Centralized versus Decentralized
A Study on Tangible Control over Devices in the Living Room

M. Kwak

Eindhoven University of Technology
Department of Industrial Design

ABSTRACT

Technology is moving to the background and interoperability between devices increases. The handles for users to explore, make and break connections between devices seem to disappear in overly complex menu structures displayed on small screens.
Two prototypes have been developed that introduce a tangible approach towards exploring, making and breaking connections between devices in the living room. One provides a centralized approach (SCD1), the other a decentralized approach (SCD2).
Industrial Design students and graduates (N=12) have performed tasks and were asked to explain and grade one out of three methods: SCD1 (image 1), SCD2 (image 2)and bluetooth pairing.
Findings suggest that users are better able to project their mental model of how the system works on SCD2 and that a tangible solution is not necessarily a better one.

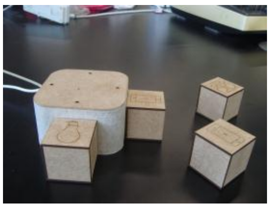


image 1 SCD1

image 2 SCD2

keywords **- tangible interaction; mental models.**

INTRODUCTION

Users are currently occupied with finding the right cables to connect devices and have to deal with cables that allow for connections that aren't possible. Even more, some possible connections never get explored, simply because cables don't allow for it.
Bluetooth technology solves part of the problem but introduces overly complex menu structures and devices without proper interfaces.

In 'the Internet of Things' [1] and 'Shaping Things' [2] a world is sketched in which each everyday object has an identity and is connected to the internet. In this world technology has moved to the background and interoperability between devices has increased.
Provided these devices are able to communicate with each other and to the user, this could mean the end of compatibility problems and the hassle of using cables.
It would also mean that users will have less physical and visual handles to make sense of their environments and the devices therein.
Design can play an important role in this sense-making with paradigms like TUI [3] that believe that physical handles for digital information provide users with more freedom and control.

The SOFIA project is a European research project that targets to *"make 'information' in the physical world available for smart services - connecting physical world with information world"*[4].
Within this project a Semantic Connections Demonstrator (SCD) has been developed at the department of Industrial Design of the Eindhoven University of Technology (ID TUE), by G. Niezen and B.J.J. v.d. Vlist.
It allows users to, tangibly, explore, make and break connections between devices in the living room.

A second SCD was developed to explore the possibilities of TUI.
Where the first SCD provides users with a centralized way of exploring, making and breaking connections, the second SCD provides users with a decentralized way.
In order to see which SCD would be best suited, a usability test was set up to answer the following questions:

- Are the SCDs a better* alternative, compared to the conventional method?
- Will the users be able to work equally well with both demonstrators?

* better in the sense that exploring, making and breaking connections is easier (performance) and more satisfactory (preference). An important aspect is the mental model the participants have and how close it is to the real model of the system.

BACKGROUND

Semantic Connections Demonstrator 1
The Semantic Connections Demonstrator 1 (SCD1)[5][6] consists of a center tile and cubes with labels that represent devices in the living room. The devices are: two mobile phones with music playing ability, a sound system and an Ambient Light lamp. The center tile detects when a cube is aligned to one of its sides and is able to recognize the cubes. Four LEDs, one on each top edge of the center tile, give the user feedback:

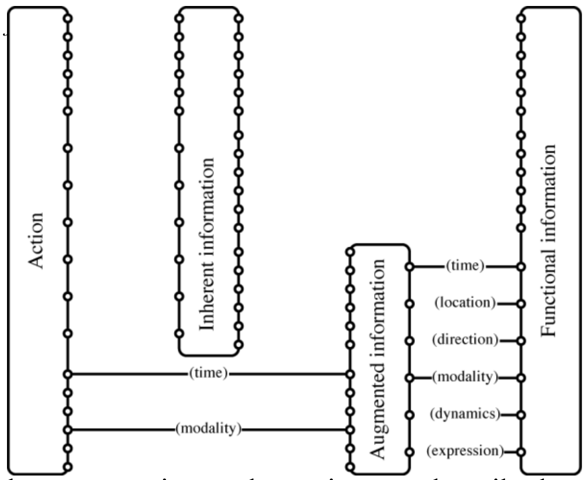
Red - No connection possible. This occurs when no relation is possible between two or more devices of which the cubes are aligned to the center tile. It also occurs when only one cube is aligned to the center tile.

