

Empirical Research Through Design

David V. Keyson * Miguel Bruns Alonso*

** ID StudioLab TU Delft
Delft, The Netherlands, d.v.keyson@tudelft.nl*

Abstract: This paper describes the empirical research through design method (ERDM), which differs from current approaches to research through design by enforcing the need for the designer, after a series of pilot prototype based studies, to a-priori develop a number of testable interaction design hypothesis which are then embedded in a working prototype and tested in context with target users. The approach builds on contextual enquiry methods such as context mapping leading to a verifiable working prototype while contributing to fundamental design knowledge. A case study is described to illustrate the application of the ERDM method.

Key words: *Research through design, design knowledge.*

1. Introduction

Research through design focuses on the role of the product prototype as an instrument of design knowledge enquiry. The prototype can evolve in degrees of granularity, from interactive mockups to fully functional prototypes, as a means to formulate, develop and validate design knowledge. The designer-researcher can begin to explore complex product interaction issues in a realistic user context and reflect back on the design process and decisions made based on actual user-interaction with the test prototype. Observations of how the prototype was experienced may be used to guide research through design as an iterative process, helping to evolve the product prototype.

Given the design challenge of developing products that may be complex in terms of function and context of use, research through design can provide a means to increase the external validity of a given design concept. However, though the research product prototype may develop into a highly desirable design through a series of rapid design iterations and validations, it is often difficult for the designer to reverse engineer the resulting final prototype and form a substantial contribution to design knowledge in terms of generalizable results. On the other hand controlled lab studies exploring a certain aspect of interaction, may lead to high internal validity but findings may be difficult to generalize and apply in context when not integrated as part of the total product experience.

A middle ground is proposed in this paper based upon ten years of research through design projects in which controlled empirical studies were integrated into working prototypes as a means to gain deeper insights into interaction issues and to contribute to design science. In this manner data-driven field findings are fed back into the theoretical model of interaction behind the prototype concepts. This approach is referred herein to as “empirical research through design”.

2. Empirical research through design

The key aspect of empirical research through design is to create experimental variability in the product prototype so as to formally test the underlying theoretical design questions at hand and in a real-world context. First, the prototype variability has to be carefully defined so as not to confound the research question at hand. For example, if the design researcher is interested in how speech versus pen input might influence interaction on a tablet PC, then the visual interface should appear identical in the speech and pen input conditions as well as operation sequences. Secondly, the design skills and creativity of the designer are critical in developing a balanced test platform. For example, users who may prefer speech input should not be steered towards pen input because of a poor speech input design. Thirdly, the designer-researcher engaging in research through design should be well informed on formal methods of statistics and research design techniques as common to the social sciences. In particular this knowledge should consider small subject numbers, a common phenomenon in design studies involving physical prototypes. Fourthly, the designer-researcher should strive to place a working prototype in context and test over time so as to increase the number of observations. And, finally the designer should collect data in an objective manner so as to not bias results.

In the following a description of the empirical research through design process is presented in more detail.

Finally, a case study involving the empirical research through design method is discussed below.

3. Applying the Empirical Research through Design Method

The empirical research through design method (ERDM) starts by formulating a hypothesis based on input from literature or inspirational tools such as context mapping (see figure 1). Continuously, various prototypes of artifacts or interfaces are developed, based on structured design principles, and are then tested based on the underlying hypotheses of interaction. Important findings are fed into a second design iteration to improve the artifacts or interfaces. This procedure can be repeated several times until the right prototype is developed by which the formularized hypothesis can be tested. The prototype is then used to run an experiment which from the user's perspective is a fully functional prototype running in context. If the research question and resulting prototype involves longitudinal usage issues it should be tested in the field, otherwise a context-simulation lab may suffice.

