a two handed control.

Fig. 1 The smart kitchen at the German Research Center for Artifical Intelligence Saarbrücken

3.2.2 Classification

For gesture recognition we use the classification framework TaKG that was developed at our lab for recognizing gestures recorded by 3D-accelerometer-signals [10]. The framework provides several common classifiers for learning and recognizing signals of an arbitrary number of dimensions, such as Multilayer Perceptron Neural Networks (NN), Support Vector Machine (SVM) and multidimensional Dynamic Time Warp (DTW) [11]. Furthermore several methods for feature extraction are provided. In [8][10] we used the framework in different application scenarios to recognize pantomimic gestures performed with the WiiMote-Controller by exploiting the accelerometer signal of the WiiMote. TaKG is responsible for learning new gestures and organizing them into user specific training sets. Within a set, the information for every trained gesture is listed including the measured signal data and a gesture tag denoting the gesture. The main API contains the following functionalities: Load data for a special user, learn and delete gestures and classify new recorded gestures. A gesture classifying request returns the gesture tag of the gesture in the training set with the highest similarity to the gesture which has been provided together with the request. Another option is to ask for a ranked list of all trained gestures. SVM and NN provide just a ranking, the DTW algorithm describes similarity based on Euclidian distance which gives the advantage that an ill performed gesture can be dismissed.

4 Interaction Concepts

In the related work [5][7] we observed that for any single interaction with a home appliance a distinct gesture was defined and trained to the interaction system. This inevitably leads to a very large set of gestures to remember and to a difficult and unintuitive interface. Goal of this study is to find appropriate concepts to reduce the amount of different gestures a user has to learn and to remember for his interaction with the smart home environment. This is crucial, especially for the elderly and

Fig. 2 Right: The WiiMote controller and Nunchuk. Left: Accelerometer axes.

people with a reduced cognitive capacity. The basic idea to achieve this goal is a context sensitive system. A system that knows the appliance that is actually in the user's focus needs a much smaller number of distinct gestures than a system without this information. With this additional information only one gesture, if you just consider the hand movement, can result in different interpretations, depending on the appliance that is controlled in the moment of execution: e. g. a gesture for the interpretation of activating something can switch on the television or the lamp in the living room. A gesture that is a metaphor for switching through different selections of a setting can be used to zap through television channels but also to increase the ventilator speed of a kitchen hood: One gesture - Several meanings! This is the idea our first interaction concept is built on.

4.1 Gesture control in context

Gesture-control in context offers the advantage that one gesture can trigger functionalities on several appliances. These functionalities have a similar meaning and can be grouped in function categories. Typical function categories we applied are listed in table 1:

It remains the task to indentify the appliance that is actually in focus. Originally the plan was to take advantage of the infrared-sensor of the WiiMote and a special flashing - pattern of infrared LEDs that are assembled at the different appliances. With a specific flashing frequency for every appliance it would be possible to detect the appliance a user is pointing at with the Wiimote. Unfortunately this installation was not ready early enough for the evaluation and remains an open task for future work. To simulate this system behavior the user is able to select the appliance that is actually in focus by switching through the devices with the directing-pad of the WiiMote. Speech synthesis gives feedback about the actually selected appliance that is then in the user's focus.

4.2 Gesture triggered scenarios

In [7], it is discussed that a speech controlled system should provide the feature to chain multiple commands in order to speed up the interaction. We wanted to adopt this idea for gesture controlled systems and created scenarios that can be triggered by a single gesture. One scenario exactly describes a situation in which a special behaviour of the smart environment is expected. For example, a gesture to instruct silence has influence on different appliances, e. g. a running television is muted and the speed of a fan with a high noise level is lowered. Another gesture that stands for 'go to bed' switches off all running appliances that are not needed while the user is sleeping or that would even be annoying. Obviously the idea of scenarios is also

Table 1 predefined categories

- turn on	turn on television turn on standard light turn on fan turn on light of the hood	(impact: appliance turned on)
- turn off	turn off television turn off standard light turn off fan turn off light of the hood	(impact: appliance turned off)
- increase	increase television volume increase fan speed increase light setting	(impact: appliance setting increased)
- decrease	decrease television volume decrease fan speed decrease light setting	(impact: appliance setting decreased)
- next choice	television channel	(impact: switch to next channel)
- previous choice	television channel	(impact: switch to previous channel)
- toggle mute	television	(impact: toggle between mute/unmute)

a benefit for other modalities. They could also be triggered by voice commands or GUI interactions.

5 User Tests

The user study compares the two interaction concepts we introduced in section 4 in practice. The test users have to solve a number of interaction tasks with the smart home that are embedded into an everyday life story. During these tests we measure the users' response times separated into cognitive and physical parts. The time the user needs from hearing the task until performing the gesture is the cognitive response time. The time need for performing the gesture is the physical response time. Additionally, we assessed the users' subjective ratings of the system through a questionnaire.

