A Pen User Interface for Controlling a Virtual Puppet

Jean Vanderdonckt
Université catholique de Louvain (UCLouvain), LouRIM
Louvain-la-Neuve, Belgium
jean.vanderdonckt@uclouvain.be

Radu-Daniel Vatavu
MintViz Lab, University Ştefan cel Mare of Suceava
Suceava, Romania
radu.vatavu@usm.ro

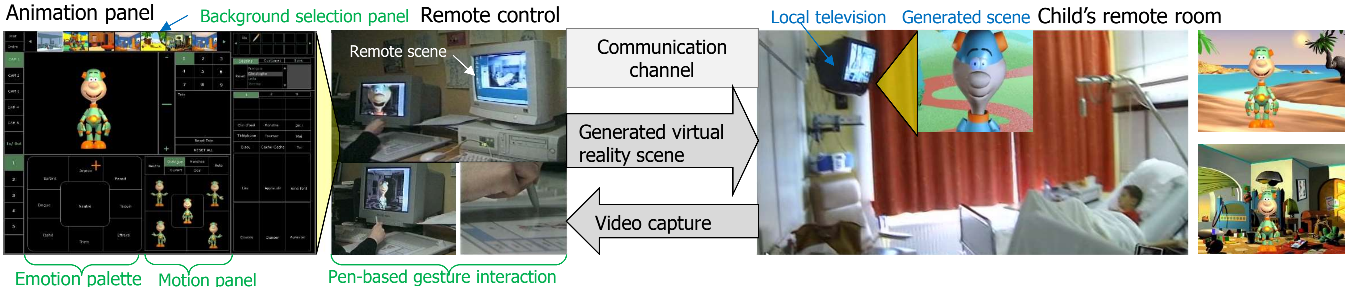

Figure 1: Context of use for our application: a psychologist, from a remote location, controls a virtual puppet via an animation panel and pen gestures (left) generating a Virtual Reality story displayed in the hospital room of a child (right).

ABSTRACT

We present an application and corresponding user interface where the user, a psychologist, controls a virtual puppet (a cartoon-like character in Virtual Reality) that displays human-like emotions for the benefit of hospitalized children. A crossing-based UI enables specification of emotional states for the virtual puppet, which are chosen from a widget UI control showing several categories of body poses, movements, and emotions. Input is performed with pen gestures that trigger animated transitions between the states of the virtual puppet. Gesture articulation parameters, such as gesture speed and the pressure applied on the touchscreen, control the rendering of the various animations. We discuss a practical application for the pediatric departments of hospitals.

CCS CONCEPTS

• **Human-centered computing** → **Graphical user interfaces**; **Gestural input**; *Pointing device*; *Touch screens*; • **Software and its engineering** → Virtual worlds software;

EICS '20 Companion, June 23–26, 2020, Sophia Antipolis, France

ACM ISBN 978-1-4503-7984-7/20/06…$15.00
https://doi.org/10.1145/3393672.3398637

KEYWORDS

Pen gesture input, Virtual puppetry, Digital puppetry, Virtual Reality, Remote manipulation, User interface.

ACM Reference Format:
Jean Vanderdonckt and Radu-Daniel Vatavu. 2020. A Pen User Interface for Controlling a Virtual Puppet. In *ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '20 Companion), June 23–26, 2020, Sophia Antipolis, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3393672.3398637

1 INTRODUCTION

Virtual Reality (VR) is expected to deliver a sense of deep presence by displaying photorealistic scenes and characters, such as virtual puppets [6, 12, 14] that exhibit human-like behavior [16]. It is known that humans tend to anthropomorphize virtual characters by projecting their own emotional states [18]. Although simulations of human-like emotions can be rendered in ways that look realistic, there is always the risk of psychological rejection when users notice discrepancies with their own expectations of human-like interactions, a phenomenon known as the "uncanny valley" [15]. Therefore, the behavioral design of virtual characters should compromise between displaying human-like emotions at a level that is sufficient enough to trigger emotional responses and life-like behavior in a ultra-realistic virtual scene, which may induce a negative response from the user [17]. In this paper, we propose an application and user interface in which the user, a psychologist, searches for the optimum balance between the two when manipulating a cartoon-like character for the benefit of hospitalized children; see Fig. 1.