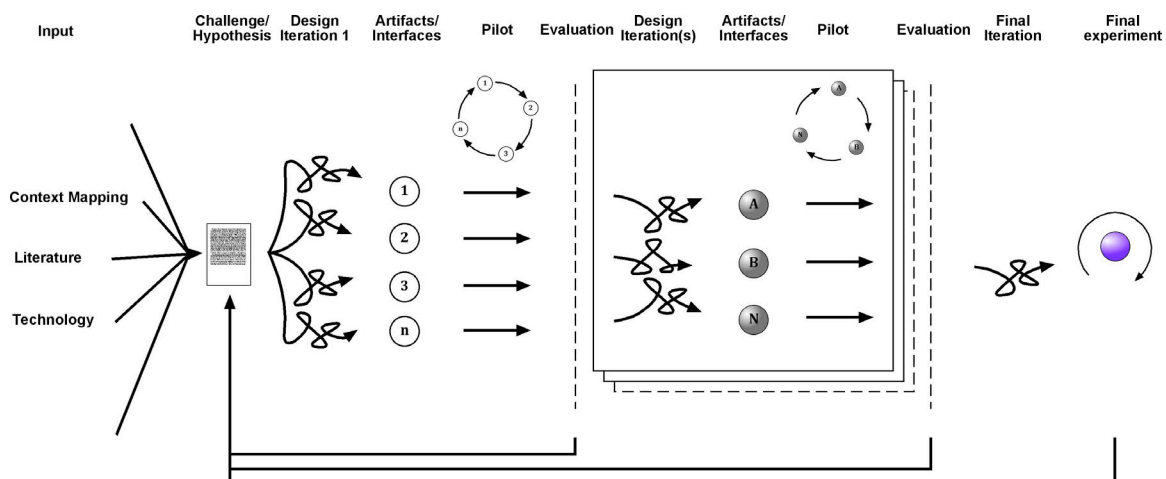


Figure 1. Empirical research through design process

3.1 Prototype variability

Prototypes in the ERDM serve as a means to empirically test underlying theoretical hypothesis, rather than acting mainly as a reflective mechanism or as a means to only organize user feedback. Typically the term “research through design” is used to refer to explorative design driven research [12], but lacks the formal manipulation, testing, and iterative development of design parameters. Furthermore, unlike Zimmerman et al. [12] ERDM delineates “Fuzzy front end” techniques as a means to gain deep user experience via techniques such as context mapping [9], with the actual process of creating a phenomenological or perceptual model which can then be developed into design dimensions and integrated and tested in-context, using the working prototype. The ERDM, though defined here for the first time, has been applied in a range of design projects, which began with the modeling phase. For example, Ross & Keyson [7] developed an atmosphere perception and projection model for living room environments, given that no formal perceptual model could be found for the experience of home atmospheres. In work by Rozendaal, Keyson & De Ridder [8] a model of perceived richness during human-product interaction was developed based on the parameters of engagement and control. Work by Vastenburg, Keyson, & de Ridder, focused on developing a phenomenological model of interpretability for in-home messages as a factor of message urgency and activity [10].

3.2 Balanced test platform

To formally test the design questions in a prototype, from the user’s perspective the product should appear very similar. Typically a within subject design is used to reduce the number of required subjects and to control for individual differences. Secondly an existing product may be used as a control condition. For example, in an experiment by Bruns & Keyson [1] it was hypothesized that interacting with music in an expressive and tangible way would increase the appeal of the interface. Therefore two different interfaces were compared one being a graphical interface, as implemented on the Apple iPod, and a second being a tangible expressive interface named MusicCube, a prototype which had been developed after a series of design iterations. Two variants of the MusciCube were utilized in the study, one providing speech feedback to indicate song name and mode, the other version, provided no explicit feedback, other than the actual song being played at a perceived volume level. The finishing and appearance of the prototype was of such high quality that naïve users experienced the MusicCube as an off-the-shelf product. The prototype was fully functional in terms of the proving all the basic features of a music player. The finishing allowed the prototype to compete on a functional, aesthetical and appeal level with a successful product such as the iPod.

