Combining Motion Capture and Digital Human Modeling for Creating Instructions in Industrial Settings

Ulrike Schmuntzsch, Ulas Yilmaz, and Matthias Rötting

TU Berlin, Department of Psychology and Ergonomics, Chair of Human-Machine Systems
Marchstr. 23, 10587 Berlin, Germany
{usc,uyi,roetting}@mms.tu-berlin.de

Abstract. In this paper, a hybrid framework for creating an instruction video by means of motion capture technologies will be explained. In this video an animated pedagogical agent named *Anastasia* (animated assistant for tasks in industrial applications) provides human operators with assistance while performing maintenance tasks in IPS^2. Firstly, the paper contains a description of the creation process of an animated pedagogical agent which will be illustrated step by step on *Anastasia*. Secondly, the motion capture technology in form of a data glove will be presented. Thirdly, the concept and implementation how to improve realism of *Anastasia* by using the data glove will be introduced.

Keywords: animated pedagogical agent, instruction video, human modeling, motion capturing, data glove, wearable computing technologies, smart clothes.

1 Introduction

The growing global competition with emerging BRICS nations forces internationally operating industrial companies to work out a unique selling point and to gain a competitive edge. To keep up, high technology companies develop extremely customized industry solutions combining products and services. These, so called Industrial Product-Service Systems (IPS^2) are characterized by a variety of different users, contexts of usage and highly specialized products and services. Human operators working in these settings are in charge of the "supervisory control", which includes monitoring the system process and performing tasks that cannot be automated efficiently such as maintenance or overhaul (Sheridan in [12]). This constellation oftentimes forces human operators to unexpected and unpracticed interventions which could be risky for man, machine and environment (Reason in [9]).

As to the initially illustrated situation, an urgent need of coherent and understandable user support exists by taking different user populations and contexts into account (Uhlmann et.al. in [17] and Schmuntzsch in [13]). This project focuses on the design of multimodal instructions in IPS2 so that the given user support for human operators provides practical and intuitive information on actions, technical

V.G. Duffy (Ed.): DHM/HCII 2013, Part II, LNCS 8026, pp. 124–133, 2013.

processes and potential hazards. A prominent and successful method, used for instance in online markets and computer games, is to provide instructions to a multitude of different users by an animated character. However, there are only few examples when it comes to technical contexts (Ziegler and Zülke in [21]). One important reason might be the complexity and great adaptability in industrial settings which makes it difficult as well as time and cost consuming to apply animated characters for industrial tasks. However, through huge technical progress and reduced prices, motion capture technologies, traditionally applied in film and gaming industry and ergonomics research, nowadays are increasingly used in further areas (Bergler in [1], Brodie et.al. in [2] and Xsens in [20]). These increased opportunities are used in our project to combine the motion capture technology with the modeling of an animated pedagogical agent, who demonstrates how to perform a maintenance task in an instruction video. The realism of finger movements is a key to more realistic instruction videos. That's why, in this study, the focus is on increasing the realism of hand in the instruction video by integrating a motion capture system into the design.

The article is organized as follows. Hereafter in Section 2, the design process of the animated character and the therefore used software tools are presented. Section 3 gives insides how the finger movements are captured using the data glove. The mapping procedure and the overall framework are clarified in Section 4. The article concludes with a discussion of the main experiences and challenges of the realization process.

