ROYAL INSTITUTE OF TECHNOLOGY

DEGREE PROJECT AT CSC, KTH 2010

DD143X

Wizard-of-Oz studies

Wizard-of-Oz studier

Elina Meier

870729-0667
Klasrovägen 41C
19149 Sollentuna
elinam@kth.se

Degree project in: Computer science

Supervisor: Johan Boye

Examinator: Mads Dam

Wizard-of-Oz Studies

Abstract

This degree project will explore the Wizard-of-Oz approach to perform empirical studies. The scope of the project includes areas of application, strengths, weaknesses and ethical considerations. The psychological aspect will also be covered and many examples of past experiments will be discussed. In spite of the ethical arguments and limitations against the methodology, the conclusion is that it is a good technique that should be considered as a choice of methodology as long as the requirements of the methodology are met, the experiment must for example involve a task that a human is able to simulate in a reasonable manner.

Wizard-of-Oz studier

Abstrakt

Examensarbetets syfte är att utforska Wizard-of-Oz metodologin som ett alternativ att utföra empiriska studier. Arbetets omfattning inkluderar tillämpningsområden, fördelar, nackdelar och etiska synpunkter. Dessutom tas den psykologiska aspekten i beräkning och många exempel av tidigare studier och experiment inom området diskuteras. Trots de etiska argumenten samt begränsningar inom metodologin är slutsatsen att det är en bra teknik som borde övervägas som ett alternativ under förutsättning att metodologins krav är uppfyllda. Experimentet måste till exempel involvera en uppgift som en människa kan simulera på ett trovärdigt sätt.

Acknowledgements

I would like to thank my supervisor Johan Boye for valuable input and advice. I would also like to thank Mats Wirén for his time to answer my questions and share his results and experiences of conducting a Wizard-of-Oz experiment at TeliaSonera.

Table of Content

1. Introduction

This thesis essay will discuss and evaluate the Wizard-of-OZ methodology.

1.1 Background

Technology advancements during the 20th century allowed for the rise of computer science as we know it today (Datorhistoria och datorutveckling 1999). The use of computers and computer run systems is increasing in popularity as it becomes less costly and time consuming to rely on computers rather than humans. Human psychology, dialogue and interaction are, however, very complex phenomena that are not always obvious. Therefore the interest in studying human interaction with computers arose. One method that can be used to perform such studies is the Wizard-of-Oz methodology. The name comes from the book written by L.F Baum (1900) of the same title. (Wei-Haas 1985) In the story, the wizard makes images of him selves and leads the characters to believe that the images are the wizard him self. Therefore the Wizard-of-Oz methodology includes the idea of a wizard and participants that are duped with an image of some sort. The first researcher to officially name the methodology Wizard-Of-Oz was J.Kelley (Kelley 1984) in his research about natural language interfaces (Wei-Haas 1985). Although the origin is not completely unambiguous, clearly the idea of the paradigm was formed somewhat earlier, for example in experiments conducted by Chapanis (Wei-Haas 1985). The Wizard-of-Oz methodology has, since Kelley, been used in many areas of application, not only for researching purposes and in the context of natural language interfaces. In fact, in 1984 the methodology was introduced as a useful tool in software development by Wixon, Witherside, Good and Jones (Wixon 1984) and the methodology has also proven to be of interest to other areas.

1.3 Purpose of the degree project

The purpose of this degree project is thus to give an account of the Wizard-of-Oz methodology and to explore the different areas of application. The methodology can, for instance, be used to test a hypothesis, improve a user interface or evaluate an idea of a software product. As the methodology most often is applied to experiments that are interested in the reaction and thinking of humans in a specific situation, we are also interested in how computer science and human psychology relate to one another with regard to this particular topic. More specifically, how is knowledge about human psychology used within the methodology and can the methodology itself be a useful tool within experimental psychology? The Wizard-of-Oz technique will hence be evaluated and ethical considerations will be discussed.

1.4 Methodology

The degree project will rely much upon exhaustive research on the Wizard-of-Oz methodology. It will be objective in its nature as well as source-critical. Many examples of previous publications within the area will be given as to cover as many different aspects as possible. The text will contain conclusions and opinions of the author based on the research and these will not have references. An interview with Mats Wirén will also be carried out as

to be able to provide the reader with first-hand information of an example experiment conducted by Wirén and his colleagues at TeliaSonera.