3.3 Statistical Methods

In each of these studies techniques such as, semantic differential scaling and multiple analysis of variance (MANOVA), common to the social sciences were used to formally analyze the field data and draw conclusions, leading to improvements in the underlying phenomenological models. Semantic differential scaling provides a means to group several factors under a small number of dimensions explaining a high percentage (e.g. >70%) of the underling variance in the data derived from attitude or Likert scales. MANOVA has the strength that continuous data, which are derived from objective measures can be contrasted against or grouped with categorical data, such as results from quantitative surveys. Small in field studies may be limited in terms of large amounts of qualitative data, however objective data over time, such as the number of times a certain feature may been used could be substantial in terms of data points. In short, by examining the interactions between

categorical and continuous data, the shortcomings of small user number studies may be compensated to some degree. Other methods such as regression analysis and cluster analysis have also been applied in past studies involving ERDM.

3.4 Prototype in context

The test environment of a prototype in a laboratory setting may be constructed in such a way as to assist the user in imaging a certain real-world context of use. However some dimensions such as the acceptability of a product can often only be tested and explored in context. For example, product acceptability and harmony in relation to existing living patterns was explored by means of an asynchronous interactive communication system, which was designed based upon the metaphor of a bulletin board. End-users were actively involved in various stages of the design process and a pilot study with a working prototype was conducted in the field, showing a high degree of user acceptance [11]. Also aspects, such as time are important factors for testing in context. Prolonged experiments with interfaces or artifacts that need to be integrated in people's daily routines can only be tested in context. For example an interface to control and communicate peoples well-being throughout their daily routines [5] had to be carried by the participant for two weeks.

3.5 Objective observation

Research through design projects often involve integration of technology in the prototype design. Embedding sensors in these products allows gathering objective information of use. For example in the voicemail browser gestural aspects such as playing time, percentage of errors, and search time were recorded by the system [8]. But also when concepts are tested in a Wizard of Oz setting, which is often the case in the preliminary iterations of the research through design process, the experimenter should record objective measures in addition to the subjective data obtained from questionnaires.

4. ERDM Example: PauseBuddy

To illustrate how ERDM can build on insight generation work, the design case of Pause [6] is presented below. Pause Buddy is a prototype product, which was developed to reduce office stress by stimulating short breaks and social interaction pauses. To create novel interactive products in which context and social interaction play a central role in the user experience, industrial designers have elaborated upon various techniques. Combining these methods brings the design process from “fuzzy front end” to a working field prototype. A range of design techniques and research methods fed the ERDM, including literature study, probing, interviews, context mapping, and personas described by Keyson & Ottens [6]. The different techniques formalized the hypothesis that initiated the research through design process.

4.1 Hypothesis

The insight generation studies led to the hypothesis that to reduce office stress, the product had to: offer the user enough rest, provide social support, stimulate physical movement, and allow for sufficient control and overview. Based on the hypothesis, several ideas and solutions to the challenge were generated inspired by the results from the different research methods. Continuously the ideas were sketched and transformed into prototypes and improved throughout various design iterations, which were tested in a “Wizard of Oz” setup, as described below.

4.2 Iteration 1: Posture reflector

Each participant was asked to complete a Dutch version of the Spielberger State-Trait Anxiety Index (STAI), to quantify individual degree of anxiety proneness. In the first iteration a prototype named posture reflector was

build that reflected the posture of the participant by changing from a standing man to a collapsed man when the participant sat in the same position for a too long time. The posture reflector was red to attract attention and was placed on the desk of three participants. By means of ropes underneath the table the researcher manipulated the posture reflector. The researcher observed the different participants and recorded their behavior. When participants just sat working at their computer without moving, laughing, walking away, or chatting, the posture reflector collapsed. The experiment lasted for one working day and at the end of the day the participants filled out their experiences in a questionnaire.

Findings

Participants did not immediately notice when the posture reflector collapsed. This was attributed partly to the fact that the posture reflector was out of their sight, and partly because they were not paying attention to it. When they observed that the posture reflector collapsed, they changed their posture or they took a break. But no participant became more conscious of his or her posture through via product prototype.

Conclusions

The strengths of this prototype were its physical appearance and the physical feedback, through bending. The product did not attract enough attention due to its position on the desk. The prototype was not moveable, for which it was not possible to place it closer to the participants after the test had started. Therefore, besides bending, other means for attracting attention were considered thought to be needed.