2 RELATED WORK

Since cartoon-like characters that display emotions trigger more responses in humans than those without an emotional/behavioral component, it is important to convey emotion effectively. From the wide literature available on this topic, we identified two references useful to us. The first one, the Hierarchical Model of Affect (HMA) [10], decomposes human emotions into three layers (see Fig. 2a):

(1) *Emotions* are short-term human affects that last for a short time, after which they vanish from perception.
(2) *Moods* are negative or positive medium-term affects that are conveyed by the underlying emotions.
(3) *Personality* is a long-term affect produced by the dominant moods over a longer time period.

This model is useful to understand how to assign moods or a personality to a virtual character, but does not specify which of the emotions are basic affects. To address the latter aspect, we use the Arousal-Valence model [18] that projects emotions on two dimensions (Fig. 2b): arousal shows how exciting an emotion is and valence how positive it is.

Gesture-based User Interfaces

Since our virtual character is controlled by means of gesture input, we overview in the following aspects regarding gesture-based interaction. A *flick* is represented by a quick stroke in a specified direction: *e.g.*, a left flick to delete an item from a list, flicks up and down to scroll a list of items, and diagonal flicks are used to indicate targets [5, 7]. When successive flicks are combined, they form unistrokes, such as the gestures used in *marking menus* [26]. Zhao & Chignell [26] demonstrated how a suite of flicks can form a marking menu, *e.g.*, a Chinese style glyph, which positively influenced user performance for item selection tasks. The visual appearance of marking menus assists with their navigation, while the grouping and categorization of menu items according to their visual salience increases remembrance. In most applications involving *stroke-gesture input*, gesture templates are employed for classification purposes [5, 7, 21, 25, 25]. However, the same gesture type can be articulated in various ways in terms of its location inside the active sensing area, its size, orientation, and so on [2, 20]. Independence of gesture recognizers with respect to such articulation details offers great input flexibility for end users [19, 20]. Integrating pen-gestures in graphical user interfaces can be achieved in different ways, such as by package, toolkit, or API [4].

Digital Puppetry

Gestures have been considered for manipulating virtual puppets. For example, Euclide [6] exploited a combination of multiple devices (data glove, MIDI music board, keyboard, and mouse) for the multimodal control of a puppet in VR. Barrientos [3] developed an application to transfer gestural attitudes to an avatar; the avatar behaved like a virtual puppet manipulated by strings and the system tracked the motion gestures of the user and projected them onto the avatar. Oshita [17] developed a human-like virtual character with expressions controlled by the pressure and tilt specified with an interactive pen: the character could walk, run, and turn along a specified trajectory according to the speed of the movement induced on the pen.

The Virtual Puppetry system [14] deployed computer-assisted storytelling systems by exploiting motion sensing and gesture control to enable multiple users to narrate and interact with virtual characters in the VR scene. The Virtual Marionette [12] was a study on digital puppetry taking the form of an interdisciplinary approach that brought together the art of puppetry and digital animation. In the Mani-Pull-Action system [13], Leap Motion was used to acquire 3-D gestures performed with the hand to control a cinematic interactive application inspired by the art of marionettes [24]. This setup mimics the genuine gestures performed by an artist manipulating a real puppet, leaving the hands and fingers free over the Leap Motion, thus representing probably the best match between the reality and the virtual environment for such an artist. Without entering into this level of sophistication, other devices, such as ring devices [9], could be explored to add the third dimension.

