Experiences of elderly with emotionally beeping ECAs

Embodied by a Nabaztag™

P.R. van der Graaf
p.r.vandergraaf@student.utwente.nl

ABSTRACT
The goal of this research is to investigate how elderly people experience an ECA using beeping sounds as expressions of emotional states. In a possible future scenario is that these ECAs will be responsible for basis diagnostic decision-support tasks concerning our well-being. Qualitative research is done on how elderly people experienced ‘living with’ a Nabaztag as companion. This Nabaztag performed the role of the ECA. Prior to the research, an online questionnaire was held to establish a classification of beeping sounds into different emotional states. In the second stage of the research, subjects had to classify a series of beeps into emotional groups before and after they took part in a scenario play with the Nabaztag. During these sessions, the Nabaztag communicated his emotional state by sounding appropriate beeps while with one group of subjects and communicated its emotional state with random emotional beeps while with the other group. Qualitative evaluation of living with a Nabaztag showed that a Nabaztag with an appropriate response to events was much more positively experienced than a Nabaztag making a random response. The elderly people enjoyed having a Nabaztag making beeps accompanying them. Because of the Nabaztag’s reactions to certain events, a ‘belief’ of a Nabaztag’s primitive mind was created.

Keywords
ECA, Nabaztag, emotions, emotional state, human to ECA interaction, elderly people, beeps, R2D2.

1. INTRODUCTION
Each year, elderly people represent a larger proportion of our western-worlds population. For example, in the year 2030 there will be over 4 million people over 65 years of age in the Netherlands out of the total population of 16 million. This means that almost a quarter of the entire population of the Netherlands [1]. Because of this, the health-care budgets will rise disproportionally. Not enough people will be available to nurse this group of elderly people. A solution to lower the costs of some healthcare related issues for these elderly people is to accompany these people with an electronic agent, equipped with measuring instruments to take care of some tasks related to the elderly person’s wellbeing. For some simple basis diagnostic decision-support, these electronic agents could replace labor of a human nurses. Embodied conversational agents, ECA [2], are computer-generated characters that demonstrate some of the many properties humans have in face-to-face conversation. ECAs have the ability to produce verbal and nonverbal communication. In these communications they should be able to communicate their emotional state. Having face-to-face communications is a skill that many older adults do not lose while going downhill and showing more signs of old age [3]. This is one of the reasons why ECAs will be a good fit to take care of these health-care related tasks [4].

An example of an ECA is R2D2, a famous robot from Star Wars, an epic space opera franchise initially conceived by George Lucas [5]. This R2D2 communicates his emotional states by making beeping sounds. People claim to ‘understand’ R2D2’s feelings by just listening to the beeping sounds the robot produces. This research is about recognizing and classifying these ‘emotions’ in the different beeping sounds the ECA will produce. Sounds used in this research are sounds like this R2D2’s beeps. Programming and designing a full-scale ECA is not part of the research. The AI is replaced by the researcher, using the the Wizard-of-Oz method [6]. This way the focus of the research will not have to be to develop an intelligent system, but will be on recognition of beeps as emotions and the subject’s experience using a Nabaztag / ECA. The ECA is embodied by a Nabaztag™.

1.1 Relational agents
Relational agents are computational artifacts designed to build and maintain long-term social-emotional relationships with users [7]. These agents can either be purely software animated, humanoid or other creature-like characters, but they can be also embodied in many physical forms like robots, hand-held computers, avatars and even toys. Important to the concept of these embodied conversational agents (ECA) is that relationships are built up from a series of interactions from user to agent. The agent should create a history of these interactions and manage future expectations using this database. Relationships can be constructed without language, but language is crucial for the development and management of human relationships [8].