4.3 Iteration 2: Pause Stimulator 1

The second iteration prototype was named pause stimulator. The pause stimulator reflected the pause behavior of people, and was meant to make people aware of taking enough pauses. When participants did not take pauses for a certain time the product would slowly turn from green to yellow and finally to red. After a long enough break the prototype turned green again. The pause stimulator made it possible to observe which colleagues did not take time to pause, because the products were placed on their desks one could see the colors, and thus possibly invite a colleague to take a break together. In short the Pause Stimulator, was thought to stimulate users to pause sufficiently and engage in social interaction with colleagues during the triggered breaks.

For the experiment three participants were given the pause stimulator to place on their desks. When a participant had been focused on working for 10 minutes, without talking to colleagues for example, its product turned to yellow and after another 10 minutes it turned to red. The researcher observed the participants and recorded their behavior by filling in an event table. The experiment lasted for one working day and at the end of the day the participants filled reported their experiences via a questionnaire.

Findings

The main finding from this pilot study was that participants did not feel obliged to take a break. On the contrast, the pause software that was intended as a prevention mechanism for repetitive strain injury often caused stress because the participants felt forced to take a break instead of being able to decide by themselves as to when to take a break. Furthermore, be required to take a break every 20 minutes appeared to be too frequent according to all the participants. Unfortunately, the prototype did not stimulate social interaction with colleagues, but more with the researcher, which could be attributed to not having the same schedule. Also the feedback nuances were considered to be too abrupt.

Conclusions

Having a physical representation was again considered to be very important as well as the fact that participants were not obliged to take a pause. The time between two pauses was considered too short and would need to be longer. Also the social interaction stimulated by this product needed to be improved.

4.4 Iteration 3: Pause stimulator 2

The second pause stimulator reflected the pause behavior of people. The difference with the previous pause stimulator was that the participants could choose a color by which other people would not be able to determine pause behavior. During the pilot study each participant was given a pause stimulator to place on her desk. She could choose a personal color for the relaxed mood and one for the pause mood. After sitting for 30 minutes the product changed to white, which was the in between phase. After another 30 minutes the product changed to the pause mood and was only switched to the relaxed mood after the participant stood up from her chair and left the work place for 5 minutes. Again the researcher observed the different participants and recorded their behavior. The experiment lasted for one working day and at the end of the day the participants filled out their experiences in a questionnaire.

Findings

This time there was an hour between two pauses, which seemed to be fine. The participants liked the unobtrusive way of feedback and did not feel obliged to pause. However the interface was considered to be very subtle, and could be thus overlooked. The participants did not mind whether their colleagues could see when they had to pause. The feedback was considered not to be personal, as it did not refer to the stress level, which was considered to be more personal.

Conclusions

The timing for pauses was better in the second pause stimulator. The product reminded the participant to take pauses. The possibility of choosing personal colors did not make a difference, since taking pauses is not related to a mood. Again, the physical appearance seemed to be important.