2. Wizard-of-Oz

2.1 The purpose and need of the methodology

The purpose of the Wizard-of-Oz methodology is to simulate an interaction between a computer and a human. This allows for instance software engineers to evaluate the human-computer dialogue of a system that has not yet been implemented. The idea is most often to lead the participant into the belief that he is interacting with a computer when it is in fact another human at the other end. The participant's reaction to the interface to what he believes is a computer is noted by the experimenter who is now able to predict the reaction to a situation where a computer is indeed at the other end. The experimenter responds to the user input either by direct communication of some sort or by choosing pre-determined output messages. Another thing worth mentioning is that the experimenter can also observe and note the behavior of the participant to see whether the dialogue, for example the wording and phrasing, changes as he thinks a computer is the subject to his interaction. The second example would serve as a good indicator in determining whether humans use for example fewer expressions of emotions in the interaction with computers, but the methodology does not restrict the interfaces to only natural language. According to Kelley (Kelley 1984) regular "office systems" can use one of the following input modes. Menus, command language or natural language and the Wizard-of-Oz methodology is applicable to all of these. The choice of input mode that should be used for a system depends on the level of knowledge of the user. Fraser and Gilbert (Gilbert 1991), however, explore the different design options using natural language which are:

1. Only the subject uses natural language
2. Only the wizard uses natural language
3. Wizard translators – which translates the user input into another type of dialogue
4. Both interactants use natural language

It is assumed that natural language can appear in both spoken and written forms and Fraser and Gilbert assert that although speech is more natural than writing, there are certain tasks that spoken language are not suitable for. In these cases written language should be considered more natural. Whichever design method that is used, the Wizard-of-Oz methodology is a useful technique in developing user-friendly systems and/or simulating man-machine interaction and dialogue. Weaknesses and strengths of the methodology are discussed further in *section5*.

2.2 Areas of application

There are many possible areas of application of this methodology. Linguists interested in human dialogue, experimental psychologists and software engineers are some, amongst all, that could take advantage of it. The psychological aspect will be discussed in *section 4* and will therefore be disregarded in this section. Since the methodology was formed, it has been used for many different purposes and the most common are accounted for below.

2.2.2 To evaluate a user interface and/or an application before implementation

To be able to design and evaluate a user interface of a system or the system itself still at the planning stage is, as mention previously, both an advantage from an economical point of view, but it also provides the developers a good insight of how the user prefers to interact with the system and can thereafter reconsider the design if necessary. The experimenter can, for instance, evaluate how the interface is perceived by the user, if there are any misunderstandings, how fast the user performs certain tasks or how the user behaves. Price, Dahlstrom, Newton and Zachary (Price 2002) used a Wizard-of-Oz approach to evaluate how students would like to interact with a system that can code on command by the user using natural language. They convinced students that they were interacting with such a system when in fact a programmer at the other end was coding the requests of the students. Price, Dahlstrom, Newton and Zachary argue that "If a researcher wants to learn how people will use a program, it is necessary to convince them that they are actually using that program". Therefore the Wizard-of-Oz technique is a good choice of methodology and post-session interviews with the students also showed that the students reacted very positively to the idea of the application that was not yet developed.

2.2.3 To collect high quality data

Another area of application of the methodology is to collect high quality data of some sort. Cheng et al (Cheng 2004) used the technique to collect human-computer dialogues in high-stress situations with the ultimate goal to enable natural interaction when a driver is operating a car or some other machinery. They believe this would "help reduce the user's overall cognitive load". Another example will be covered in *Section 3* where data is collected to train a speech recognizer.

2.2.4 To iteratively design a user interface

Kelley's (Cheng 2004) study in 1984 is an example of how the Wizard-of-Oz methodology can be used to iteratively design a user interface. His goal was to develop a natural language computer application: CAL, *Calendar Access Language* and conducted an empirical study prior to the development in order to design the user interface. The iterative methodology comprises of at least three phases: pre-experimental, first and subsequent phase. This particular study comprised of six phases.