Green - Connection exists. This occurs when two or more devices are connected and their cubes are aligned to the center tile.

Green pulse - Connection possible. This occurs when a relation is possible between two or more devices that are aligned to the center tile. This also occurs when the cubes of two or three connected devices are aligned to the center tile and a fourth is introduced with which a relation is possible.

To make or break a connection between devices, the cubes of these devices have to be aligned to the center tile, after which the center tile has to be picked up and shaken. After this it is not necessary to keep the cubes aligned to the center tile to maintain connections.

The Semantic Connections Demonstrator 1 is a centralized solution.



between action and reaction are described as:
Time - The product's reaction and the user's action coincide in time.
Location - The reaction of the product and the action of the user occur in the same location.
Direction - The direction or movement of the product's reaction (up/down, clockwise, right/left and towards/away) is coupled to the direction or the movement of the user's action.
Dynamics - The dynamics of reaction (position, speed, acceleration, force) is coupled to the dynamics of the action (position, speed, acceleration force).
Modality - The sensory modalities of the product's reaction are in harmony with the sensory modalities of the user's action.
Expression - The expression of the reaction is a reflection of the expression of the action.
Furthermore, Wensveen [7] distinguishes between three types of feedback and feedforward; functional, augmented and inherent.

Feedback is 'the return of information about the result of a process or activity' [American Heritage Dictionary]. Functional Feedback is *"the information generated by the system when performing its function"*. Augmented Feedback is information generated by an additional source, not directly related to the system and its function. Inherent Feedback was defined by Laurillard as *"Information provided as a natural consequence of making an action. It is feedback arising from the movement itself."*.
Feedforward is the information provided to the user before any action has taken place. Inherent Feedforward communicates what kind of action is possible and how one is able to carry this action out.
When an additional source communicates what kind of action is possible it is considered Augmented Feedforward. Functional Feedforward communicates the more general purpose of a product.

There are many improvements one can consider for the SCD when putting it in the Interaction Frogger framework. It was decided, though, to stay as close to the original concept as possible; for research purposes it is best to change as little as possible in order to be able to clearly identify what exactly causes change in user behavior (if users' behavior actually changes).
By removing the center tile and moving its functionality in the cubes, the SCD's functionality would increase. This would also allow for direct connection, removing the necessity for having to shake anything in order to make or break a connection.

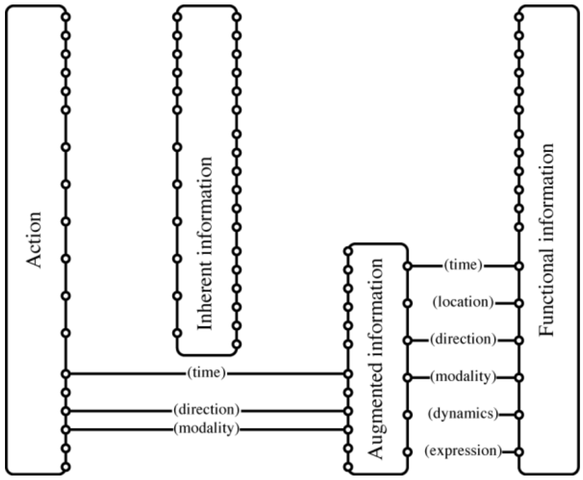


figure 2 The couplings of SCD2 in the Frogger Framework (remote)

When looking at the SCD as a remote control (figure 2), the changes improve SCD with regard to:
Direction - With the center tile removed, the direction of making and breaking connections (although done remotely) corresponds better.
Modality - With the shaking interaction removed, the modality of making and breaking connections corresponds better.

