Heuristics for NUI Revisited and Put into Practice

Vanessa Regina Margareth Lima Maike[1(✉)], Laurindo de Sousa
Britto Neto[1,2], Siome Klein Goldenstein[1],
and Maria Cecília Calani Baranauskas[1]

[1] Instituto de Computação, Universidade Estadual de Campinas (UNICAMP),
Campinas, Brazil
{vanessa.maike,siome,cecilia}@ic.unicamp.br,
laurindoneto@ufpi.edu.br
[2] Departamento de Computação, Universidade Federal Do Piauí (UFPI),
Teresina, Brazil

Abstract. Natural User Interfaces (NUIs) represent a strong tendency for interaction with new computational technologies. They also represent a big challenge for designers, since delivering the promised feelings of naturalness is not trivial. In this paper, we revisit a set of 23 heuristics for NUI applications within the context of three experiments to evaluate the design of two scenarios of using NUI as assistive technology systems. While using the initial set of heuristics, they also were evaluated. Results of the experiments led to a leaner set of 13 NUI heuristics, with a compliance scale ranging from –4 to 4. The heuristics in the revisited set were defined, described and illustrated in the context of the experiments, so that they can be useful for designers and evaluators.

Keywords: Natural user interfaces · Design · Usability · Accessibility

1 Introduction

New technological devices have experimented input and output methods based on gestures, touch, and voice, considered more "natural" for interaction than the conventional mouse and keyboard. The new forms of interaction provided by Natural User Interfaces (NUIs) should evoke the feeling of naturalness in their users, by fitting the executed task to its context, and by meeting the user's capabilities [11]. This naturalness makes it possible to address a wide variety of contexts. For instance, Nebe et al. [5] propose using a multi-touch table for the management of disaster control, allowing several users to interact with a map at the same time. Renzi et al. [9] propose a serious game with a gesture-based interface to teach music concepts for children. Ringland et al. [10] show how a NUI to create paintings on a projected surface can help children with neurodevelopmental disorders. Finally, Bolton et al. [1] present an exergame that uses virtual reality goggles, a Kinect, and a stationary bicycle so that users can exercise while they play a game based on the concept of delivering newspapers with a bike.

Each of these examples employ different input and output methods with distinct purposes and for varied types of users. They all try to make the interactions between users and computers more natural and seamless.

Despite the numerous examples in literature of studies involving NUIs, there is a debate [7, 8] around the use of the term "natural" and its implications in interaction design. We believe, however, that this is an indication that successfully designing a NUI is a challenge that involves more than considerations regarding input and output technologies. Therefore, to face this challenge, a set of 23 heuristics for NUIs were proposed [4]. These heuristics were the result of a systematic literature review that also aimed at finding the state of the art of the use of NUIs to assist people with disabilities. In this paper, we present the results of applying these heuristics in practical contexts of design and evaluation of different NUI applications scenarios. In Sect. 2 we present a description of the experiments we conducted and the main results obtained. In Sect. 3 we show how the original set of 23 heuristics was revisited based on analysis of their use, and we present for each new heuristic a description and an example of use. Finally, in Sect. 4 we present concluding remarks.

2 The Heuristics in Practice

The heuristics proposed by Maike et al. [4] were applied in three different experiments involving NUI scenarios; Table 1 presents a summary of each experiment. The three experiments also had a common feature: they were all preliminary studies with the goal of finding critical system bugs, technical issues and usability problems in their tested systems.

First, let us detail Experiment 1. It followed these steps:

1. Thirteen participants were registered in the database, with five pictures for each, from different angles and distances. Therefore, two participants were left out to act as unknown.
2. One of the participants volunteered to act as a blind user. Before being blindfolded, this person received instructions on how to access the GFR software through voice commands, and on how to aim the smartwatch to capture people's faces. He was also instructed that his goal was to find and recognize (by their name or as unknown) the people that would be in front of him.
3. In silence, other four participants were placed in front of the blindfolded user, and at least one of them was not registered in the database.
4. The timer started counting and the blindfolded user accessed the GFR application. For each person she found, she must say aloud who she believes that person is, based solely on the feedback received from GFR. Timer stops when the user signalizes she believes she has achieved her goal.
5. Steps 3 and 4 were repeated twice for the same blindfolded user and two different sets of four individuals to be recognized.
6. Steps 2 to 5 were repeated four more times with a different participant acting as the blind user and different sets of people to be recognized.