2 Animated Characters

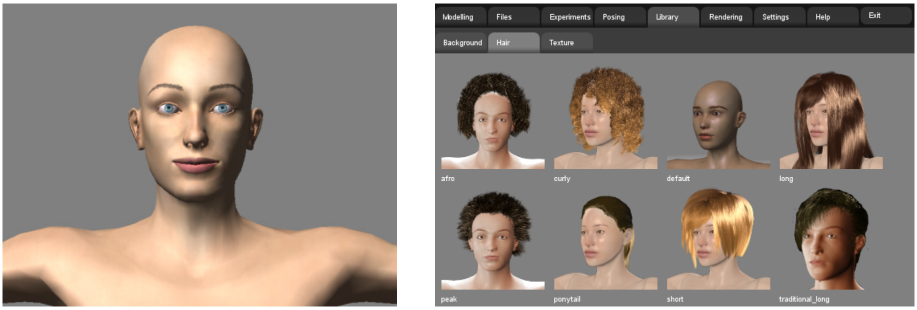
2.1 Theoretical Classification

The enormous variety of animated characters can be subdivided into two basic categories: humanoid and nonhumanoid types. For the first group, the humanoid characters, exist numerous terminologies, such as virtual human (Rickel in [10]) or digital human (Lopiccolo in [7]), digital actor or synthetic actor (Thalmann in [16]), animated pedagogical agent or virtual (interface) agent (Rickel and Johnson in [11]). Even though they all seem similar, slightly different meanings are noticeable. For the purpose of this work, an animated pedagogical agent seems appropriate. Due to having sufficient knowledge about the content and reacting properly towards user inputs, animated pedagogical agents are used to facilitate the learning processes (Lester in [6]). Apart from that group, virtual (interface) agents also suggest an interesting alternative. These characters can be split into avatars, assistants and actors. Relevant to us are especially actors because they represent a computer-based human acting in a virtual setting whose behavior cannot be influenced by the user (Mase in [8]). Functioning as instructor, advisor, motivator or companion, animated characters are widely used in different application fields, mainly in computer games and entertainment, e-commerce and learning environments. There, they not only support users by giving stepwise instructions, but also by evoking the feeling not to be left alone with a problem (Fröhlich and Plate in [3]). In several user studies, it has been shown that humanoid characters can increase user satisfaction and entertainment value (Lester in [6]). Furthermore, it has also been found out that a true-to-life representation of a virtual character leads to an increased user motivation and readiness to work (Johnson et.al. in [5]). Thus, one can state that its mere presence creates a pleasant working atmosphere for users. Working in technical contexts means learning oftentimes complex factual as well as practical knowledge. For these kinds of settings, there is much less examples for user support (Ziegler and Zülke in [21]). A relatively well known example is the animated pedagogical agent *Steve* who functions as tutor teaching students how to use or overhaul complex machines (Rickel and Johnson in [11]). Particularly for such industrial tasks, animated pedagogical agents provide many advantages. For instance, the interactive demonstrations make it possible to teach users how to perform physical tasks such as repairing equipment. That way, different sequences of events, which have to be carried out to fulfill a task, can be reproduced stepwise. Another important factor, as to Rickel in [10], is the agent's gaze and gestures. Due to their familiarity and human likeness they seem to be very suitable as attention guides. Further advantages of animated pedagogical agents are their expressed emotions and personality as well as the associated story and character. This helps to increase the learning motivation as well as to capture and to hold attention. All the mentioned aspects can be seen as key factors to trigger user's emotions and to gain a better understanding of the required technical procedures and knowledge.

2.2 The Creation Process of *Anastasia* and Her Technical Equipment

As to the above mentioned advantages of using animated pedagogical agents especially for teaching technical procedures, in the project it was decided to model a female character and to name her *Anastasia* which also stands for animated assistant for tasks in industrial applications (Schmuntzsch et.al. in [14] and [15]). This acronym indicates the function and the field of application of *Anastasia*, as well.