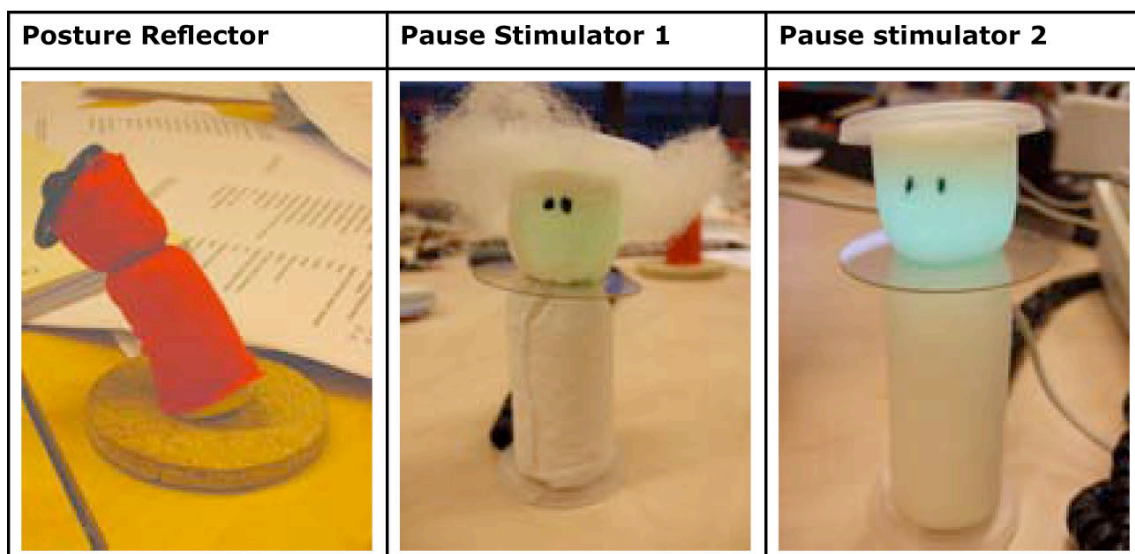


Figure 2. Design iterations for the Pause Buddy

4.5 Final design

The final prototype of the Pause Buddy was constructed as a round ball with LEDs, a small vibration motor, and limited range Bluetooth technology with signal strength detection to detect if another Pause Buddy is in the vicinity of 50 cm. The device shows one blinking red LED light to indicate when the user needs to take a pause of at least 3 minutes. The user could decide whether or not to pause alone or with someone else. If the user did not react to Pause Buddy after 15 minutes it would begin to vibrate and would display three blinking red lights. Based on a user-defined threshold the Pause Buddy indicated whether it was time to take a social interaction break via a yellow emitting light.

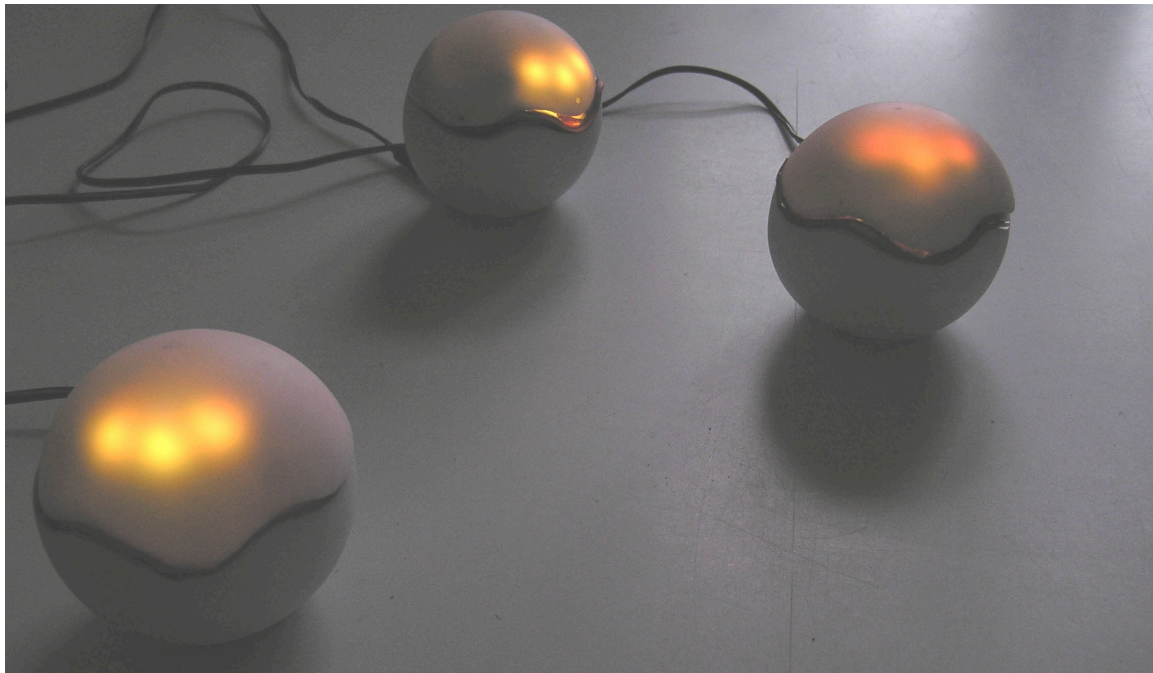


Figure 3. Final prototype Pause Buddy

4.6 Final evaluation

The Pause Buddy was tested within two companies, with three participants per company, over a period of four days. The Pause Buddy stimulated 51% of their pauses [6]. There was only one kind of pause cue, triggered by a blinking red LED, suggesting users to take a macro-pause, being a break longer than three minutes. The user could decide whether to pause alone or with someone else. There was no rule that users had to take a social pause, only that the user had to pause 7 times a day for at least three minutes and leave the desk chair. An ignore function was implemented when participants did not want to take a break.