1. Task analysis
2. Deep structure development
3. First run of OZ simulation
4. First-approximation language processor
5. Second run of OZ, using intervention
6. Cross validation

The idea is to continue to design the user interface while running simulations using interventions by the experimenter and as the design progressed, less and less intervention is needed. This text will not cover the phases in detail and any reader further interested in how the experiment was performed is referred to Kelley's own publication, (Kelley 1984).

2.2.5 To test a hypothesis

Another way of using the Wizard-of-Oz methodology is to test a hypothesis, often related to, but not restricted to man-machine dialogue. Hauptmann and Rudnicky (Rudnicky 1988), for example, were able to conclude that the number of pronouns referring to the dialogue participants is higher for speech than for typing. This was a part of a Wizard-of-Oz study with participants interacting with an email system. As long as it is possible to set up a simulation environment applicable to the Wizard-of-Oz methodology, any hypothesis can be tested.

3. Example experiment

3.1 Context

TeliaSonera as the first company in Scandinavia wanted to implement a new call-routing system for their customer service using natural language with open questions. (Wirén 2007) Thousands of customers call in everyday to ask questions and it is of course better for both the customer and the company if the call is directed to the correct agent immediately instead of being transferred a couple of times first. To have employees asking the customers questions and then directing them is obviously more expensive than to implement a call-routing system that will do the work automatically. Why else would most companies choose the automatic choice? Prior to the experiment TeliaSonera had already a live system that uses touch-tones but they wanted to test the idea of system that instead uses natural language. (M. Wirén 2010) There were several problems with the existing system. There could for example only be a maximum of about four or five choices at each level due to difficulties for the customers to memorize them. Neither could there be too many levels as this would be considered too frustrating and therefore it was difficult to find a unique combination of the choices to all the different 80 end destinations. It also occurred that the customers had navigating problems and could not find their intended destination or had difficulties mapping their problem to the choices at a given level. At the same time the company had to consider the customers reaction to such a new system using natural language, and the application would also have to be trained with training data in order to be able to correctly recognize speech and key words.

3.2 Purpose

This is a good example of a situation where the Wizard-of-Oz methodology can be used with success. TeliaSonera conducted an experiment with two purposes. (E. E. Wirén 2007) The first one was to collect speech data that would be used to train the speech recognizer. The second goal was to obtain data that could be used to guide dialogue design of the intended application. The experiment was also a way of testing their idea to see how the customer would react to the system and if it was indeed a good idea.

3.3 Alternative methods considered

According to Wirén there are two alternative methods to the Wizard-of-Oz approached that can be considered for the described purpose and these are accounted for below.

3.3.1 Human-human dialogues

The method comprises the recording of human to human dialogue, and in this case, the recording of dialogues between real customers and service agents. The method is of course comparably inexpensive, less time consuming and easy to conduct. Another advantage is that the customers are not affected in any way. The disadvantage, on the other hand, is that the dialogue between humans tends to be different than that of humans and machines, and

therefore the collected data is not representative enough. If the customer is speaking to a human, it is not possible to collect any data on how the customer would converse when a machine is at the other end.

3.3.2 Automated applications

Using an automatic application would be to integrate a fully automatic system for call-routing, in this case in a live environment, which records the dialogue repeatedly until the system comprehends what the customer says and then directs all of the calls to the same end point i.e. a customer service agent. The system will thus not direct any calls as they all go to the same destination, but just records the conversation to collect the desired data. This method, however, could have a negative customer impact .The customer can be asked multiple times to describe their issue and it is possible they perceive this as very frustrating and even if a human is allowed to take over when a customer clearly has problems getting out of the "recording loop" the customer may consider it a bad experience calling to the customer service. It is moreover possible that if the customer pauses for a short period of time to think, the automated system will perceive the response delay as an end-of-speech and therefore interrupts the customer. Yet, this approach is significantly less expensive compared to the Wizard-of-Oz approach.