Table 1. Tested system and participants of each experiment

Experiment	Tested System	Participants
1	Gear Face Recognition (GFR) [2], a face recognition software installed in a first-generation Samsung Gear smartwatch. Developed especially for visually impaired users, this software application has the goal of helping them find and recognize people in their surroundings. The software offers several sound cues to help the user when framing a person's face (a voice says "recognizing") and identifying that person from the database (a voice says the recognized person's name) or identifying the person as unregistered (voice says "unknown").	15 Human-Computer Interaction (HCI) researchers from the University of Campinas, Brazil.
2	Gear Face Recognition (GFR) with a few improvements in the face framing audio instructions. In addition, the way to access the application was made easier.	23 graduate students from a Human Factors class in the University of Campinas.
3	A face recognition software that uses the Microsoft Kinect placed on top of a helmet to detect and recognize people in the surroundings of a visually impaired user. The system provides 3D auditory cues to indicate who someone is and where they are located in relation to the user. Besides the Kinect, the system also requires a laptop with high processing power to run the face recognition software. This laptop is placed inside a backpack so the user can carry it hands-free.	9 Human-Computer Interaction (HCI) researchers from the University of Campinas, Brazil.

At the end of Experiment 1, the set of 23 heuristics [4] was used during the debriefing session to discuss the design of the GFR system. Additionally, the heuristics themselves were discussed, so that we could figure out if their writing was clear, if they were understandable and if they actually made sense in the context of designing Natural User Interfaces (NUIs) applications. During this debriefing session, the participants reached a consensus regarding a grade for each heuristic; the scale used for the grade was the same proposed by Nielsen [6] for his usability heuristics: from 0 (not a problem) to 4 (meaning a catastrophic problem).

Regarding the GFR system, the main problem pointed out by the participants was that the audio cues to help the user in framing someone's face needed to be more helpful. Regarding the heuristics, participants suggested that the scale of grades could include, besides problems, a positive aspect, i.e., how much the system was in accordance with the heuristic. Additionally, some heuristics could be grouped together since they were understood as semantically similar.

As for Experiment 2, it followed these steps:

1. Students were divided into five groups of four or five participants. Each group was asked to elect a member to act as a blind user, and another to be "unknown" in the database. The remaining members of each group were then registered in the database: three pictures for each person, from different angles.
2. The participant elected to be blindfolded received instructions on how to operate the GFR system. A group of non-blind users was silently placed in front of her. Her task was to find and recognize all the people who are in front of her, assisted only by the GFR system.
3. The timer started counting and the blindfolded user accessed the GFR application. For each person she found, she must say aloud who she believed that person was, based solely on the feedback received from GFR. Timer stopped when user signalized she had achieved her goal.
4. Steps 2 and 3 were repeated four more times for a different blindfolded user and different groups of people to be recognized.

At the end of Experiment 2, the participants received the set of 23 heuristics [4] to analyze during the experiment. As an after class activity, they were asked to discuss the GFR in the context of the heuristics and reach a consensus for the grades of each heuristic. The applied scale of grades was the same as in Experiment 1, but it also included grading how much the system is adherent to the heuristic: from –1 (adheres the heuristic in a superficial manner) to –4 (completely adheres the heuristic). After the participants submitted their heuristic evaluations, a debriefing session was conducted. During this session, participants suggested that, for improving the heuristics, the option "not applicable" was included in the grading scale.