Permission for remotely video recording the participants at their workstation using a web cam was asked for. Furthermore, a questionnaire containing 25 questions, on a seven point Likert scale, was included with the package. The scales were based on measures of Appeal [2], including ergonomic and hedonic factors and an overall impression dimension. As can be seen in Figure 4, subjective appeal ranged from slightly positive with mean rankings higher than 3.0 to very positive for factors such as Attractiveness and Sympathy. The lower scores relating to control, value and fragility could have been due to technical problems, such as the unit overheating at one of the sites, and the occasional loss of the Bluetooth signal.

Average interaction values of the Pause Buddy

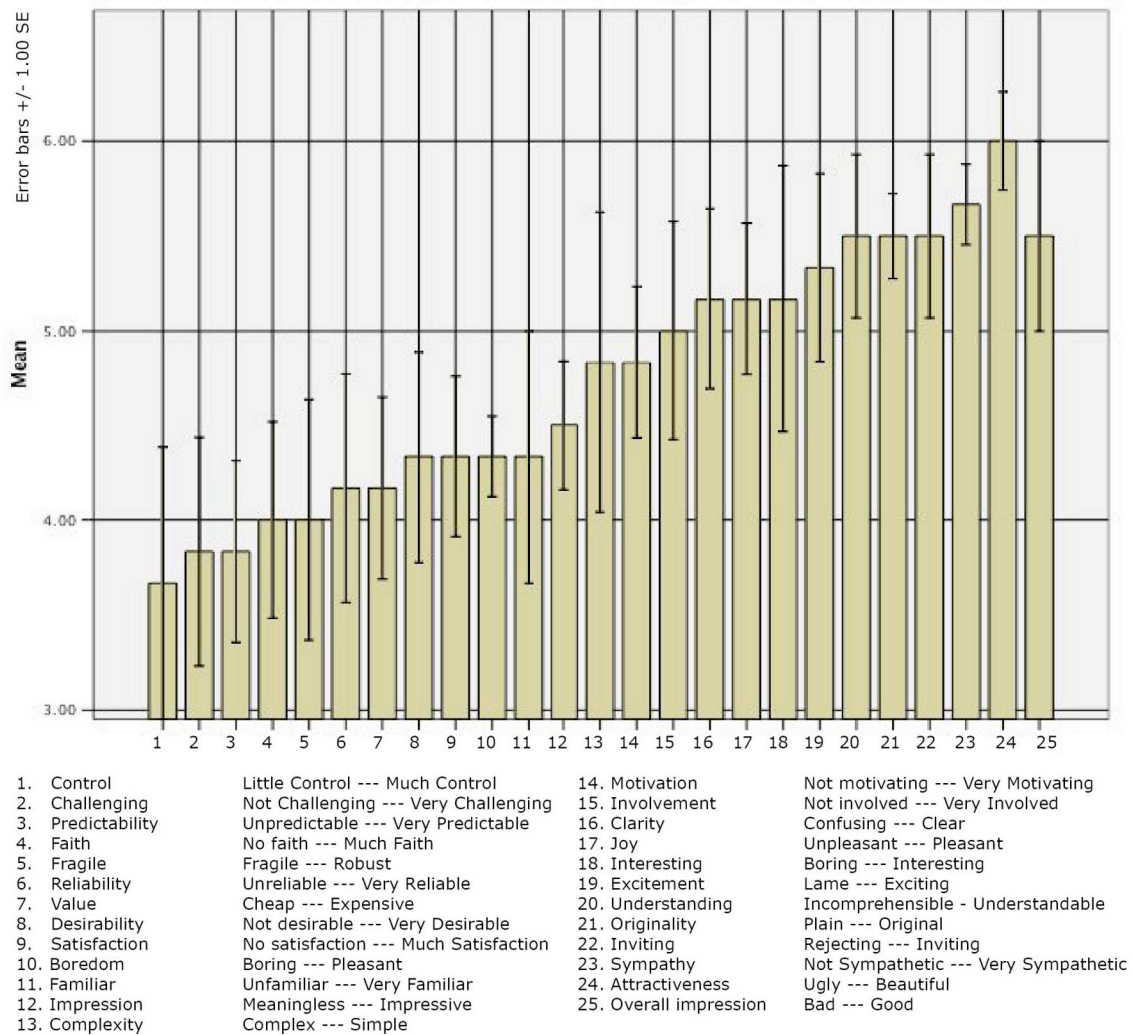


Figure 14: Results from subjective questionnaires for Pause Buddy appeal, including ergonomic and hedonic factors [17].

In 25% of the pause cases, Pause Buddy was taken along when the Pause Buddy indicated to the office workers that a social pause should be taken. Ten percent of the social pauses were taken without taking Pause Buddy along. In 65% of the pause cases, the participant left or “walked away” from the desk and took a pause. It cannot be ascertained if these were social pauses, however, 16% of these pauses occurred while Pause Buddy was flashing (see Figure 5).