3.4 Design

The experiment was conducted with real customers as participants in a live environment. Acting as wizards did ten experienced customer service agents who had in advance attended a course on how to use the system and how to act as wizards. The agents were also a part of developing the GUI, in particular, of the application used in the experiment. During five weeks 42,000 dialogues were recorded and the wizards used pre-recorded audio clips to answer the customer. The wizards could at any time interrupt and take over the call if they noticed that the customer seemed to have problems. When the customer had spoken there were three possible actions that the wizard could take. (E. E. Wirén 2007)

1. If the customer described the problem in a sufficient way the wizard would direct the call to the correct end destination.
2. If the customer failed to describe the problem in a sufficient way the wizard would play another prompt that encouraged the customer to give more information about the problem.
3. If the customer had problems the wizard could thus choose to take over the call.

During the weeks of data collection several different experiments were also conducted within the context. These tested for example the difference between aspects of the customers' answers when using different styles of prompts.

3.5 Results

The experiment and the data collection were successful. The call-routing system was implemented at TeliaSonera and the system is still being used. High quality data was collected and therefore they were able to train the speech recognizer and design the system most probably in a more user-friendly way than they would have been able to do without the

collected data. After the speech recognizer had been trained with the data from the live environment the accuracy was measured to 85%. (E. E. Wirén 2007)

3.6 Evaluation of the choice of methodology

In the interview Wirén said that he still believes the Wizard-of-Oz methodology was the best available choice of methodology but it was of course a very expensive and time consuming project. The idea was that although the initial cost was high, there would be less expenses and work at the end and the quality of the system would be higher. A minimum of negative customer impact was also a priority and at the end no customer impact could be shown. This was a significant advantage of using the Wizard-of-Oz methodology but it is necessary to emphasize that the wizard could take over a call if the customer experienced any problems. As the experiment was performed in a live environment with real customers, the experiment did not lack realism and the usual problems with a laboratory environment could be avoided. The wizards were real customer service agents and were therefore able to quickly grasp what the customer's issues were. Although this is good as they are used to dealing with customers and know what information that is needed, Wirén noticed that at the beginning of the data collection period the wizards would start playing the next prompt too early because they knew what the customer wanted to say before he or she was finished speaking. This was however not a critical problem because the wizards quickly learned to act their role.

4. The psychological aspect

4.1 The psychology of human dialogue versus computer dialogue

The science of psychology can of course be used within the Wizard-Of-Oz methodology as the goal is to observe human behavior, reaction or dialogue. The dialogue between people and computers is not easy to define. It takes place in many different forms, such as touch tone, command language, natural language and many more. What seems to characterize the dialogue, however, is that it is much more consistent than human dialogue, making the structure of the model a bit more simplistic. Dahlbäck, Jönsson and Ahrenberg (Dahlbäck 1993) discuss this topic and their ultimate claim is that "it is natural for any human engaging in a dialogue to adapt to the perceived characteristics of the dialogue partner." When speaking to a child for instance, people adapt their language to a level where the child is able to understand. Therefore one should never use human dialogue as the ideal example of man-machine dialogue nor try to develop a natural language interface that aims to resemble this dialogue. Obviously it is valid to argue against this claim and assert that the more man-machine dialogue resembles human dialogue, the better it is and to an extent this is probably true. People presumably prefer to use a dialogue that they already know and with the least limitation in how they can express themselves, but I will have to agree with Dahlbäck, Jönsson and Ahrenberg in that there are many differences between the two types of dialogues. Their most convincing argument is that human dialogue follows the "rules of politeness". R. Lakoff (Lakoff 1973) concluded that in human dialogue people tend to follow three rules. Do not impose, give options and make the receiver feel good. Although Reeves and Nass (Nass 1996) conclude the opposite, namely that people treat computers as real people, one could argue that people do not feel a need to strictly follow these rules when speaking to a machine. Lakoff has also claimed that the way women speak can be distinguished from the way that men speak. One example is that women tend to apologize more. If even the dialogue differs between men and women, then it is not difficult to imagine that the dialogue between humans and computers differs from the human dialogue. People thus adapt their language to the machine and interfaces should therefore be designed according to the adaption of the user to the system and not according to human dialogue. The Wizard-of-Oz methodology does not disregard this psychological aspect and is consequently a good technique for experimental purposes. Having an understanding of the topic could therefore be an advantage when designing an experiment. Perhaps a psychologist could or should be consulted when designing and evaluating a Wizard-of-Oz experiment.