5.1 Test Scenario

For the evaluation we invented a story of typical everyday life situations and confronted the test persons with the exercise to operate the appliances in our smart environment according to the situation in the story. The involved appliances are television, hood with fan and light setting and a floor lamp. Every test takes around 90 minutes and is separated into three phases.

1. Test of the context sensitive interaction concept. In this phase the test person is instructed to devise personalized gestures for every single function category. There are a total number of 7 categories (table 1 gives an overview of the categories used and their influence on the different appliances). First the test persons paint the movement path on a paper and additionally describe it with words. Then they try to rank the difficulty of devising the gesture. After defining the gestures the system is trained with three performances of every gesture, so the total number of trained gestures per person is 21. In order to get the user familiarized with the interaction concept all trained gestures are tested with some exercises. Here we have the possibility to detect confusions between gestures for different categories that are too similar and in case of a confusion, replace them. After this adaption phase the user has to solve tasks that are embedded into the following everyday life story (parts of the story that require a user reaction are underlined):

   It is 7 o' clock p.m. I returned home from work and I am tired. Now I would like to relax, watch television and enjoy life. I enter the kitchen. The kitchen is dark, so I switch on the floor lamp. I also turn on the television. I take a drink out of the fridge and watch the TV program. I do not like what is on so I switch to the next channel. This program is better and I increase the volume.
   20 minutes later, I get hungry. Today I will prepare spaghetti Bolognese. I fill the pot with cold water and put it on the hob. I spend the waiting time watching TV. After some minutes the water is boiling. I would like to put the spaghetti into the pot. The steam is very strong and the light at the hob is too low. I can barely see what I am doing so I turn on the fan and the light of the extension hood and set the fan to maximum speed. The spaghetti are in the pot. I wait and continue to watch TV. Suddenly the phone in the kitchen is ringing. I pick up the phone. Because of the noise of the fan and the TV I can hardly hear the person at the other side of the phone. I lower the fan speed to level 1 and mute the television. After five minutes the phone call is over. I set everything back to the previous state so I unmute the TV and increase the fan speed. Finally the spaghetti are ready and I turn off the extraction hood.
   After the meal I am tired. I switch off all the devices in the kitchen and go to bed.

2. Test of the shortcut interaction concept. The procedure is very similar to the previous one. We predefined five shortcuts for the story (see table 2). In the beginning, the test person is introduced to the interaction concept of shortcuts and has to devise personalized gestures for the six scenarios. Again, the gesture is documented by painting the movement path and describing the gesture with words. Then every gesture is trained to the system three times resulting in a total training set of 15 gestures per person. The user plays the story again with the instruction to use shortcuts this time.

Table 2 The six control scenarios for our story

Amusement	Turn on television and floor lamp
Start Cooking	Turn on the light of the hood and increase the fan speed to maximum
Finish Cooking	Turn off the light and fan of the extraction hood
Silence/Stop Silence	Mute the TV and decrease the fan speed of the hood/Reset TV and hood to previous settings.
Go to bed/leave home	Turn off all appliances.

3. In the last phase the user has to answer a questionnaire. Here we want to find out the subjective impressions of the test persons. We asked for the cognitive and physical challenge of the tasks and difficulties during the test. Furthermore we wanted to know, whether it is more difficult to devise gestures for the context sensitive or the scenario based interaction concept. A last part collects statistical information like age, gender and technical affinity.

5.2 Evaluation

The user test serves primarily to compare both introduced interaction concepts and evaluate their efficiency by observing the reaction times and the subjective impressions of the test persons. A total number of 13 test persons took part in the test, six were male and seven female. Eight persons were between 18-30 years, two between 31-60 and three persons older than 60 years. The questionnaire revealed that 84.6 % of the test persons were convinced by the context-sensitive interaction concept. 69.3 % rated the shortcut based system as helpful or very helpful (figure 3).

Comparing the response times according to the age of the test users (figure 4) reveals that there is no gap between users younger than 30 and younger than 60 years. Nevertheless, it is worth mentioning that the users between 30-60 were technology-friendly. The users older than 60 years took significantly longer to solve the tasks with response times twice as long as the times of other the users. One reason for this is that the older users never have seen or used a gesture based control before.

Figure 5 compares the response times for both interaction concepts of our user study. It depicts the median of all measurement. Generally, it is clear that both, the cognitive and the physical reaction time, are shorter at the scenario based interaction concept. The observed time is the time needed to execute one shortcut which already contents multiple commands of the context based approach. So shortcuts already solve problems faster naturally because several commands are chained together to one. Finally, we examine the correlation between the user's efficiency in using the gesture based smart home control and their subjective rating of the control method. For this, the user should rank their confidence with the system on a scale from one to five (very easy to very difficult). The graph in figure 6 charts this subjective rating related to the response time of the users. The result is a marginal correlation of 0.425.