3 CONTEXT OF USE OF OUR APPLICATION FOR HOSPITALIZED CHILDREN

The context of use envisioned in this paper for our application concerns the pediatrics departments of hospitals, where children undergo operations that may cause trauma. *Our hypothesis is that children will freely and frankly talk to a virtual character in order to share with it their pains, concerns, and worries [11].* From a remote location (see Fig. 1), a psychologist initiates the discussion with the child with a voice that is altered (to resemble the voice that the virtual character would have) and operates the virtual character (taking in our case the form of a kind extraterrestrial being, chosen to avoid too much anthropomorphism) displayed on the TV set available in the child's hospital room [11]. Equipped with a headset, a graphic tablet, and a stylus (Fig. 1, left), the psychologist-animator can see the child's reactions at all times through a live video feed. The conversation is implemented via animations of the virtual character (whose lip movements are synchronized with the voice of the animator), a microphone, and an animation widget control panel from the user interface available to the psychologist-animator. The latter permits a wide range of expressions, attitudes, and emotions for the virtual character to display [11].

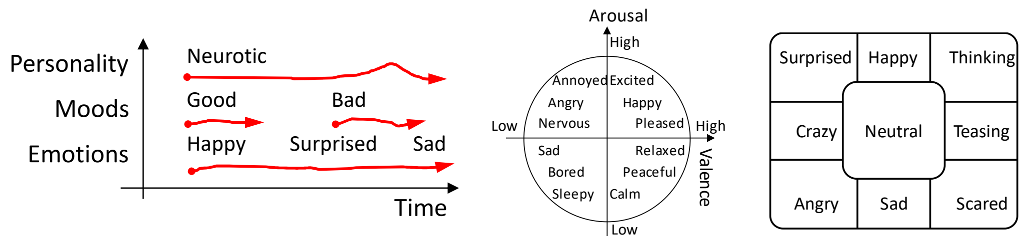

Figure 2: The Hierarchical Model of Affect over time [10], the Arousal-Valence plane [18], and the resulting emotion palette.