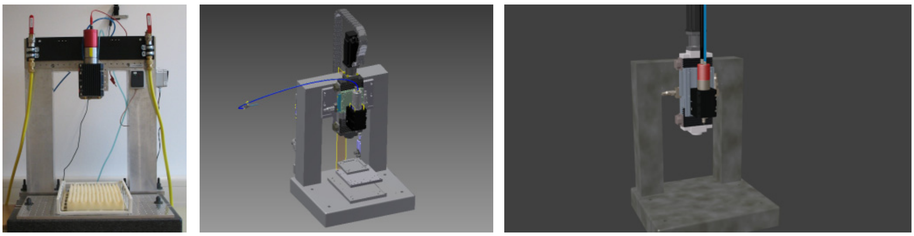
In order to finally realize the instruction video in its present form several steps had to be carried out. This meant to realize various 3D concepts, to acquire data from the data glove, to send remote animation scripts over network, and to integrate all the existing software components. Here, a mixture of different software tools such as *MakeHuman*, *Autodesk Inventor* and *Blender* had to be used. *MakeHuman* is free software for the modeling of three dimensional humanoid characters. It is released under an open source license and available for various operating systems such as *Windows*, *Mac OS X* and Linux (www.makehuman.org). *Autodesk Inventor* is 3D mechanical solid modeling design software for creating 3D digital prototypes used in the design, visualization and simulation of products (usa.autodesk.com/autodesk-inventor). Finally, *Blender* is a free and open source 3D computer graphics software product used for creating animated films, visual effects, interactive 3D applications and video games (www.blender.org). Creating *Anastasia* the first step was to design a model of an animated assistant with the software tool *MakeHuman 1.0 alpha 6.0* as shown in Figure 1a. Then, the created model was exported as so called "Blender Exchange" in mhx file format.



(a) (b)

Fig. 1. (a) The human model of *Anastasia* in *MakeHuman* and (b) various hair options

For various reasons, such as inter-software inconsistencies, nonstandard file format definitions, and integration challenges, creation of the hair of Anastasia requires special attention. Thus, the hair of Anastasia was also created in MakeHuman using one of the hair templates provided within the software as shown in Figure 1b. It was then exported as Wavefront file in obj file format with some manual settings and several small adaptations in hair and clothing, in order to make the animated assistant look natural and familiar. The human operator interacts at the workstation with several other technical components, such as machine equipment and tools. Modeling these components was also an essential part of the creation process of the instruction video. Thus, a model of the real micro milling machine (see Figure 2a) was created in Autodesk Inventor 2012 and adapted to Blender as shown in Figure 2c. Tools, such as the Allen key and the jaw spanner were also designed with the same program and adapted to Blender. During the adaption process, the created models were first exported as CAD models in stl file format, and then modified parts were combined in Blender, so that the final model could be colorized and animated realistically.



(a) (b) (c)

Fig. 2. (a) Real micro milling machine; (b) micro milling machine designed in *Inventor* and (c) imported into *Blender*

Further technical equipment used in the instruction video, such as power supply, operating computer, spindle box, laptop, tables and backgrounds, were directly designed in *Blender*.

2.3 Animating *Anastasia*

After creating the virtual character *Anastasia* and her technical equipment the further animation process of the instruction video took place in two steps. In the first step, the main body movements of *Anastasia* and the animation regarding other system components were created. This was realized manually in *Blender* "Pose Mode" as shown in Figure 3a by moving *Anastasia*'s body parts and position in the environment. Whereas in the second step, finger movements of *Anastasia* are acquired using the data glove, and integrated into *Anastasia*'s motions. This process will be explained in detail in Section 3. Both created body poses and finger movements were saved as key frames. Oral explanations synchronous to *Anastasia*'s actions are integrated as third person narration, and finally the instruction video was exported as a video file, a screenshot of which is shown in Figure 3b.