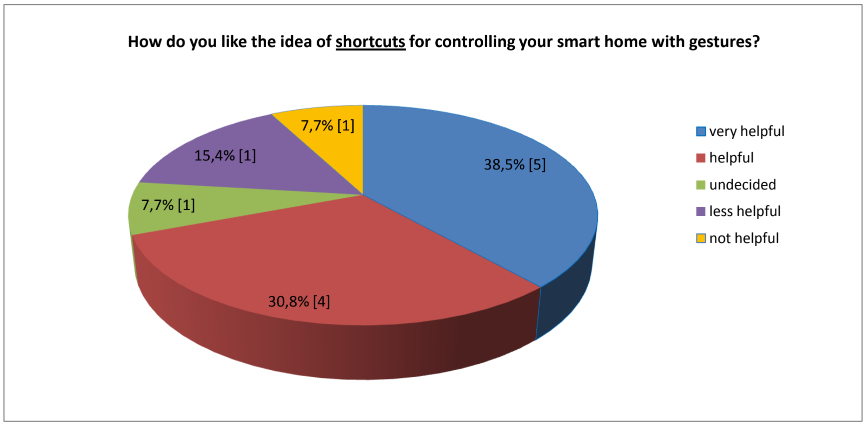

Fig. 3 How do you like the idea of shortcuts for controlling your smart home with gestures?

6 Conclusion and future work

In this paper, two interaction concepts to simplify the gesture based interaction with a smart home environment are presented. For hand movement recognition we exploit the three dimensional accelerometers of the WiiMote from Nintendo and use machine learning algorithms to learn and classify performed hand gestures. The result of the classification process is used to control appliances in a smart home, in detail an extraction hood, a floor lamp and a television. The priority of the interaction concepts is to reduce the number of gestures a user has to learn and to remember without constraining the number of control functionalities. To this end, the first concept takes the actual control context into account and the second introduces scenarios to combine simple control commands to one. Integration is realized with the ISO 24752 standard middleware Universal Remote Console (URC). Both implemented interaction concepts are evaluated in a user study with 13 test persons of different ages. The result of the study shows that 85 % of the test persons are confident with the first concept and 70 % persons with the second one. During the study we measured the cognitive and physical response time of the users and found that the median of the reaction time in the second concept is slightly smaller than in the first one. Additionally, a correlation between response time and subjective user confidence with the system was found.
In the future, we would like to develop better methods to detect the appliance that is actually in the user's focus, e.g. eye tracking, pointing gestures or context reasoned from the interaction history.