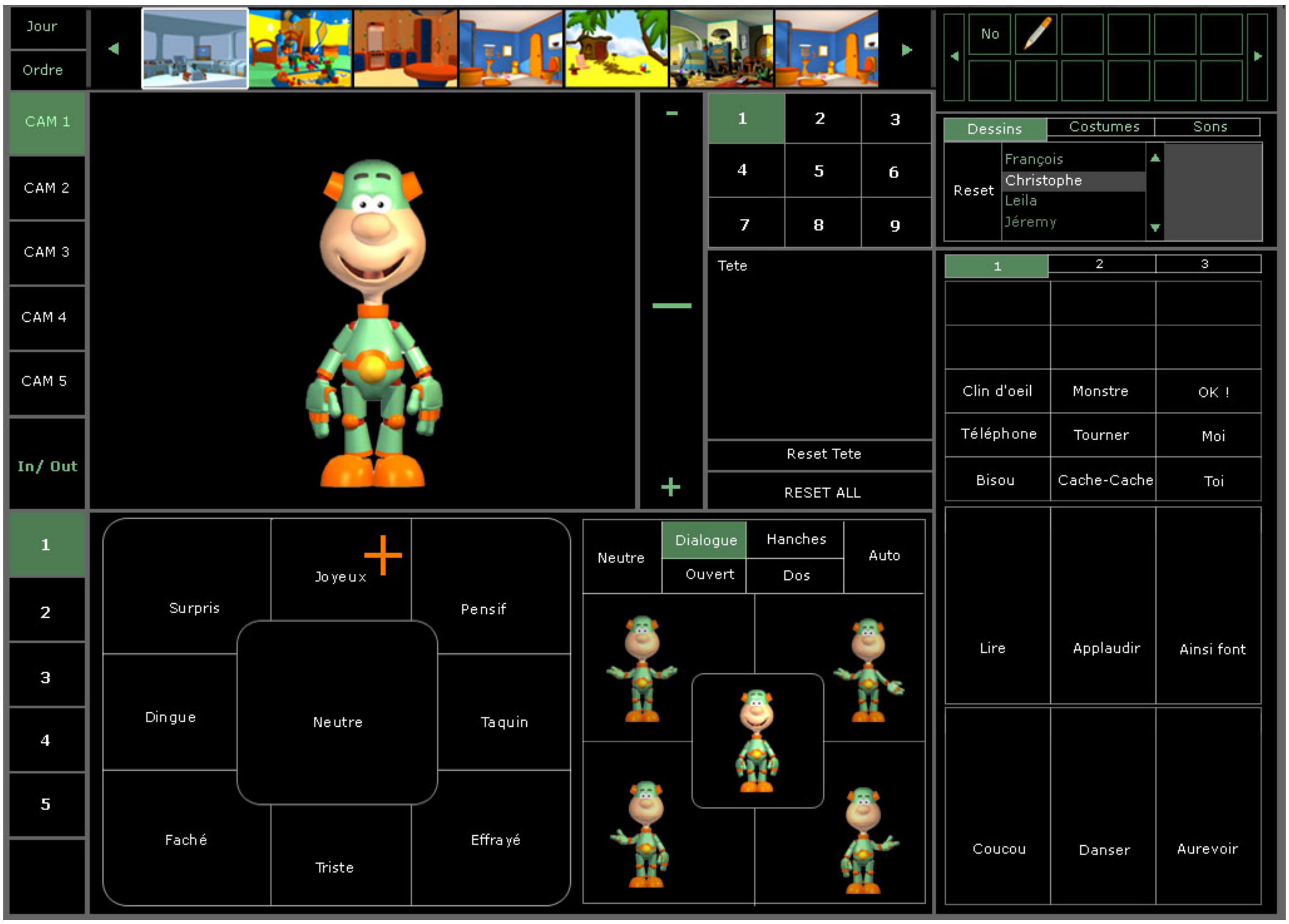

Figure 3: The animation control panel showing various emotional states that can be displayed by the virtual character.

4 A PEN-GESTURE USER INTERFACE FOR VIRTUAL PUPPET CONTROL

Fig. 3 shows a screenshot of the animation panel for controlling the virtual puppet (also shown in Fig. 1 from the first page). The user interface is composed of several interaction units: at the top, the background selection panel that sets the context for the virtual character (*e.g.*, in the bedroom, on a planet, inside a spatial vessel); the video camera selection panel (left) selects the hospital room and a video camera from that room; the emotion palette to convey the emotions of the virtual characters via pen gestures; and a motion panel to specify limb movements (*e.g.*, the virtual puppet raises its hand), and predefined animations (*e.g.*, applaud, laugh, run). A rendering of the virtual character is shown in the top-left panel for the animator. To control the virtual character, the animator can employ the following palette of gestures.

Zero-D gestures are represented by single taps by which the animator selects an emotion from the emotion palette (Fig. 3). The emotions were inspired from Fig. 2 by arranging them in a circular pattern, as follows: North (N) is happy, North-East (NE) is thinking, East (E) teasing, South-East (SE) scared, South (S) sad, South-West (SW) angry, West (W) crazy, and North-West (NW) is surprised. A tap performed in the central area of the palette returns the virtual character to its emotion-neutral state. While the animator is speaking, the virtual character shows lip movements that correspond to vowels. When an emotion is selected, a small pause between utterances inserts the specified emotion. Repeating a tap stresses the emotion displayed between consecutive utterances. A tap performed with more pressure maintains the emotional state for a prolonged time duration and with more emphasis. Taps are also used to specify the background of the virtual scene (Fig. 1, top left). Predefined animations of the virtual character, *e.g.*, applaud, walk, jump, open arms, close hands, are also executed by taps performed in distinct regions of the user interface.