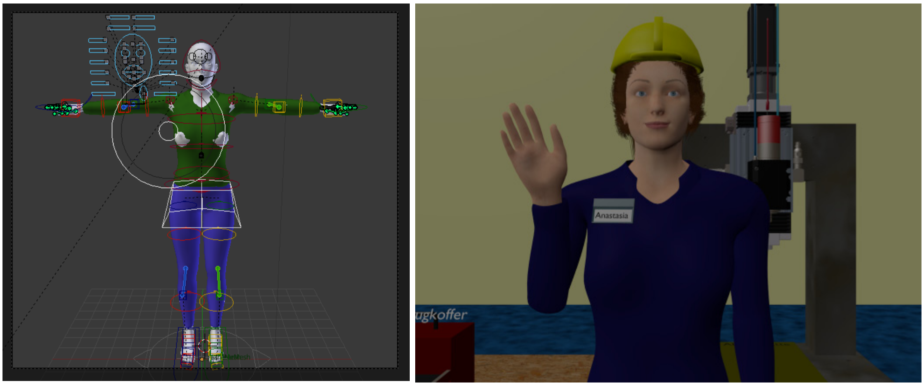


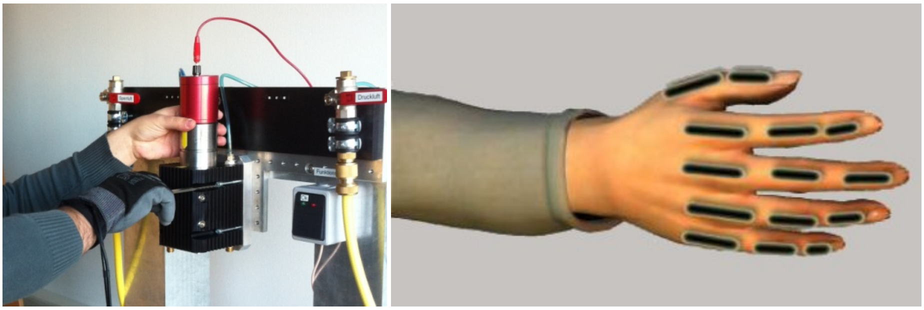
Fig. 3. (a) *MakeHuman* model imported into *Blender* and shown in "Pose Mode" where possible body movements are illustrated; (b) beginning of the instruction video

3 Motion Capture

Motion capture technologies are being used in ergonomic research and in the film industry for quite a long time (Bergler in [1], Brodie et.al. in [2] and Xsens in [20]). IPS[2], on the other hand, is a rather new field of application. Growing product complexity and increasing heterogeneity of users and contexts highlight the importance of understandable and illustrative user support such as instruction videos. Creating these videos, one can make use of motion capturing in order to increase realism, as it is widely done in film industry. However, most of the time, the focus in such videos, is on the hand, respectively on the fingers. The more realistic are the fingers modeled the more realistic is the instruction video. That's why, the focus of the study is on increasing the realism of hand in the instruction video by using a data glove to capture finger movements of the human operators. In the following two subsections, the data glove used in the study will be described. Furthermore, it will be explained in detail, how the calibration for increasing realism was managed.

3.1 Data Glove

Depending on the sensors being used, the data glove systems available on the market can be grouped into four categories: optical, mechanical, inertial and bend (VRealities in [18] and Inition in [4]). Each system has its advantages and disadvantages. For instance, optical systems are relatively cheap, but the occlusion problem is inevitable. Mechanical systems are relatively robust, but they are ergonomically hard to use. Inertial systems are precise, yet considerably expensive and very sensitive to magnetic interference. Our working environment has high magnetic interference and as the human operator moves around freely, system components are being continuously occluded. Consequently, the cheapest easy-to-use system that fits our working environment was the *X-IST Wireless DataGlove* that has bend sensors on fingers (X-IST in [19]). The data glove and the micro milling machine while the human operator is replacing the spindle are as shown in Figure 4a.



(a) (b)

Fig. 4. (a) The micro milling machine and the data glove while the operator is replacing the spindle; (b) the distribution of bend sensors on finger joints

In *X-IST Wireless DataGlove* the sensors are located on finger joints as shown in Figure 4b; two sensors being on the thumb and three on each of the rest. Each sensor delivers a 10 bit value that corresponds to the relative bend of the finger bone at that instance. The working frequency is 60 Hertz. Note that, with *X-IST Wireless DataGlove* hand rotation is poorly captured through inertial sensors and finger spread is not to be acquired at all.