Finally, Experiment 3 followed these steps:

1. Verify which participants were and which were not registered in the database, since the registration process is time consuming and was made in advance.
2. One participant volunteered to act as a blind user. She received instructions on how the system works, and her main goal: finding and reaching a specific person amid a group of four people. The participant then was wearing the helmet, the backpack and was blindfolded.
3. In silence, other four participants were placed in front of the blindfolded user, and at least one of them was not registered in the database.
4. The timer started counting and the blindfolded user began walking towards the group of people to be recognized, moving her head sideways to scan the room. For each person she found, she had to say aloud who she believed that person was, based solely on the feedback received from the system. The timer stopped when the

user signalized she had achieved her goal (success), or when the time reached 2 min (fail).

5. Steps 3 and 4 were repeated once for the same blindfolded user and a different set of four individuals to be recognized.
6. Steps 2 to 5 were repeated eight more times with a different participant acting as the blind user.

After the experiment, a debriefing session was conducted. The participants discussed the experiment and the main problems found, trying to analyze them with the help of the set of 23 heuristics [4]. The heuristics themselves were also dis-cussed, aiming at to regroup and rewrite them to better support evaluation. To grade each heuristic, the participants had to reach a consensus using two concurrent scales: from 0 (not a problem) to 4 (catastrophic problem), and from –1 (follows the heuristic in a superficial manner) to –4 (completely follows the heuristic). Therefore, it is the same scale used in Experiment 2, but with the possibility of pointing out problems, and measure how much the system complies with the heuristic.

The main problems pointed out by the participants during the debriefing were the need for regrouping the heuristics, since many of them had very close meanings, and the need to change the grading scale, since having negative numbers representing something positive (following the heuristic) is counter-intuitive. Therefore, the suggested grading would represent the level of compliance with a heuristic and would range from –4 (does not follow the heuristic at all) to 4 (follows the heuristic completely). In this case, 0 would be a neutral evaluation, i.e. there are no indications of neither problems nor heuristic compliance.

3 The NUI Heuristics Revisited

The previous section described the use of the heuristics for NUI [4] in three different experiments; each experiment pointed out to improvements that were necessary in order to make the heuristics more understandable and useful. In this section we will present the regrouping and, in some cases, rewriting of the 23 original heuristics. First, we show our criteria in evaluating the need for change in a heuristic. Then, we will give a general view of before and after. Finally, we will present the new heuristics in detail, with practical examples of use.

3.1 Change Criteria

The changes in the heuristics were based on both quantitative and qualitative analysis of the experiments results. The quantitative analysis come from Experiments 1 and 2; since both experiments tested the same system but with distinct groups of participants, we decided to compare the grades from these experiments. Hence, we placed the grades from the HCI researchers (one grade for each heuristic) in a table, along with the grades from the Human Factors students (one grade for each group, five in total). Additionally, we colored the grades in a scale of gray: the smaller the number (i.e., the more the system followed the heuristic), the whiter the table cell; conversely, the higher the

number (i.e., the more critical a problem was), the darker the cell. The result is in Fig. 1, where the grades of the HCI researchers are the bottom line of each table. It is important to note that the heuristics regarding "Multiple Users" are not shown in Fig. 1 because the tested system is not in that category.

Our main goal with the comparison in Fig. 1 was analyzing the interpretations given for each heuristic by finding divergence or convergence in the grades. This way, a column with a predominant tone of color (clear or dark) shows convergence in the grades, suggesting the heuristic had homogeneous interpretation. Likewise, a column with no color tone predominance indicates that there was divergence in the heuristics interpretation, suggesting possible problems with its writing.

Interaction								
Operation Modes	"Interactability"	Accuracy	Responsiveness	Identity	Metaphor Coherence	Distinction	Comfort	Device-Task Compatibility
-1	0	3	-2	-3	-4	-3	3	-3
0	0	3	2	-3	-3	-2	3	-3
0	0	3	2	3	-3	-2	4	-3
2	3	3	2	3	-2	0	4	-1
3	3	4	2	3	0	0	4	0
3	3	2	1	2	0	0	2	3

Navigation				User Adoption					
Guidance	Wayfinding	Active Exploration	Space	Engagement	Competition	Affordability	Familiarity	Social Acceptance	Learnability
-4	-2	-3	2	-4	-3	-2	-3	-1	-3
0	0	-2	2	0	-2	-2	-2	0	-3
0	0	0	2	0	-2	0	-2	0	-3
0	0	0	2	0	0	2	-2	0	-2
3	2	3	3	0	0	3	-2	3	0
2	3	3	0	2	0	2	1	0	0

Fig. 1. Specialists' evaluations in Experiments 1 (bottom row) and 2 (first five rows)

The qualitative analysis draws on the comments, suggestions and discussions from the debriefings of the three experiments. These data allowed deeper insights into how the specialists actually understood each heuristic, often corroborating with the quantitative data and possibly providing a reason for the divergence in interpretations. Some examples of this will be given in Sect. 3.3.