1-D gestures are represented by simple flicks with various orientations that enable the animator to convey a particular emotion from the palette by ensuring smooth transition from the neutral state to that emotion. Since the palette has zones that can be crossed, the principles of crossing-based user interfaces [1] are used: as soon as the flick enters one zone, the corresponding emotion is produced. Emotions are dependent on the gesture speed and pressure controlled by the psychologist-animator: transitioning from the neutral state to an emotion zone determines the speed of the animated transition between the states of the virtual character, while the exerted pressure (with three distinct states) controls the intensity of the emotion displayed. The faster the flick gesture, the shorter the animated transition between states. The more pressure exerted, the more emphasized the emotion is. Combining speed and pressure results in a large number of options for the psychologist-animator. For example, the animator can depict a short "angry" state followed by a long "happy" one before returning to neutral.

2-D gestures are geometrical shapes exploiting the full surface of the emotion palette (Fig. 4). Gestures that render emotions (see gesture category no. 1 in Fig. 4) start from the neutral zone in the direction of any of the emotional states from the circular palette. The virtual character stays in the specified emotional state until further commands are performed. To change the emotion, the animator either selects another emotion directly (by tapping on it) or performs a gesture toward the new emotion (by means of a unistroke: see gesture examples no. 3 and 4 from Fig. 4). Gestures that start from the neutral zone, enter into an emotion zone, and then return to neutral (category no. 2 in Fig. 4) convey a particular emotion whose duration and intensity are regulated by gesture speed and pressure. These gestures are intended to convey a transient emotion, such as being astonished, angry, or sad. When the virtual character remains in a particular emotion zone, *e.g.*, "happy," an emotion-transition gesture (category no. 3 in Fig. 4) can be performed that starts from the current emotional zone ("happy"), traverses the central zone ("neutral"), and then enters another emotion (*e.g.*, "sad"). The speed of the gesture determines the speed of the animated transition. Unlike the gestures that specify emotional states directly, a smooth transition is ensured in this case. An emotion-transition gesture can be vertical (see gesture category no. 3 in Fig. 4) or horizontal (no. 6 in Fig. 4). Adjacent gestures represent a particular type of emotion-transition: instead of considering a target emotion, the adjacent emotion, either in the clockwise or counterclockwise direction (categories no. 5 and 6 in Fig. 4), is selected without passing through the neutral zone. Since there is no return to neutral, the transition between the two emotions is applied directly. Repeated adjacent gestures (categories no. 7 and 8 in Fig. 4) transition more than one adjacent emotional zone. For instance, a gesture can start from "happy," go through "surprised," and finish with "happy" again, enabling the psychologist-animator to create complex emotions.

Gesture speed was computed as the magnitude of the instant velocity between uniformly sampled points (isochronicity [19]). Linear gestures were recognized by computing the orientation of the flicks from one of the eight possible directions [7]. The gestures illustrated in Fig. 4 were implemented as unistrokes. Multistroke gestures were also recognized by turning them into unistrokes when the duration between their component strokes was under a threshold.

5 TOWARD AN EVALUATION OF THE APPLICATION FOR VIRTUAL PUPPET CONTROL AND HOSPITALIZED CHILDREN

It was not our goal in this work to perform a thorough evaluation of the user interface, but we rather focused on the technical engineering aspects of our application. Prior work has showed encouraging preliminary results for conveying emotions via cartoon-like characters [18] for the specific context of pediatric departments [8, 11], which is also our context of use. In this section, we briefly present preliminary evaluation aspects for our application:

- The cartoon-like character elicited curiosity, interest, and enthusiasm. Children became progressively attached to the virtual character to the point that some actually stole the TV remote control when they left the hospital, believing that the virtual character would be present on their TV set as well.