Overview of the way of pausing

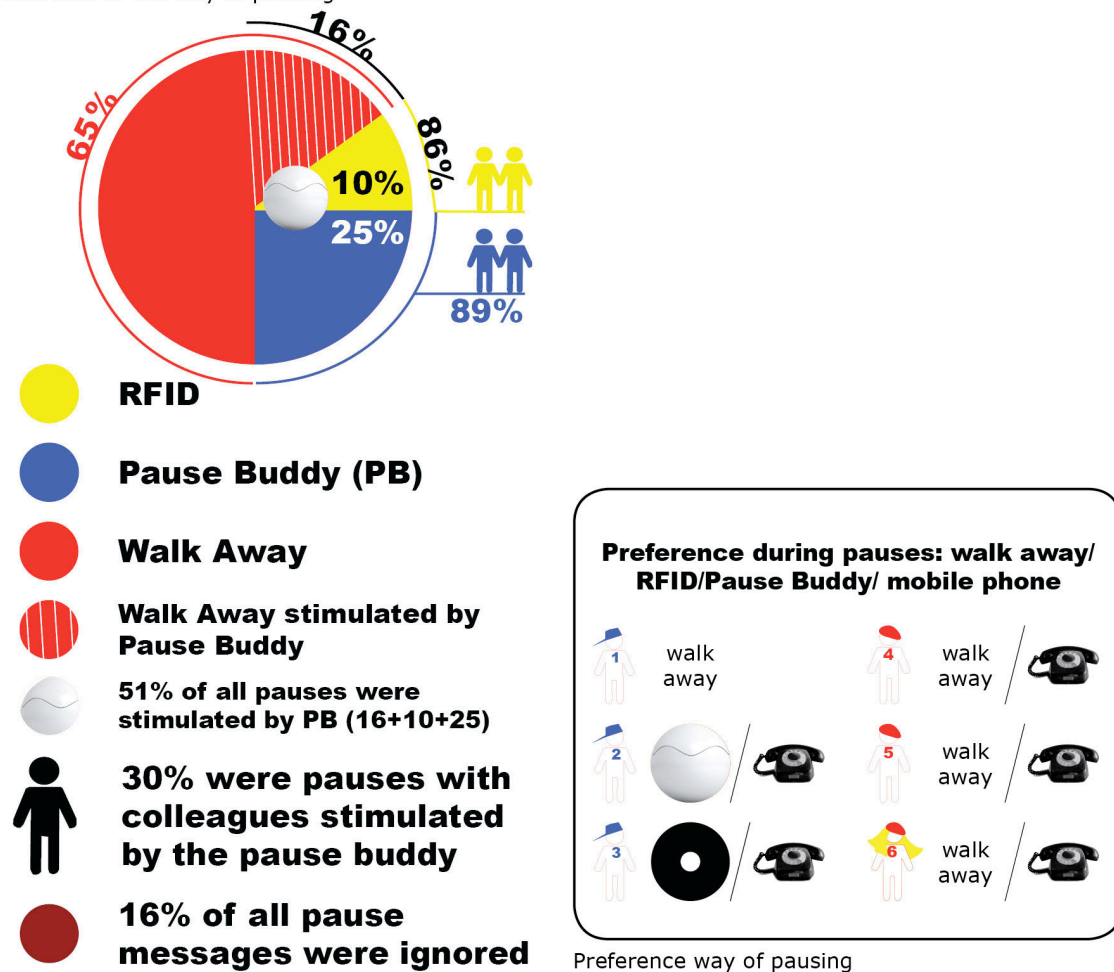


Figure 5. Percentages for each type of pause

5. Conclusions

The example of the Pause Buddy illustrated the application of the empirical research through design method. The different steps were taken within this project, starting by formulating the hypothesis based on knowledge obtained from literature and context mapping studies. Various design iterations explored the results obtained from these inspirational sources, by building and testing different prototypes, and yielded the final design of the Pause Buddy, which was tested and evaluated in a real-world context. To control the different prototypes a Wizard of Oz method was used. The researcher recorded objective measures on pausing behavior.

In the current experiment, the design researcher was able to attribute the observed field studies back to the design manipulations, thus validating the results and building on fundamental design knowledge. In the current case, EDRM led to new insights on how to design for and stimulate stress-reducing pauses in the workplace. Future work on EDRM will focus on the role of methods such as experience sampling [3] and the day reconstruction method [4] to support pilot work leading to the formal testing and development of design knowledge. In particular such methods enable the researcher to gain in-context user insights, which in turn can guide the research design process.

6. References

- [1] Bruns Alonso, M. and Keyson, D.V. (2005). MusicCube: Making Digital Music Tangible. In *Proceedings of CHI'05*, ACM Press pp 1176-1179.
- [2] Hassenzahl, M., Platz, A., Burmester, M. and Lehner, K. (2000) Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. *Proceedings of CHI'00*, ACM Press pp 201-208.