3.2   Before and After

Prior to detailing the new set of heuristics Fig. 2 illustrates the process of change. As shown in Fig. 2, the two heuristics Accuracy and Responsiveness (indicated by the number 1 in the image), were removed. Although both the quantitative and the qualitative analysis did not suggest any confusion related to the interpretation of these heuristics, in all three experiments they pointed to problems regarding algorithmic and technological issues only. For instance, in Experiment 2 many of the participants reported lack of precision in the face recognition software for the Accuracy heuristics

(hence, the dark tone of its column in Fig. 1). Conversely, in Responsiveness, they reported delays in the audio feedback provided by the system.
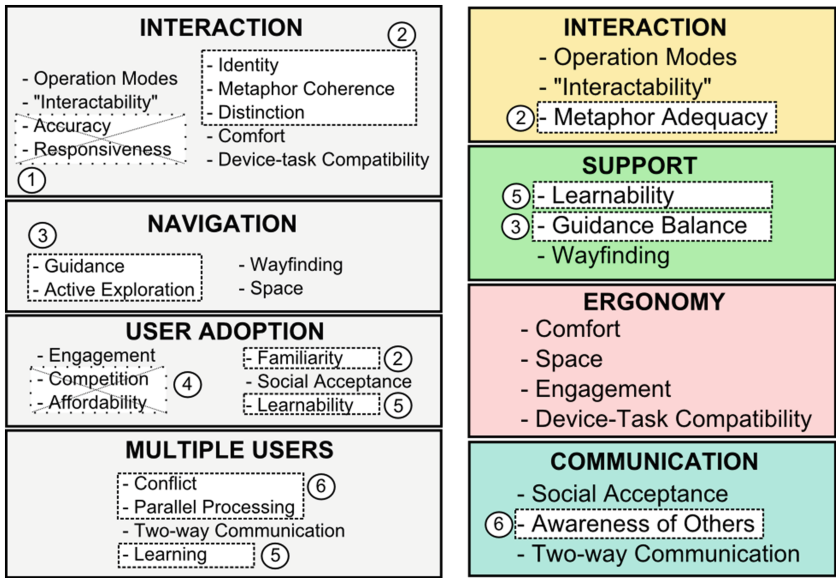

Fig. 2. To the left, the original set of heuristics and, to the right, the new set of heuristics

The four heuristics indicated by the number 2 in Fig. 2 (Identity, Metaphor Coherence, Distinction and Familiarity) were grouped together mainly because, during every debriefing, there was a clear confusion regarding their difference in meaning. Looking at Fig. 1, we can see that these heuristics seem to have the same scores, except for Identity, which seemed to be the representative of the system's interaction metaphor problems. However, the qualitative data shows that the four heuristics were used to analyze the same aspect (interaction metaphors). Therefore, they were grouped into one heuristic, Metaphor Adequacy.

Figure 2 also shows the grouping of the heuristics Guidance and Active Exploration (marked by the number 3) into one called Guidance Balance. Figure 1 suggests they both had similar scores, and the qualitative analysis reveals that both heuristics focus on the learning curve and on the balance between expert and novice users. The analysis of the HCI researchers for the Active Exploration heuristics even reads "as pointed in Guidance, there is free exploration of the system".

The number 4 in Fig. 2 points to the exclusion of the two heuristics Affordability and Competition. Figure 1 shows some divergence in Affordability, but that was because of the different views the participants had on how affordable the system was. Furthermore, the qualitative data showed that these two heuristics pointed to problems related to costs, market and technology.