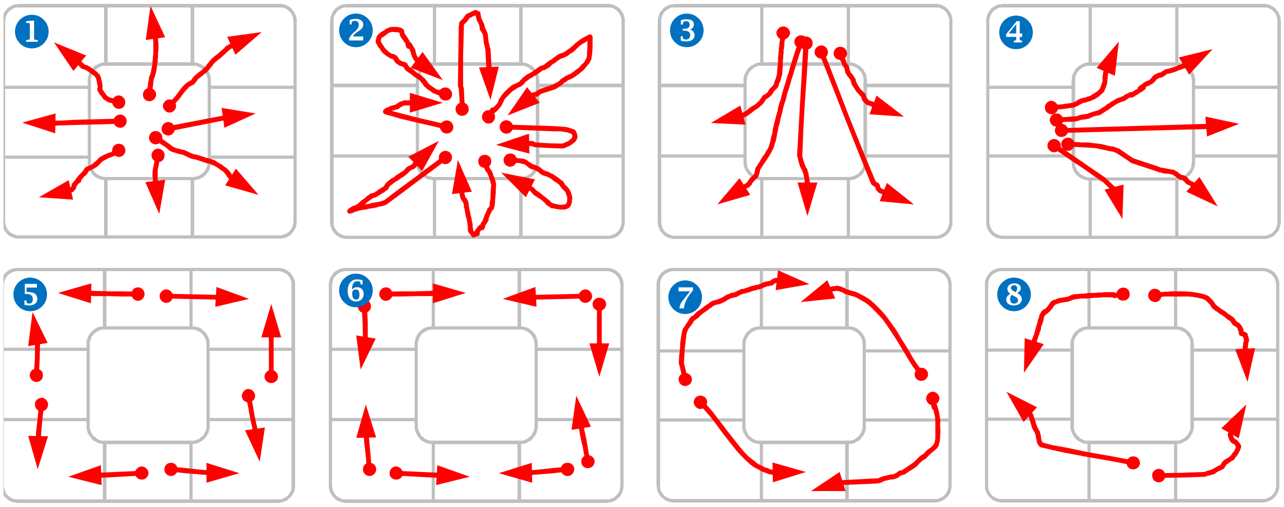

Figure 4: The gesture set used in this work to specify and control the emotions displayed by the virtual character.

Figure 5: A dragon virtual puppet may exhibit different behavior and emotions compared to a human-like character.

- The virtual character played the role of a social mediator between children and adults, *e.g.*, their parents, the psychologist, medical personnel, etc.
- The relationship established between children with the virtual character reduced the children's stress, especially in critical conditions after surgical operations, and improved the children's perceptions of pain.
- By distracting children from the context of a hospital environment and from their own illness and by changing their moods, the virtual character made possible and facilitated medical acts which were regularly perceived as painful by children or difficult to perform on children by the medical personnel.

6 CONCLUSION AND FUTURE WORK

We presented an application and gesture-based user interface for expressing human-like emotions conveyed by a virtual character. We designed three classes of gestures to depict individual emotions and transitions between multiple emotional states. Future work will address evaluations of our gesture set as well as of the user experience of our application. The set of gestures designed for this work is based on a simplified version of the Arousal-Valence model. We did not apply standard methodology for user-defined gestures [22, 23] to inform our set of gesture commands, but rather designed the gestures to enable fluid transitions between various emotional states arranged in a circular pattern.

Future work will also look at incorporating a larger set of emotions, but without increasing considerably the number of gesture commands to display them. While some of the gestures that could be designed for this purpose might be transferred from other contexts of use [20] (*e.g.*, gestures for smooth transitions between emotional states and gestures showing variations in speed and pressure), user studies and experiments to evaluate the opportunity of this design option are needed.

Finally, only individual moods are conveyed through our user interface, but the ways in which they could be combined to render more complex aspects of the personality of the virtual character, according to the Hierarchical Model of Affect (Fig. 2), represent interesting directions for future work as well. For example, different moods and personalities could be attributed to different virtual characters, *e.g.*, the dragon shown in Fig. 5 may exhibit behavior and emotions that are different from those of a human-like character.