It is also possible to look at the SCD as the main product to interact with (figure 3), instead of the devices that can be connected with the SCD. This reveals more improvements:
Information about time, location, direction and modality are augmented and inherent when the center tile and shaking interaction are removed. For SCD1 only location is inherent, and time, direction and modality are augmented (figure 4).
Aside from the lack of expression and dynamics these changes result in a framework that is the same as the ideal frogger framework (figure 5). This should mean that the couplings are more meaningful.

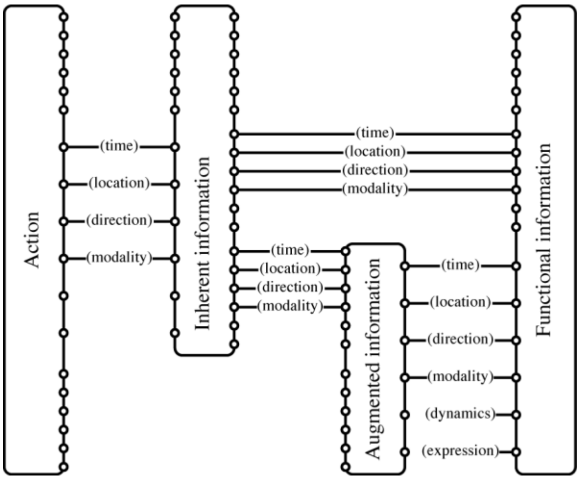


figure 3 The couplings of SCD2 in the Frogger Framework (main product)

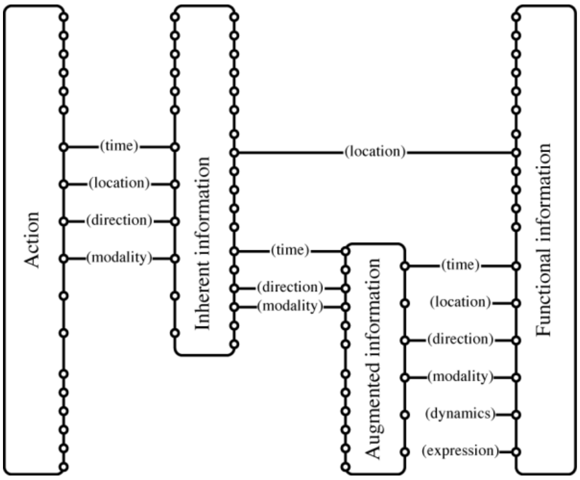


figure 4 The couplings of SCD2 in the Frogger Framework (main product)

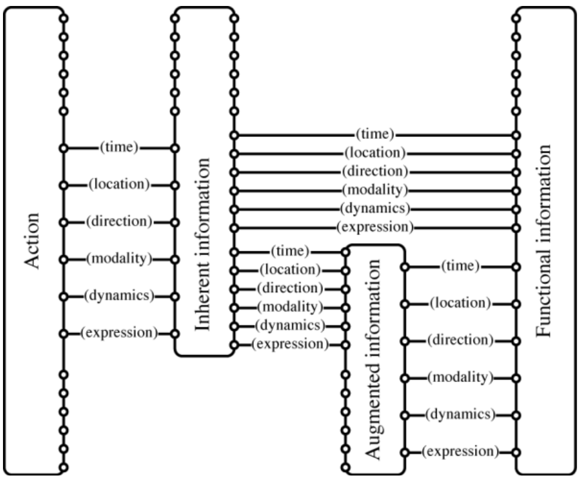


figure 5 'Through a combination of enrichting the action possibilities which exploit the human repertoir of actions and the inherent feedback based in the richness of the physical world the quality and number of possible meaningful couplings between action and function are increased'[7]

Inspired by Siftables [8] the cubes were transformed to tiles, because tiles have a clear top and bottom. This does still afford stacking, but hopefully users wo uld understand that they the tile were to be aligned.