3.2 Calibration

The data glove delivers the maximum bend values, when the hand is open, and the minimum values, when it is closed to an ideal fist. However, not everyone's finger joints form a perfect line when the hand is open, or they form a 90 degrees angle when the person makes a fist. So, depending on some anatomical and the habitual differences, these bend values differ for the same hand gesture of different operators. In order to normalize bend data, an interactive interface through which the human operator calibrates the data glove was implemented. With that, it is possible to get a

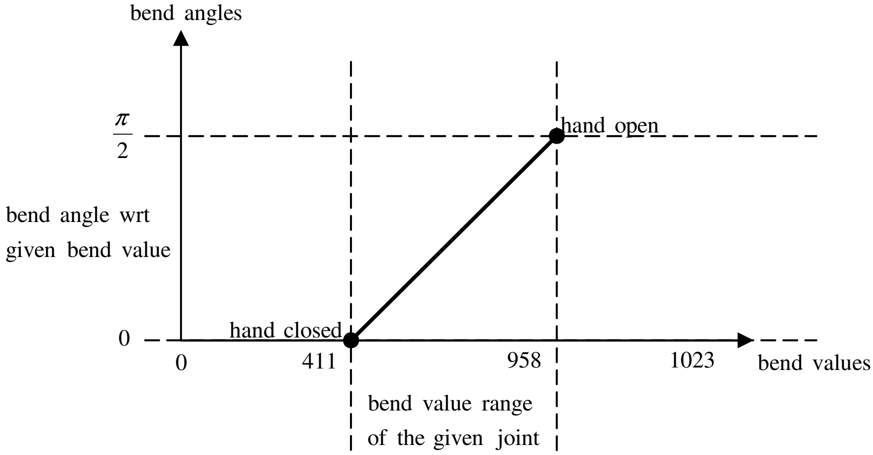


Fig. 5. A personalized mapping function converts bend value of joints to normalized bend angles

personalized mapping function that converts bend values to bend angles, which are always the same for the same hand gesture. This procedure is illustrated in Figure 5 for a single joint. The bend values occupy a portion of the possible bend spectrum, where the minimum bend value corresponds to closed hand and the maximum to the open hand. By calibrating the data glove, the best range is determined, so that the mapping function is reasonable.

As shown in Equation 1, a linear mapping function is used in the study. However, it is also possible to replace this function with partial linear functions or nonlinear functions for a better performance with different data gloves.

$$angle(v) = \begin{cases} 0 & v < v_{min} \\ \pi/2 & v > v_{max} \\ \dfrac{v - v_{min}}{v_{max} - v_{min}} \pi/2 & otherwise \end{cases} \tag{1}$$

Through the developed interface, vmin and vmax can be set manually. Furthermore, the following calibration methods are also implemented in the designed interface. Thereby, one can first perform an automated calibration and then manually refine the extracted calibration parameters.

- Greedy range extraction
- Distribution analysis
- Gesture based calibration

In greedy range extraction the human operator is asked to make possible hand gestures, which are supposed to be done during the specified instructions. This is recorded for a given amount of time. At the end, the minimum value and the maximum value, which occur in the data set, are extracted to be the v_{min} and the v_{max}, respectively. In order to avoid outliers, a median filter can be applied to the data by selecting the n^{th} extremes instead of the first ones. In distribution analysis method, human operator is asked to open and close his hand continuously for a given amount of time. Then, the minimum and the maximum value, which lie in some predefined standard deviations of the mean, are extracted to be the v_{min} and the v_{max}, respectively. In gesture based calibration the human operator is asked to make some specific hand gestures, such as open hand, fist, holding a specific tool, etc. The statistical mode of bend values is recorded during each gesture. At the end, the minimum and the maximum of these recorded values are selected to be the v_{min} and the v_{max}, respectively.