ACKNOWLEDGMENTS

The virtual puppet depicted in this paper is Doctoon, a virtual character designed by NeuroPlanet (http://www.doctoon.com). Demo videos of this application are available at https://www.youtube.com/user/Parnamont/featured, https://youtu.be/DGTH-XO9Bkk, and https://youtu.be/6FhPrd8EgBk.

REFERENCES

[1] Johnny Accot and Shumin Zhai. 2002. More than Dotting the i's — Foundations for Crossing-Based Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. Association for Computing Machinery, New York, NY, USA, 73—-80. https://doi.org/10.1145/503376.503390

[2] Lisa Anthony, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2013. Understanding the Consistency of Users' Pen and Finger Stroke Gesture Articulation. In *Proceedings of Graphics Interface 2013 (GI '13)*. Canadian Information Processing Society, CAN, 87–94.

[3] Francesca Barrientos. 2000. Continuous Control of Avatar Gesture. In *Proceedings of the 2000 ACM Workshops on Multimedia (MULTIMEDIA '00)*. Association for Computing Machinery, New York, NY, USA, 5–8. https://doi.org/10.1145/357744.357747

[4] François Beuvens and Jean Vanderdonckt. 2012. Designing Graphical User Interfaces Integrating Gestures. In *Proceedings of the 30th ACM International Conference on Design of Communication (SIGDOC '12)*. Association for Computing Machinery, New York, NY, USA, 313–322. https://doi.org/10.1145/2379057.2379116

[5] Alessandro Carcangiu and Lucio Davide Spano. 2018. G-Gene: A Gene Alignment Method for Online Partial Stroke Gestures Recognition. *Proc. ACM Hum.-Comput. Interact.* 2, EICS, Article Article 13 (June 2018), 17 pages. https://doi.org/10.1145/3229095

[6] Céline Coutrix, Giulio Jacucci, Anna Spagnolli, Lingyi Ma, Matti Helin, Gabriela Richard, Lorenza Parisi, Stefano Roveda, and Prayag Narula. 2010. Engaging Spect-Actors with Multimodal Digital Puppetry. In *Proc. of the 6th Nordic Conf. on Human-Computer Interaction (NordiCHI '10)*. Association for Computing Machinery, New York, NY, USA, 138–147. https://doi.org/10.1145/1868914.1868934

[7] Adrien Coyette, Sascha Schimke, Jean Vanderdonckt, and Claus Vielhauer. 2007. Trainable Sketch Recognizer for Graphical User Interface Design. In *Proceedings of IFIP TC 13 Int. Conf. on Human-Computer Interaction (INTERACT '07)*, Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa (Eds.). Springer, Berlin, Heidelberg, 124–135. https://doi.org/10.1007/978-3-540-74796-3_14

[8] Amélie Druez. 2005. *Approche de l'influence de DocToon sur le vécu de l'hospitalisation d'enfants de 7 à 12 ans.* Master Psychology thesis. Free University of Brussels.

[9] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 623–635. https://doi.org/10.1145/3196709.3196741

[10] Sumedha Kshirsagar. 2002. A Multilayer Personality Model. In *Proceedings of the 2nd International Symposium on Smart Graphics (SMARTGRAPH '02)*. Association for Computing Machinery, New York, NY, USA, 107–115. https://doi.org/10.1145/569005.569021

[11] Benoît Labaye, Nadia Guérin, and Sébastien Dohogne. 2001. DocToon© — A Mediator in the Hospital of the XXIst Century. In *Virtual Storytelling Using Virtual Reality Technologies for Storytelling*, Olivier Balet, Gérard Subsol, and Patrice Torguet (Eds.). Springer, Berlin, Heidelberg, 171–180. https://doi.org/10.1007/3-540-45420-9_19

[12] Luís Leite. 2012. Virtual Marionette. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI '12)*. Association for Computing Machinery, New York, NY, USA, 363–366. https://doi.org/10.1145/2166966.2167049