4.2 Experimental psychology and Wizard-of-Oz

Some psychologist themselves are also interested in the interaction between computers and humans. As mentioned, human dialogue and interaction is a complex phenomenon and psychologist are of course also interested in studying human interaction with computers as to be able to acquire a better understanding of how the human mind functions. We have previously discussed how psychology can be used within the Wizard-of-Oz methodology. Having an understanding of how people think can above all be helpful when designing an

experiment or when drawing conclusions about the results. Now we will examine whether experimental psychologists themselves can use the methodology to bring research about the human mind forward. Cognitive psychologists in particular believe in the computational theory of the mind. (Horst 2005) It is based on the assumption that our mind functions similar to a computer in the sense that we receive input from the environment, the input is thoughts that are processed in the mind and then our behavior is the output. The purpose of the process is to find the best possible output that will lead to our desired goal. Cognitive psychologists therefore study the mind in terms of a computer and perform experiments accordingly. One famous experiment in this area is ELIZA, conducted by Joseph Weizenbaum. (Weizenbaum 1966) In a Wizard-of-Oz experiment a human simulates a computer or a system. The aim of ELIZA was to let a computer simulate a human. The participants were told that they are communicating with a therapist through a chat but in reality the ELIZA program answered the questions of the participants only by following programmed rules of how to connect sentences etc. ELIZA succeeded in deceiving the participants, who had difficulties believing it was only a computer even after they had been told the truth. (Weizenbaum 1966) The result supports the idea that it is possible to simulate human dialogue in a somewhat convincing manner by means of a computer program. The Wizard-of-Oz methodology, on the other hand, would enable the psychologist to study the human mind without the limitations of an existing system. Of course the methodology itself puts restrictions on what type of experiments that can be performed, as will be discussed further in *section 5,* but hypotheses similar to those tested by Hauptmann and Rudnicky (Rudnicky 1988) as well as Kennedy (Kennedy 1988) are examples of hypotheses that experimental psychologists could be interested in. In their experiments, they amongst all concluded that the total number of words used when conversing by typing is significantly less than when conversing by speech and the total number of words used when conversing with a computer is significantly less than when conversing with a human. The Wizard-of-Oz methodology is most certainly applicable to these hypotheses and it is therefore possible to assert that the Wizard-of-Oz methodology is or should be used within experimental psychology. Furthermore, going back to the conclusion drawn by Weizenbaum we can take this thought further and suggest that one possible application of the Wizard-of-Oz methodology in the future could be prior to the development of computers or robots that are supposed to imitate the behavior of humans. This is an ongoing and growing research area. If a human cannot tell the difference between a computer and a human, then we are very close to artificial intelligence.

5. Evaluation

5.1 Strengths

As mentioned previously, there are many areas of application of the Wizard-of-Oz methodology and therefore many advantages with the methodology regardless of the purpose of the experiment. Green and Wei-Haas state that the technique is "an efficient way to examine user interaction with computers and facilitates rapid iterative development of dialogue wording and logic" and this claim is certainly enhanced by the results of Wizard-of-Oz experiments such as that at TeliaSonera. Dahlbäck, Jönsson and Ahrenberg seem to agree and argue that it is indeed the "best available alternative for gathering data as a basis for theories of the specific genre of human-computer interaction in natural language." Of course the methodology can not only be used for natural language interfaces, but also for any sort of intelligent interface between humans and computers, such as command language. The methodology also offers an excellent method of collecting valid training data for statistical algorithms that will use for example voice recognition. Fraser and Gilbert discuss the problem of designing a system relying solely on intuition and judgments of the designers, which according to them lead to a system with flaws reflecting the prejudices of the designers. They are in accordance with von Hahn (Hahn 1986) stating" we have no well-developed linguistics of natural language man-machine communication" and this is problematic. The use of the Wizard-of-Oz methodology is, however, a way to overcome the problem allowing for system designers to evaluate how the users want to be able to communicate with the system, what kind of functionalities that they expect etc. Obviously this is a much better alternative than to repeatedly re-design the system or in the worst case, have a system that does not fulfill its performance requirements. A significant strength of this methodology is thus that it can reduce time and costs related to the development of a system. It is moreover a good way of developing a user-friendly system that interacts well with the end user. The participants are able to give feed back to the developers so that the system can be improved. The methodology provides the developers a good prediction of how the system will work in its own intended environment. As seen above, it is additionally a method of studying how humans interact with computers. What type of language they use and if they respond differently to a computer rather than to another human and this has a great research value in the field of psychology, linguistics and probably many more.