Each side of a tile has an LED that gives the user feedback:

Red - No connection possible. This occurs when no relation is possible between the two devices of which the tiles are aligned.

Green - Connection exists. This occurs when a relation exists between two devices of which the tile are aligned.

To make a connection between devices, the tiles that represent these devices have to be aligned. To break the connection, the alignment between these tiles has to be broken.

The Semantic Connections Demonstrator 2 is a decentralized solution.

METHOD

In order to answer the questions raised in the beginning of this report, a usability test has been developed (see appendix 1). This was done in collaboration with a usability test expert, the project supervisor and the clients.

Participants

Ideally, participants would be recruited at the department of Architecture of the Eindhoven University of Technology (ARCH TUE). This would result in a homogenous group of participants, roughly the same age, with some understanding of technology and a sense of aesthetics. ARCH TUE also has an even distribution of men and women.

Due to lack of time, participants were recruited at ID TUE, in agreement with a usability test expert.

Initially, 18 participants were recruited. Due to unexpected failure of software, nine tests had to be cancelled and only three could be rescheduled. Of the 12 remaining participants 3 were women and 9 were man. The participants were between 21 and 26 years old. All but one had a BSc. in Industrial Design. One also had a MSc. in Industrial Design. Nine participants attended the MSc. program at ID TUE. Two participants attended the MSc. program at the department of Industrieel Ontwerpen (Industrial Design) of the Delft University of Technology (IO TUD).

Because of experience and familiarity with the subject, testing with these participants can have consequences. Most likely, the information gathered from the tests is biased and not representative for the target group of this project.

Design

The setup of the test is experimental, but includes two proven methods to gain insight in the participants' mental models and have the participant score the usability, respectively the 'Teach-Back protocol' [9] and the 'System Usability Scale'[10]. The action cycle by Norman [11] was also used to gain insight in the participants' mental models.

Added to these methods, the usability test tries to answer the questions raised in the beginning of this report by collecting performance data:

- user manual consults (if and how many)
- task completion time
- connection errors (if and how many)
- recovery from connection errors (time)
- correct explanation of scenarios
- correct description of envisioned performance to achieve a task

Apparatus

For the usability test the following devices were used:

- A Dell laptop (Windows XP) with wireless antenna, bluetooth antenna, audio out (3.5" jack plug), two USB-ports and the software package Eclipse installed to run code to be able to read RFID tags establish connections between the mobile phones, the laptop and the Ambient Light. The laptop, in combination with the Philips computer speaker set, also serves as the Sound System.
- A Nokia N95 mobile phone with Python installed, running a script to be able to play a sample and communicate with the laptop.
- A Nokia XpressMusic mobile phone with Python installed, running a script to be able to play a sample and communicate with the laptop.
- An Ambient Light lamp: A homebred bluetooth Arduino RGB LED lamp, with code running to be able to communicate with the laptop.
- A Samsung NV8 digital camera mounted on a tripod to film the usability test.
- A Philips computer speaker set with two satellites and a subwoofer.
- A Wireless router.

Process

Data was gathered about the usability* of the demonstrators in comparison to conventional methods of connecting devices, using Bluetooth pairing.

* usability can be divided in three aspects; efficiency, effectiveness and satisfaction.

Location and setup

A controlled setting was used to conduct the tests. The study took place in the 'Contextlab' at ID TUE. The 'Contextlab' is a room that is furnished to look like a living room, which is the context in which the SCDs would normally be used.
Participants explored, made and broke connections between two mobile phones (a Nokia N95 and a Nokia XpressMusic), a sound system and an Ambient Light lamp. This was done using SCD1, SCD2 and Bluetooth pairing.

Every session was recorded and notes were made by the moderator.

Methodology

This usability test was somewhat exploratory but also gathered assessment data about the efficiency, effectiveness and satisfaction of using the SCDs. Participants fell into three groups by the method they used to perform the tasks. Data was collected about task completion time, errors, recovery from errors and participants' satisfaction with using the method.