[13] Luis Leite and Veronica Orvalho. 2017. Mani-Pull-Action: Hand-Based Digital Puppetry. *Proc. ACM Hum.-Comput. Interact.* 1, EICS, Article Article 2 (June 2017), 16 pages. https://doi.org/10.1145/3095804

[14] Hui Liang, Jian Chang, Shujie Deng, Can Chen, Ruofeng Tong, and Jianjun Zhang. 2015. Exploitation of Novel Multiplayer Gesture-Based Interaction and Virtual Puppetry for Digital Storytelling to Develop Children's Narrative Skills. In *Proc. of the 14th ACM SIGGRAPH Int. Conf. on Virtual Reality Continuum and Its Applications in Industry (VRCAI '15)*. ACM, New York, NY, USA, 63–72. https://doi.org/10.1145/2817675.2817680

[15] Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* 146 (2016), 22 – 32. https://doi.org/10.1016/j.cognition.2015.09.008

[16] D. Monteiro, H. Liang, J. Wang, L. Wang, X. Wang, and Y. Yue. 2018. Evaluating the Effects of a Cartoon-Like Character with Emotions on Users' Behaviour within Virtual Reality Environments. In *Proc. of the IEEE Int. Conf. on Artificial Intelligence and Virtual Reality (AIVR '18)*. IEEE, 229–236. https://doi.org/10.1109/AIVR.2018.00053

[17] Masaki Oshita. 2005. Pen-to-mime: Pen-based interactive control of a human figure. *Computers Graphics* 29, 6 (2005), 931 – 945. https://doi.org/10.1016/j.cag.2005.09.010

[18] G. S. Thirunavukkarasu, H. Abdi, and N. Mohajer. 2016. A smart HMI for driving safety using emotion prediction of EEG signals. In *Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC '16)*. IEEE, 004148–004153. https://doi.org/10.1109/SMC.2016.7844882

[19] Jean Vanderdonckt, Paolo Roselli, and Jorge Luis Pérez-Medina. 2018. !!FTL, an Articulation-Invariant Stroke Gesture Recognizer with Controllable Position, Scale, and Rotation Invariances. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 125–134. https://doi.org/10.1145/3242969.3243032

[20] Radu-Daniel Vatavu. 2017. Characterizing gesture knowledge transfer across multiple contexts of use. *Journal of Multimodal User Interfaces* 11, 4 (2017), 301–314. https://doi.org/10.1007/s12193-017-0247-x

[21] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as Point Clouds: A $P Recognizer for User Interface Prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*. Association for Computing Machinery, New York, NY, USA, 273–280. https://doi.org/10.1145/2388676.2388732

[22] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3390–3402. https://doi.org/10.1145/2858036.2858228

[23] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob A. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of ACM Int. Conf. on Designing Interactive Systems (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 25. https://doi.org/10.1145/3357236.3395511

[24] Benjamin Walther-Franks, Tanja Döring, Meltem Yilmaz, and Rainer Malaka. 2019. Embodiment or Manipulation? Understanding Users' Strategies for Free-Hand Character Control. In *Proceedings of Mensch Und Computer 2019 (MuC '19)*. Association for Computing Machinery, New York, NY, USA, 661–665. https://doi.org/10.1145/3340764.3344887

[25] Jacob O. Wobbrock, Brad A. Myers, and John A. Kembel. 2003. EdgeWrite: A Stylus-Based Text Entry Method Designed for High Accuracy and Stability of Motion. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. Association for Computing Machinery, New York, NY, USA, 61–70. https://doi.org/10.1145/964696.964703

[26] Shengdong Zhao and Mark Chignell. 2007. Using Glyphs to Facilitate Transition from Hierarchical Selection to Gesturing. *IFAC Proceedings Volumes* 40, 16 (2007), 206–211. https://doi.org/10.3182/20070904-3-KR-2922.00036