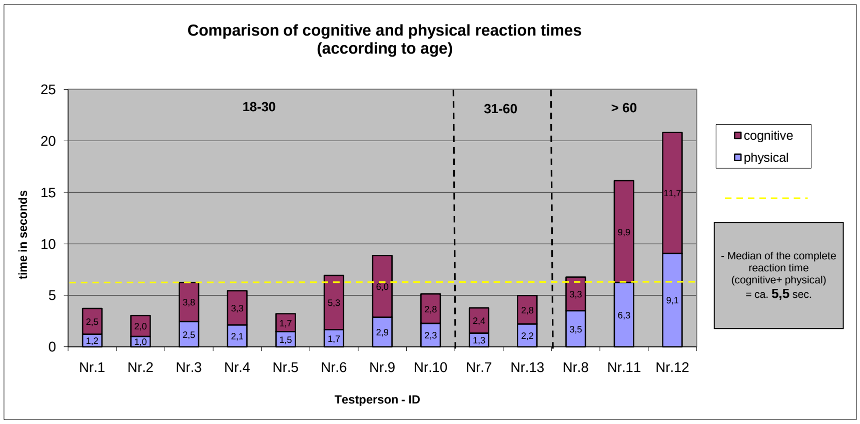

Fig. 4 Comparison of cognitive and physical reaction times according to age

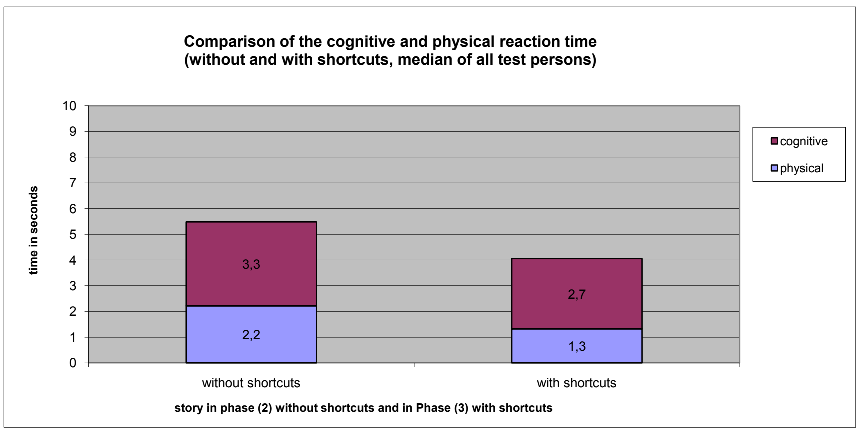

Fig. 5 Comparison of the cognitive and physical reaction time according to interaction concept

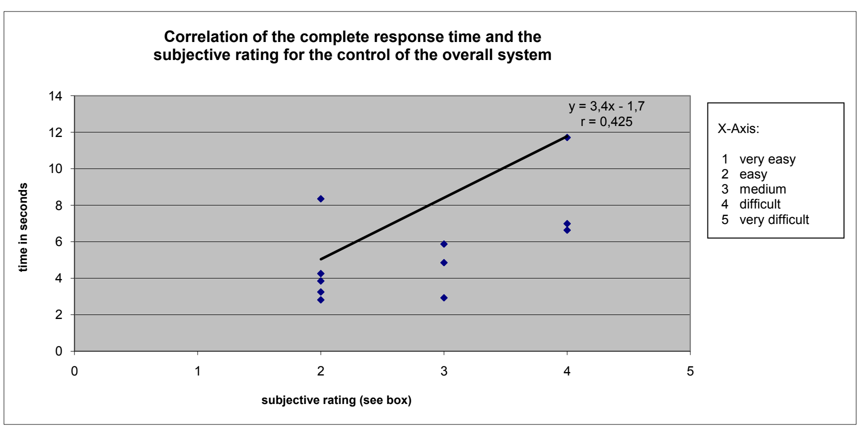

Fig. 6 Correlation of the complete response time and the subjective rating for the control of the overall system

References

1. Zimmermann, G., Vanderheiden, G.C.: The universal control hub: An open platform for remote user interfaces in the digital home. In Jacko, J.A., ed.: HCI (2). Volume 4551 of Lecture Notes in Computer Science., Springer (2007) 1040–1049
2. Burzagli, L., Emiliani, P., Gabbanini, F.: Ambient intelligence and multimodality. In Stephanidis, C., ed.: Universal Access in Human-Computer Interaction. Ambient Interaction. Volume 4555 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2007) 33–42
3. Machate, J.: Being natural: On the use of multimodal interaction concepts in smart homes (1999)
4. Song, W.G., Lim, J.T.: Design and Management in Smart Home. In: Independent Living for Persons with Disabilities and Elderly People. IOS Press (2003) 33–37
5. Rahman, A.M., Hossain, M.A., Parra, J., El Saddik, A.: Motion-path based gesture interaction with smart home services. In: MM '09: Proceedings of the seventeen ACM international conference on Multimedia, New York, NY, USA, ACM (2009) 761–764
6. Blumendorf, M., Feuerstack, S., Albayrak, S.: Multimodal smart home user interfaces. In Mukasa, K., Holzinger, A., Karshmer, A., eds.: Intelligent User Interfaces for Ambient Assisted Living: Proceedings of the First International Workshop IUI4AAL 2008, IRB Verlag (2008)
7. Westermann, T.: I'm home: Smartphone-enabled gestural interaction with multi-modal smarthome systems. In: Informatiktage. Volume S-9 of LNI., GI (2010) 137–140
8. Neßelrath, R., Alexandersson, J.: A 3d gesture recognition system for multimodal dialog systems. In: Proceedings of the 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems. Twenty-First International Joint Conference On Artificial Intelligence (IJCAI -09), in Conjunction with 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (KRPD-09), July 12, Pasadena,, California, United States, IJCAI 2009 (2009) 46–51
9. Jochen Frey, Christian Husodo Schulz, R.N.V.S.J.A.: Towards pluggable user interfaces for people with cognitive disabilities. In: Proceedings of the 3rd International Conference on Health Informatics. HEALTHINF-2010, in Conjunction with Proceedings of the 3rd International Conference on Health Informatics, January 20-23, Valencia, Spain, Springer (2010)
10. Neßelrath, R.: TaKG - Ein Toolkit zur automatischen Klassifikation von Gesten. Mastersthesis, DFKI (2008)
11. ten Holt, G., Reinders, M., Hendriks, E.: Multi-dimensional dynamic time warping for gesture recognition. In: Thirteenth annual conference of the Advanced School for Computing and Imaging. (2007)