A between-subjects design was used.

In this study, each participant worked through four phases of tasks starting with one out of three methods (SCD1, SCD2 and Bluetooth pairing).

Bluetooth pairing was tested to set a benchmark to measure the usability of the SCDs.

Participants received a brief explanation (5 min.) before the test, outside the 'Contextlab'. They were guided through the task-path by the moderator. After all tasks were completed the participants were asked to fill in a questionnaire and a brief discussion (10 min.) took place between the participant and the moderator.

<u>Pre-test arrangements</u>
Participants:
- Filled in a pre-test questionnaire (see appendix 2).
- Reviewed and signed informed consent form (see appendix 3).

<u>Introduction to the session (5 minutes)</u>
Discussed:
- Participant's experience with usability studies and focus groups
- Importance of their involvement in the study
- Moderator's role.
- Room configuration, recording systems, observers, etc.
- The protocol for the rest of the session

<u>Tasks (30 minutes, including intermediary discussions)</u>
The task-path for each method (SCD1, SCD2 and Bluetooth pairing) looked like this:
- First, users were introduced to the method and given three task-descriptions. For each description the were asked to connect the devices/configure the SCD to perform the tasks. *9 minutes*
- Second, users were given one task-descriptions. They were asked to fill in an Action Cycle diagram (see appendix 4). *6 minutes*
- Third, users were presented with three scenarios. For each scenario they were asked to explain which connections there were. *9 minutes*
- Fourth, users were asked to explain what the method was they had used and how it worked using the teach-back protocol (see appendix 5). *6 minutes*

The order of tasks was random but the same for each participant.

<u>Post-test questionnaire (5 minutes)</u>
- Participants filled in the SUS questionnaire to rate the satisfaction of using the method (see appendix 6).

<u>Post-test discussion (5 minutes)</u>
- The moderator followed up on any particular problems that came up for the participant.

Measures
To answer the questions raised in the beginning of this report both performance and preference data was collected during the test sessions.

Performance:
- task completion time
- connection errors (if and how many)
- recovery from connection errors (time)
- correct explanation of scenarios
- correct description of envisioned performance to achieve task

Preference:

- Satisfaction of using a method to perform the tasks
- Appropriateness of method to perform the tasks
- Ease of use overall

Moderator role
The moderator sat in the room with the participant while conducting the session. The moderator introduced the session, conducted a short background interview, and then introduced tasks as appropriate. Because this study is somewhat exploratory, the moderator sometimes asked unscripted follow-up questions to clarify participants' behavior and expectations. The moderator also took notes and recorded the participants' behavior and comments.

The session was digitally recorded on video using a Samsung NV8 digital camera.

RESULTS

Performance data measurements
Due to a failure with the software used, both the tests with SCD1 and Bluetooth pairing took longer than expected. The system had to be reset more than once during a single test and in some cases a complete restart was unavoidable.
Because of these set-backs it was not possible to accurately measure any performance data; Tasks took longer than necessary to complete, the system did not recognize all errors as such, the system saw errors where there were none and participants had to recover from errors they were not responsible for.

What can be said is that only the participants testing SCD2 consulted the manual to receive further explanation of the system's working. This was done during the first task when the participants had not yet fully utilized the system and no errors had occurred. In one occasion the participant consulted the manual twice.

Action Cycle Diagram
The participants clearly had problems with filling in the Action Cycle Diagram. Only a few descriptions correspond to the predefined description (see appendix 4). This can be explained by the fact that people don't consciously think about the seven steps as defined by Norman [11], during everyday activity. It is also not uncommon to go through several cycles before a goal is reached and not all of these cycles have to include all seven steps. This would require participants to fill in several diagrams or include several cycles in one diagram. Because this issue didn't surface during the pilots or the first test, it would have been incorrect to change the procedure.
All participants followed roughly the same steps in achieving their goal. All but one participant forgot to mention the breaking of existing connections in 'Action specifications'. The participants of SCD2 and Bluetooth noticed this during the execution (before they thought they had achieved the goal) and went through an iteration immediately. Of the participants of SCD1, all but one participant noticed this after the execution (after they thought they had achieved the goal). These participants went through an iteration at a later stage but were also able to achieve their goal.