5.2 Weaknesses

One important weakness with this methodology, as has been argued by many, is that there is a lack of realism (E. E. Wirén 2007). Although this is not the case with the experiment conducted by TeliaSonera, most often an experiment is performed in a laboratory environment. This means that the participants are not real users. They are participants of an experiment and will perhaps act accordingly. A real user of the system may be more determined to perform the necessary task, or perhaps more determined to perform it thoroughly, and the results of the experiment can thus be misleading. (Dahlbäck 1993) Yet, as has been concluded by Dahlbäck, Jönsson and Ahrenberg that, "If the focus is on aspects not under voluntary conscious control, the prospect is better for obtaining ecologically valid

data". This means that if the focus of study is on the use of grammar or construction of sentences, for instance, then it is not likely that the role play in this sense will affect the results. They also discuss the critique that "One should study existing systems instead of simulated ones" but agree with Tennant (Tennant 1981) that "People often adapt to the limitations of an existing system, and such an experiment does not therefore tell you what they ideally would need". Another weakness of the methodology with regard to software development is of course that all systems or interfaces cannot be simulated using the Wizard-of-Oz methodology. An example of such a system could be one that is supposed to perform a task that a human cannot perform in a reasonable amount of time, for example complex calculations. The methodology is based upon the assumption that the human will be able to simulate the system so well that the participant actually believes it is a computer. In their review of the Wizard of Oz technique, Frazer and Gilbert proclaim three pre-conditions that must be fulfilled in order to be able to conduct an experiment.

1. It must be possible to simulate the future system, given human limitations
2. It must be possible to specify the future system's behavior
3. It must be possible to make the simulation convincing

Clearly this is a huge disadvantage of the methodology as probably most possible situations will not fit into these requirements. Even with systems that can be simulated, one has to consider human factors when designing the experiment. People are, for example, slow type writers and they tend to make occasional and random errors such as spelling. If these factors are not taken into account, the simulation may be difficult to run. It is furthermore not possible to guarantee that the computer or the system will act the same way as the human did in the experiment. After all, people are flexible and inconsistent whilst computers are deterministic and consistent. In addition, the need of many participants could entail a larger scale experiment that cost a lot of money itself. Not to mention the amount of time required to perform the experiment. This is not good if the purpose of the experiment was to reduce the costs of the development, making the methodology somewhat unnecessary.

5.3 Ethical considerations

As with many experiments, there are several ethical aspects to consider. In this particular case it is relevant to ask whether it is ethical to deceive the participants or not. The experiment would be useless otherwise. If the participants, on the other hand, were to sign a form of consent in advance saying they agree to be duped, then their expectations of the experiment might alter the results. In some cases this methodology can be used without even the participant knowing he is a subject to an experiment. This can certainly be considered unethical but if nothing in the results can be tied back to a particular individual then no harm can be shown. Dahlbäck, Jönsson and Ahrenberg argue that due to the differences in the language used and the communication itself when human converses with another human compared to the conversation with a computer, it is necessary to deceive the subjects. If moreover the subjects are debriefed afterwards and they feel that they have not been exposed to inconvenience of any kind, then this ethical aspect should not be considered too much of an obstacle. On the other hand, guidelines declared by the Swedish Vetenskapsrådet (Vetenskapsrådet 2010) state that it is not ethical to deceive participants of an experiment, and although researches receiving government funds are obligated to follow these, they are only guidelines. (M. Wirén 2010) A controversial experiment that has raised a lot of