The number 5 in Fig. 2 shows that the two heuristics Learnability and Learning were grouped as Learnability. Although originally one was meant for every type of system and the other was specific for interfaces with simultaneous multiple users, the way they were written was semantically very close. Furthermore, the quantitative and the qualitative data show that the users fully understood the heuristic.

Finally, number 6 indicates the grouping of two heuristics (Conflict and Parallel Processing) from the Multiple Users major group. Although we did not have experimental data about them, closer inspection reveals they were semantically close, becoming the heuristic Awareness of Others.

In summary, from the original set of 23 heuristics, based on the experimental quantitative and qualitative data regarding its use, a new set of 13 heuristics was generated. It is important to note that the changes made were either removing or grouping heuristics together; no new heuristics were added.

3.3   The Heuristics in Detail

This subsection presents the details of each one of the 13 NUI heuristics, within the following format: number, name, description and example of use. The descriptions were based on both the original description from Maike et al. [4] and on the analysis of the experimental data.

[NH1] Operation Modes. The system must provide different operation modes (visual, auditory, tactile, gestural, voice-based, etc.). In addition, the system must provide an explicit way for the user to switch between the modes, offering a smooth transition.

Example of Use: For the system tested in Experiments 1 and 2, the operation modes were: voice command (to run the application), pressing the smartwatch's physical button (also to run the application), dragging the screen (to close the application) and moving the arm to point the camera and frame someone's face. The evaluation for this heuristic pointed to problems related mostly to the transition between the modes. The experts concluded that the modes were competing with each other, since there was a delay to open the application, there was no sound feedback to inform the successful closing of the application, and the framing with the arm movement was difficult.

[NH2] "Interactability". In the system, the selectable and the "interactable" objects should be explicit and allow both their temporary and permanent selection.

Example of Use: In Experiments 1 and 2, participants pointed as "interactable" objects, the smartwatch's physical button, its camera and its screen. In Experiment 3, the HCI researchers said the people in front of the Kinect were the "interactable" objects.

[NH3] Metaphor Adequacy. The sets of interaction metaphors the system provides should make sense as a whole, so that it is possible to understand what the system can and cannot interpret. When applicable, there should be a visual grouping of semantic similar commands. In addition, the interaction metaphors should have a clear relationship with the functionalities they execute, requiring from the user a reduced mental load and providing a sense of familiarity. Finally, the metaphors should not be too similar to one another, to avoid confusion and facilitate recognition.

Example of Use: In Experiments 1 and 2, one of the interaction metaphors was the visual feedback the system provided while framing a person's face. When a face was detected, the system placed a rectangle around it and a voice said "framing". This audio clue did not translate completely the metaphor of the rectangle, which represents the focus functionality of a digital camera, which usually displays a rectangle in the screen to say that the image focus is being adjusted. Additionally, the evaluations also pointed that since the system is embedded in a smartwatch, a device that resembles a normal wristwatch, there is a natural sense of familiarity in using the system.

[NH4] Learnability. There has to be coherence between learning time and frequency of use. Therefore, if the task is performed frequently then it is acceptable to require some learning time; otherwise, the interface should be usable without much learning effort. In addition, the design must consider that users learn from each other by copying when they work together, so it is important to allow them to be aware of each other's actions and intentions.

Example of Use: In Experiment 1, the same person acted as a blind user more than once. This allowed us to measure the execution time of each iteration, and the results [2] showed that this time greatly decreased after the first round. Therefore, the system was easy to learn after a few minutes of use.

[NH5] Guidance Balance. There has to be a balance between exploration and guidance, to maintain a flow of interaction to both the expert and the novice users. To enhance transition from novice to expert usage, active exploration of the set of interaction metaphors should be encouraged by the system. Finally, it is important to provide shortcuts for the expert users.

Example of Use: The system tested in Experiments 1 and 2 provided both visual (rectangle around a face) and auditory guides (voice saying "framing", the name of the recognized person or "unknown"). In this sense, the user has freedom to explore freely, but to achieve her goal she will have to follow these feedbacks. In addition, the differentiation between novice and expert users is in how they interpret the feedbacks. For instance, it takes some time to understand that when the system says "framing" it is necessary to keep the arm still, so the system can finalize the recognition.