- [3] Hektner, J., Schmidt, J. A., and Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
- [4] Kahneman D., Krueger A.B., Schkade D.A., Schwarz N., Stone A.A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306, 1776.
- [5] Keyson D.V., Goverts S.T., Hoogh, de M.P.A.J., Varkevisser M. (2006). Monitoring and communicating stress via a mobile device. In *Proceedings of the 1st International Conference on Persuasive Technology for Human Well-Being*.
- [6] Keyson, D.V. and Ottens, H.J. (2009) Contextual Interaction Design: The case of Pause Buddy. In *Proceedings of AmI 2009* (in press).
- [7] Ross, P.R., and Keyson, D.V. (2007). The case of sculpting atmospheres: towards design principles for expressive tangible interaction in control of ambient systems. *Journal of Personal and Ubiquitous Computing* 11(2), pp 69-79.
- [8] Rozendaal, M.C., Keyson, D.V., and de Ridder, H. (2009). Product features and task effects on experienced richness, control and engagement in voicemail browsing. *Journal of Personal and Ubiquitous Computing* 13(5) pp 343-354.
- [9] Sleeswijk Visser, F., Stappers, P.J., Van der Lugt, R., and Sanders, E.B.N. (2005). Contextmapping: experiences from practice. *CoDesign: International Journal of CoCreation in Design and the Arts*, 1745-3755, Volume 1, Issue 2, pp 119-149.
- [10] Vastenburger, M.H., Keyson, D.V., de Ridder (2009) Aware Home Notification Systems: A Study of Cost - Benefit Factors in a Simulated Home. *International Journal of Human-Computer Studies* 67, pp 814-826.
- [11] Veldhoven van E.R, Vastenburger, M.H., Keyson, D.V.: Designing an Interactive Messaging and Reminder Display for Elderly. In *Proceedings of AmI 2008* pp 126-140.
- [12] Zimmerman, J., Forlizzi, J., Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of CHI'07*, ACM Press pp 493-502.