discussions about ethics is that of Milgram (Milgram 1974). In his experiment participants were tested on how much pain they were willing to cause another human when told so by an authority. Results showed that most people were in fact willing to harm another person when instructed to do so by the authority, while watching him or her suffer. Of course nobody was actually hurt during the experiment. The participants were only deceived to believe so. This experiment would not be allowed today due to modern ethical standards. Comparing to the Wizard-of-Oz methodology it is arguable that the participants are deceived in both cases and the methodology should therefore be considered unethical. Yet, the distress caused to the participants in the experiment of Milgram is significantly higher than that of any Wizard-of-Oz experiment coverer in this text and it is therefore easier to agree with Dahlbäck, Jönsson and Ahrenberg that no inconvenience can be shown from the participant's point of view and thus the experiment is not to be considered too unethical to not conduct it at all. Another thing worth mentioning is the ethical aspect regarding the handling and processing of data. This is probably not specific to the methodology but rather to the nature of the experiment. As with all personal data there are both legal and ethical aspects to take into account. If data collected during a Wizard-of-Oz experiment is not processed in an ethical manner, then it is more reasonable to claim that the researcher or the experiment is unethical, since the Wizard-of-Oz methodology it self does not restrict nor encourage the processing of data in any direction.

6. Conclusion

As seen above, Wizard-of-Oz is clearly a powerful technique to perform empirical studies involving a simulation of a man-machine interaction. For all the areas of application it is a good or perhaps the best available choice, depending somewhat on the experiment of course. The psychological aspect is interesting and speaks for the methodology. Man-machine dialogue clearly differs from human dialogue and this is emphasized by and used within the it. It has also been stressed that Wizard-of-Oz can be applied to experimental psychology. The ethical aspect is, however, a strong argument against Wizard-of-Oz but I have not come across any argument convincing enough to assert that the methodology should not be practiced. Neither do limitations such as lack of realism convince me that the methodology is not valid. A number of experiments have been covered in this text and all have had more or less successful results. The restrictions that the methodology puts on the nature of the experiment are obviously a greater problem. The methodology simply cannot be used if the task cannot be simulated in a convincing manner though I can still argue for the methodology in situations that fulfill the requirements. Hopefully we will be able to simulate even more situations in the future, making the Wizard-of-Oz approach even more powerful.

References

Cheng, Bratt, Mishra, Shriberg, Upson, Chen, Weng, Peters, Cavedon, Niekrasz. ”A Wizard of Oz Framework for Collecting Spoken Human-Computer Dialogues .” 2004.

Dahlbäck, Jönsson and Ahrenberg. ”Wizard of oz studies –Why and how.” 1993.

Datorhistoria och datorutveckling. KTH. 1999. http://www.e.kth.se/~e97_fne/datorhistoria/ (Accessed on 22 04 2010).

Gilbert, Fraser and. ”Simulating speech systems.” 1991.

Hahn, Von. ”Pragmatic considerations in man-machine discorse.” 1986.

Horst. ”The Computational Theory of Mind.” 2005.

Kelley, J. ”An iterative Design Methodology for User-Friendly Natural Language Office Information Applications.” 1984.

Kennedy. ”Dialogue with machines.” 1988.

Lakoff. ”The Logic of Politeness; or minding your p’s and q’s .” 1973.

Milgram. ”Obedience to Authority: An Experimental View.” 1974.

Nass, Reeves and. ”How people treat computers, television, and new media like real people and places.” 1996.

Price, Dahlstrom, Newton and Zachary. ”Off to see the Wizard: Using a “Wizard of Oz” study to learn how to design a spoken language interface for programming.” 2002.

Rudnicky, Hauptmann and. ”Talking to computers: an empirical investigation.” 1988.

Tennant. ”Evaluation of Natural Language Processors.” 1981.

Wei-Haas, Paul Green and Lisa. ”The wizard of oz: A tool for rapid development of user interfaces.” 1985.

Weizenbaum. ” ELIZA- A computer program for the study of natural language communication between man and machine.” 1966.

Vetenskapsrådet. *Vetenskapsrådet.* on 2010 April 2010. http://www.vr.se/ (accessed on 1 May 2010).

Wirén, Eklund, Engberg and Westermark. ”Experiences of an In-Service Wizard-of-Oz Data Collection for the Deployment of a Call-Routing Application.” 2007.

Wirén, Mats, intervjuad av Elina Meier and Peter Decker. (on 27 April 2010).

Wixon, Witherside, Good and Jones. ”Building a user derived interface.” 1984.