[NH6] Wayfinding. At any time, users should be able to know where they are from a big picture perspective and from a microscopic perception. This is important regardless of user proficiency with the system, i.e., novice and expert users need both views of the system.

Example of Use: In Experiments 1 and 2, the big picture perspective is the search for faces to scan, which also involves knowing how many people are in the environment and how big it is. The microscopic perception is the framing of one person's face, to find out who she is. In this sense, the feedbacks the system offers are more helpful to the microscopic perception than to the big picture.

[NH7] Comfort. Interacting with the system should not require much effort from the user and should not cause fatigue.

Example of Use: The system tested in Experiments 1 and 2, with the smartwatch, received several negative evaluations from the experts due to fatigue caused by keeping the arm raised for a long period. They noted, however, that there were the issues of lack

of practice from the users and low compatibility between the experiment and the real use. In contrast, the system tested in Experiment 3, with the Kinect, did not present physical discomforts, neither from the helmet nor from the backpack.

[NH8] Space. The location where the system is expected to be used must be appropriate for the kinds of interactions it requires and for the number of simultaneous users it supports.

Example of Use: In Experiment 3, a problem that happened many times was that, when the blindfolded user came too close to someone (around 60 cm), the system would stop detecting that person. In this sense, to fully comply with the heuristic the system would have to emit a warning before the user left the ideal distance from the person (which is around 1,20 m).

[NH9] Engagement. The system should provide immersion during the interaction, at the same time allowing for easy information acquiring and integration.

Example of Use: For the system tested in Experiments 1 and 2, the task of framing people's faces and finding out who they are could be more fun once the fatigue issue is resolved.

[NH10] Device-Task Compatibility. The system has to offer kinds of interactions that are compatible with the task for which it is going to be used.

Example of Use: In Experiment 3, the task of using the Kinect for locating and recognizing people proved to be very compatible with this device, given the lack of comfort issues and satisfactory success rates. In literature, however, there are examples of bad task compatibility for the Kinect, such as those reported by Cox et al. [3] who used it as a mouse cursor to select objects in a screen. This way, the user had to keep the arm raised and control the cursor by moving the arm or the hand. In this case, compared to other devices the authors found the Kinect presented high fatigue, low efficiency and high error rates.

[NH11] Social Acceptance. Using the system should not cause embarrassment to the users.

Example of Use: For the system tested in Experiments 1 and 2, participants pointed out that the smartwatch should not cause embarrassment because it is very similar to a regular wristwatch. In fact, they noted that, given its novelty and cost, it can be seen as a symbol of status.

[NH12] Awareness of Others. If the system supports multiple users working in the same task at the same time, then it should handle and prevent conflicting inputs. Therefore, users must be able to work in parallel without disturbing each other, but having awareness of the others.

Example of Use: Nebe et al. [5] present the multi-touch table they have built and how it is used in a scenario of disaster control management. In that case study, multiple users work simultaneously on a map displayed on the table. Each user can have their own tangible object (a puck) to interact with the map. Placing the puck on the map can zoom in, make markings on the map or create a personal window for the user on the screen, so each person can execute their own tasks in parallel without disturbing the group view of the map.

[NH13] Two-way Communication. If multiple users are working on different activities through the same interface, and are not necessarily in the same vicinity, the system must provide ways for both sides to communicate with each other.

Example of Use: Yang et al. [12] present a study in which participants used a multi-touch screen interface to collaborate remotely. In a ludic activity, one participant shared what she was doing in the multi-touch screen and a group of other participants, in a remote location, had to guess what was the task being executed. The participants in the remote location could not communicate back to the person performing the task, so one of the reported results was that participants wished they could do that through the system's interface.