Teach-Back protocol
While it is possible to draw conclusions concerning the actual mental models of participants, the protocol was mainly used to see if there were notable differences between the methods.
Although there were some differences between the participants individually, amongst the methods the drawings and explanations were roughly the same. None of the participants went into details about what happened in the background, but instead focussed on the matters 'at hand'. Three participants (2 for SCD1 and 1 for Bluetooth pairing) mentioned extending the current system with more devices (more mobile phones and a TV). One participant (SCD1) was able to conclude that the connected devices were networked, the rest

explained the connections in a hierarchical way.

In one of the examples given in [9] the researchers were able to conclude that participants tend to draw little when the system is transparent. If it is less transparent they are likely to make more detailed drawings to better support their story.
In this test the level of detail amongst the methods was roughly the same.

System Usability Scale
Due to the setbacks both the tests with SCD1 and Bluetooth pairing took longer than expected. Added to that, the participants consciously experienced the setbacks. Combined, this almost certainly influenced the grades given by participants in the SUS questionnaire. Therefore the SUS scores are not reliable.
Nevertheless a t-test was considered. While a t-test is usually not meant for methods like SUS, because the measurements of the outcomes aren't on an interval or ratio level, it is often done without being objected. Unfortunately, the N was too little to perform a proper t-test and no significant deviation was found.
Instead a randomization test for two independent samples was done, but this also didn't show a significant deviation.

Observations and post-test discussion
Because the quantitative part of the test was unreliable, the focus of the results shifted towards the qualitative part.
Throughout the test, none of the SCD2 participants had trouble working with that method. During the post-test discussion they only wondered what was happening in the background. This was not because they hadn't been able to perform certain tasks but because they suspected more was going on than visible to the user.
Throughout the test, none of the Bluetooth pairing participants had trouble working with that method. They all mentioned that they were familiar with this way of connecting devices but had never experienced Bluetooth working this well.
The only real trouble for the participants working with SCD1 was the initial experience with that method. It was not clear what the relation was between the center tile and the cubes and all four interpreted the puling green LED as a 'working connection'. One participant initially thought the LEDs were lasers which could 'read' the cubes when placed on top. Another participant thought it was only necessary to align the 'main' device to the center tile and align the other devices to the 'main' device.
During observations and post-test discussions it became clear that all but one participant were not able to get from the method that the connected devices were networked. The tasks given and the methods at hand led them to conclude that the connections were hierarchical and participants mainly followed one of two modes of arranging connections:
Linear (from one device to the next) - This was seen with SCD1 and SCD2.
Centralized (from one device outwards) - This was seen with SCD2 and Bluetooth pairing.
Some participants sporadically arranged connections with SCD1 in a way that indicated they took it for a network, but they explained verbally that they expected the system to make a hierarchy out of their arrangement. Some participants also explicitly mentioned that certain connections should not be possible while in fact they were.

DISCUSSION

Because of the setbacks, this discussion will also focus on the actual usability test.

Results
The most striking results came from the observations and post-test discussions with the participants. The fact that all but one thought and worked in hierarchies is an interesting one. SCD1 was designed to

convey a different way of thinking, but instead participants projected their hierarchical way of thinking on the method. By making connections between no more than two devices at a time they did not use the full capacity of the system, took longer to perform the tasks and were slightly annoyed by the ‘extra’ work. Also, for those who thought in centralized hierarchies (one device in the center, the others around it) there was no way of projecting this thought in SCD1.
This is where the power of SCD2 shows, because it allows any way of thinking (hierarchical, ontological, linear, centralized). The participants found meaning in the arrangement of the tiles and the location of the tiles in relation to each other. For the system this doesn’t matter; a connection is a connection and if devices are connected, they are networked.
This leads to conclude that SCD2 is better fit for the job. Because of the setbacks, it is not possible to say whether SCD1 and SCD2 are better than Bluetooth pairing, although it appears that SCD2 is. It also appeared that participants were better able to perform the tasks with Bluetooth than with SCD1 but this can be attributed to the fact that they had experience with Bluetooth pairing and connecting devices using a GUI.