4 Mapping Process and the Making-of of the Instruction Video

4.1 Mapping Process

The designed interactive interface reads bend values from data glove using the serial connection and converts these values into bend angles which can be used with the *MakeHuman* model of *Anastasia*. Most of *Anastasia*'s movements regarding the instruction video are already prepared in *Blender*. Hence, a *Blender* plug-in was written, in order to connect our interface with *Anastasia* over TCP/IP. An illustration of this flow is shown in Figure 6.

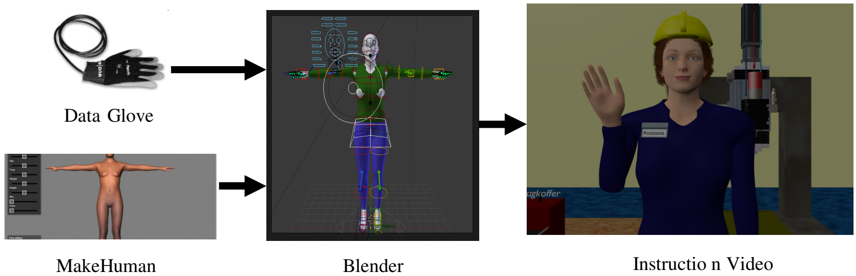
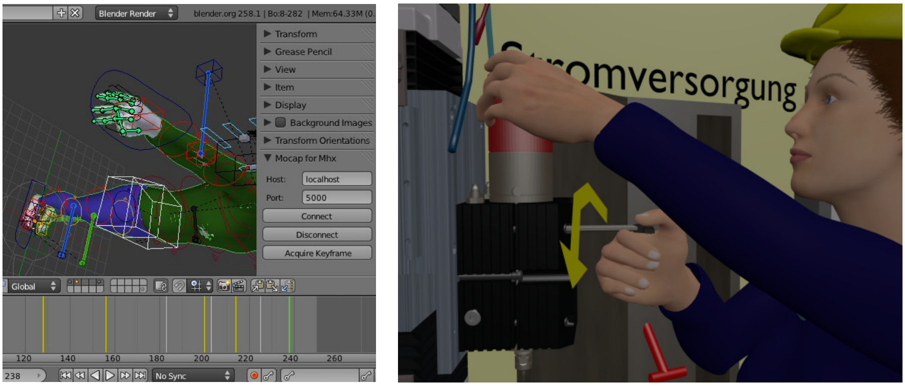


Fig. 6. System overview of capturing hand movements from serial port to *Anastasia*

The plug-in can connect to any server that provides proper Python scripts which can be used for animating mhx models in *Blender*. The designed interface, naturally, provides Python scripts for animating the right hand of *Anastasia*. Once the connection is established, bend angles of the right hand are read over TCP/IP and shown directly on the model. The user captures the motion and inserts the key frame online to the preferred position in the instruction video as shown in Figure 7a.



(a) (b)

Fig. 7. (a) The user inserts a new keyframe using the plug-in and (b) the effect of the inserted keyframe shown while *Anastasia* is using Allen key on the micro milling machine

4.2 Instruction Video

The instruction consists of many work stages. So, the entire animation was exported in several short avi video sequences (codec H.264, audio codec: mp3, resolution: 1280x800) to assist the user at its specific problem. As an example for one instruction video, a screenshot of *Anastasia* using Allen key on the micro milling machine, is shown in Figure 7b.