4   Conclusions

Natural User Interfaces (NUIs) represent a strong tendency for new computer systems, as well as a challenge for designers, since delivering the promised feelings of naturalness is not trivial. In this paper, we showed three practical experiments using a set of 23 NUI heuristics as a tool for evaluating the design of two distinct assistive technology systems. During the experiments, participants also evaluated the heuristics themselves. Results of the experiments led to a leaner set of 13 NUI heuristics, with a compliance level ranging from –4 to 4.

This new set of heuristics is result of revisiting the previous set with both quantitative and qualitative analysis. Since two experiments tested the same system but with completely different groups of participants, we were able to look for divergences in the interpretations of the heuristics, so we could find the ones that needed to be rewritten or regrouped. This quantitative analysis was supported by the qualitative evaluation of the participants' justifications for their grades, which gave insight into how they understood each heuristic and, hence, what improvements were necessary.

Therefore, the experiments provided us both with the view of the heuristics in practice and with the opportunity to improve them. They also allowed us to enhance the description of each heuristic by providing an example of use taken straight from the experiments, whenever possible. Some heuristics were not applicable to the experiments, so we see as necessary future work applying the new set of heuristics in systems that support multiple users working simultaneously on the same interface. Additionally, we also believe further experiments and empirical uses of the heuristics will point to design principles that can be helpful in guiding designers in early stages of NUI applications design.

References

1. Bolton, J., Lambert, M., Lirette, D., Unsworth, B.: PaperDude: a virtual reality cycling exergame. In: Proceedings of Extended Abstracts on Human Factors in Computing Systems, CHI EA 2014, pp. 475–478. ACM, New York (2014)

2. Britto Neto, L.S., Maike, V.R.M.L., Koch, F.L., Baranauskas, M.C.C., Rocha, A.R., Goldenstein, S.K.: A wearable face recognition system built into a smartwatch and the visually impaired user. In: 17th International Conference on Enterprise Information Systems, ICEIS 2015, Barcelona (2015, to appear)

3. Cox, D., Wolford, J., Jensen, C., Beardsley, D.: An evaluation of game controllers and tab-lets as controllers for interactive tv applications. In: Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI 2012, p. 181. ACM Press, New York (2012)

4. Maike, V.R.M.L., de Sousa Britto Neto, L., Baranauskas, M.C.C., Goldenstein, S.K.: Seeing through the Kinect: a survey on heuristics for building natural user interfaces environments. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part I. LNCS, vol. 8513, pp. 407–418. Springer, Heidelberg (2014)

5. Nebe, K., Klompmaker, F., Jung, H., Fischer, H.: Exploiting new interaction techniques for disaster control management using multitouch-, tangible- and pen-based-interaction. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part II, HCII 2011. LNCS, vol. 6762, pp. 100–109. Springer, Heidelberg (2011)

6. Nielsen, J.: Usability Engineering. (1994)

7. Norman, D.: Natural user interfaces are not natural. Interactions **17**, 6–10 (2010)

8. O'Hara, K., Harper, R., Mentis, H., Sellen, A., Taylor, A.: On the naturalness of touchless: putting the "interaction" back into NUI. ACM Trans. Comput. Hum. Interact. **20**, 1–25 (2013)

9. Renzi, M., Vassos, S., Catarci, T., Kimani, S.: Touching notes: a gesture-based game for teaching music to children. In: Proceedings of the 9th International Conference on Tangible, Embedded, and Embodied Interaction, TEI 2015, pp. 603–606. ACM, New York (2015)

10. Ringland, K.E., Zalapa, R., Neal, M., Escobedo, L.E., Tentori, M.E., Hayes, G.R.: SensoryPaint: a natural user interface supporting sensory integration in children with neurodevelopmental disorders. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI EA 2014), pp. 1681–1686. ACM, New York (2014)

11. Wigdor, D., Wixon, D.: Brave NUI World: Designing Natural User Interfaces for Touch and Gesture. Morgan Kaufmann Publishers Inc., San Francisco (2011)

12. Yang, J., Dekker, A., Muhlberger, R., Viller, S.: Exploring virtual representations of physical artefacts in a multi-touch clothing design collaboration system. In: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group, OZCHI 2009, pp. 353–356. ACM, Melbourne (2009)