For further research it would be interesting to see whether the hierarchical thinking of people can be generalized to scenarios other than the ones used in this usability test. This could include other or more devices and media or even completely different contexts. If it can be generalized, an interesting question would be whether solutions like the SCDs should allow for hierarchical thinking while working with ontologies, or not.

Usability test
What became clear during the test, when the first problems surfaced, is that weeks of preparation mean little when there aren’t enough people and skills to get through the test without noticeable problems. A usability test like the one presented in this report requires at least three people to be present during the tests; someone to manage the soft and hardware, someone to guide the participant through the test and someone to make detailed notes.
For the test to be successful, more participants are also required. Six participants for each method is limited, four even more so. It is not possible to collect reliable data with this number of participants. Added to that, not all the methods used the usability test were as relevant as expected. While the ‘speak-out-loud’ step of the Action Cycle diagram is really useful to get insight in what participants think when they perform tasks, the other steps are often unclear to them.
With more experience, it might also be possible to describe and understand the mental model of the participants, using the Teach-back protocol. It still was useful for this test to see that all methods equally provided the participants with information, but the full potential of the protocol was not utilized.

If this usability test were to be reproduced at a later moment, the advice would be to have at least three people included in the setup, execution and completion of the test.
For a more qualitative approach, the fourth step of the Action Cycle diagram (think-out-loud) could be considered for each task.
For a more quantitative approach, the Action Cycle diagram could be removed from the test completely, as well as the Teach-back protocol. This would allow for more tasks to be performed, which results in more data to be analyzed. A more elaborate usability questionnaire could be considered, although one has to take into account that lengthy questionnaires tire participants. Especially if combined with performing tasks, this could lead to participants not paying enough attention when answering the questions.

REFERENCES

1. Kranenburg, R. (2008). *The Internet of Things: A critique of ambient technology and the all-seeing network of RFID*. Institute of Network Cultures.

2. Sterling, B. (2005). *Shaping Things.* MIT Press.

3. Fitzmaurice, G., Ishii, H., Buxton, W.(1995). Bricks: Laying the foundations for Graspable User Interfaces. *Proc. of CHI'95*, pp. 442-449.

4. http://www.sofia-project.eu/

5. Vlist, B.J.J. van der, Niezen, G., Hu, J., Feijs, L.M.G. (2010). Semantic connections: Exploring and manipulating connections in smart spaces. *First Workshop on Semantic Interoperability for Smart Spaces (SISS10)*.

6. Niezen, G., Vlist, B.J.J. van der, Hu, J., Feijs, L.M.G. (2010). From events to goals: Supporting semantic interaction in smart environments. *First Workshop on Semantic Interoperability for Smart Spaces (SISS10)*.

7. Wensveen, S.A.G., Djajadiningrat, J.P., Overbeeke, C.J. (2004). Interaction Frogger: a Design Framework to Couple Action and Function through Feedback and Feedforward. *DIS'04*.

8. Merrill, D., Kalanithi, J., Maes, P. (2007). Siftables: Towards Sensor Network User Interfaces. *TEI'07, Chapter 2,* pp. 75-78.

9. Veer, C. van der, Carmen Puerta Melguizo, M. del (2003). Mental Models. In J.A. Jacko & A. Sears (Eds.), *The Human-Computer Handbook. Fundamentals, Evolving Technologies and Emerging Applications* (pp. 52-80). Mahwah New Jersey: Lawrence Erlbaum Associates.

10. Brooke, J. (1996). SUS - A quick and dirty usability scale.

11. Norman, D. (2002). *The Design of Everyday Things.* MIT Press.