5 Conclusions and Future Ongoing

In this paper, a hybrid framework for creating an instruction video is presented. Here, the realism of a pedagogical agent was increased using motion capture technologies in form of a data glove. The created pedagogical agent named *Anastasia* (animated assistant for tasks in industrial applications) is used to create friendly and realistic looking instruction videos in order to provide human operators with efficient assistance during maintenance tasks in IPS^2. Creating *Anastasia* and the instruction videos, various 3D design tools such as *MakeHuman*, *Autodesk Inventor* and *Blender*, as well as our own software were used to acquire data from the data glove, to send remote animation scripts over network, and to integrate all the existing software components. Even though motion capture technologies are already in use in entertainment industry and ergonomics research, IPS^2 is rather a new field of application. Growing product complexity and increasing heterogeneity of users and contexts in this field underlines the importance of realistic and illustrative user support, such as instruction videos. Currently, the focus is on the realism of finger movements in the instruction videos, since the realism of these limbs seems to be the most important issue in the application area. However, for future development it is planned to use a motion suit by which full body motion can be acquired, and virtual reality techniques through which one can improve realism in creating instruction videos in maintenance tasks.

Acknowledgments. We thank the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) for funding this research within the Transregional Collaborative Research Project TRR 29 on Industrial Product-Service Systems – dynamic interdependencies of products and services in production area.

References

1. Bergler, C.: Motion capture technology for entertainment. IEEE Signal Processing Magazine 24(6), 158–160 (2007)
2. Brodie, M., Walmsley, A., Page, W.: Fusion motion capture: a prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing. Sports Technologies 1(1), 17–28 (2008)
3. Fröhlich, B., Plate, J.: The cubic mouse: A new device for three-dimensional input. In: CHI, pp. 526–531. ACM Press (2000)
4. Inition: Hand motion capture (2012), `http://www.inition.co.uk/3D-Technologies/productsection/43`
5. Johnson, W.L., Rickel, J.W., Lester, J.C.: Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education 11, 47–78 (2000)
6. Lester, J.C.: The persona effect: Affective impact of animated pedagogical agents. In: Human Factors in Computing Systems, pp. 359–366 (1997)
7. Lopiccolo, P.: What are 'veople' for? Computer Graphics World 25(11), 4–5 (2002)
8. Mase, K.: Aspects of interface agents: Avatar, assistant and actor. International Joint Conferences on Artificial Intelligence. In: Workshop on Animated Interface Agents, pp. 33–37 (1997)
9. Reason, J.: Human Error, 2nd edn. Cambridge University Press (2003)
10. Rickel, J.W.: Intelligent virtual agents for education and training: Opportunities and challenges. In: International Conference on Intelligent Virtual Agents, pp. 15–22 (2001)
11. Rickel, J.W., Johnson, W.L.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. Applied Artificial Intelligence 13, 343–382 (1999)
12. Sheridan, T.B.: Telerobotics, Automation and Human Supervisory Control. MIT Press, Cambridge (1992)
13. Schmuntzsch, U., Rötting, M.: Multimodal user interfaces in IPS^2. In: International Conference on Digital Human Modeling in the Context of HCI International, pp. 347–356 (2011)
14. Schmuntzsch, U., Sturm, C., Reichmuth, R., Rötting, M.: Virtual agent assistance for maintenance tasks in IPS^2. Advances in Ergonomics in Manufacturing, 221 – 231 (2012)
15. Schmuntzsch, U., Yılmaz, U., Reichmuth, R., Rötting, M.: A first approach to generate hybrid animations for maintenance tasks in IPS^2. In: Intern. Conf. on Cognitive Modeling (2012)
16. Thalmann, D.: Autonomy and task-level control for virtual actors. Programming and Computer Software 21(4), 202–211 (1995)
17. Uhlmann, E., Meier, H., Stelzer, C., Geisert, C., Bochnig, H., Sadek, K.: Design of PSS based on customer requirements. In: First International Seminar on PSS, Bochum (2008)
18. VRealities: Data gloves (2012), `http://www.vrealities.com/glove.html`
19. X-IST: Dataglove wireless (2009), `http://www.vrealities.com/x-ist.html`
20. XSENS: MVN customer examples (2012), `http://www.xsens.com/xsens-mvn-examples`
21. Ziegler, D., Zühlke, D.: Emotional user interfaces and humanoid avatars in industrial environments. In: IFAC World Congress (2005)