1.2 Nabaztag
The ECA used in the research is embodied by an electronic rabbit. (Nabaztag™) A Nabaztag is a Wi-Fi enabled electronic device in the abstract shape of a rabbit, Figure 1. A Wi-Fi connected Nabaztag, communicating via the manufacturer’s, Violet, website [9, 10] can receive information from all different preconfigured selectable applets. Applets include functionality like reporting the weather, reading from news feeds, playing internet-radio, reading the owner’s mail etc. These applets are fully customizable and programmable via the Nabaztag API [10]. During the research, the text to speech (TTS) engine was used to let the Nabaztag introduce itself to the user. The ability to stream music was used to play the selected emotional beeps.
Some people paint their Nabaztag, give the Nabaztag behavior via the website, or even buy clothing or other accessories for the Nabaztag. The Nabaztag is equipped with a microphone and voice recognition software to communicate with its user. Basic applets, to install via the Violet website, include; voice commands for reading mail, telling the weather or opening a stream to internet radio. Equipping your agent with behavior, caring about him and therefore buy or make it clothes is the start of a real relationship between ECA and its user. The Nabaztag is the selected embodiment for the agent in this research because the Nabaztag is easy to approach and program, and because it is a very functional and feature-rich robot. It is much more approachable than for instance the AIBO [11] due to the Nabaztag’s simple design and easy configuration. The Nabaztag covers all the aspects of an ECA for study purposes. The TTS-engine is configurable to many different languages.

2. RELATED WORK

When people have established some sort of emotional connection with the ECA, they are more likely to listen to the ECA and value the ECA more [12]. One way of communicating emotional states could be making beeping sounds. Will elderly people recognize these different emotions made by the ECA during communication and link these emotions to human-type of behavior? And therefore feel a stronger connection to the ECA. When they don’t ‘listen’ they have feelings of guilt towards the ECA as they would have towards human beings. When a tight relationship can be established between ECA and an elderly person, ECA can help take better care of the well being of that person. Of course this 24hour available agent is not going to play the role of a fully trained health care professional, but it is cheaper and it can perform many basic decision-support tests and give the elderly person some sort of self-confidence, or even a communications partner, when, for example, living alone. Much research has been conducted on ECAs and the interaction between people and ECA [13-16]. When ECAs are given the ability to express in non-verbal ways, the way of the expressions made, and the expectations of the people about the ECA, have to match in order to be effective. This is defined in the “SE2PM: simple expression to primitive mind” principle [17].

In the research of Yamada and Komatsu [17] and Komatsu’s studies [18], the researchers used beeps to realize simple expressions. They used these beeps because very simple beeps in Komatsu’s research [18], showed that people can estimate different primitive minds by means of simple beep like sounds with different durations and alterations in pitch and tone. Sounds with shorter duration and decreasing intonations were perceived as ‘positive mind’. Sounds with increasing intonation were perceived as ‘negative mind’ regardless of its duration. Flat sounds with longer durations were estimated as ‘neutral mind’. These simple but intuitive and effective beep sounds were successfully used to have two robots that were un-life-like in their appearance and behavior, communicate their primitive minds to their users.

In this study, more complex beeping sounds are used to have a life-like robot communicate his slightly less primitive mind communicate his emotional status to the users. This should be a good fit according to the SE2PM principle, slightly more complex beeps and a slightly less primitive robot.

In the research of Bickmore [12], the researchers discovered that ECA with several relational skills, behavior explicitly designed into a computer interface to improve their working-relationship, can improve the wellbeing of their user. Results of a comparison among 33 subjects interacting near-daily with the relational system, and 27 interacting near-daily with an identical system with this relational behavior ablated, each for 30 days, indicate that the use of relational behaviors by the system significantly increase the working alliance and desire to continue working with the system. Comparison with a 31 subjects control group, working with a control system near-daily for 30 days, also indicated a significant increase in proactive viewing of health information. Bickmore’s research leads to the conclusion that having a relational ECA does improve the likeliness of the participants actually using and listening to his electronic health-care system. For instance, both relational- and control-group’s system stimulated the users to make 10.000 steps daily and upload the total amount of steps into the system. Not only the relational group uploaded their results more often to the system, but they also felt ashamed and guilty ‘unhappy for the ECA’ when they did not reach these 10.000 steps.

Having these ‘feeling’ and establishing this connection to the ECA will in the end improve the elder’s wellbeing, because a stronger patient-doctor or elder-ECA relationship is created.

3. RESEARCH QUESTIONS

The goal of the research is to find answers to the question of whether the emotional beeps are classifiable and improve the relation-establishment process between the ECA and its user. These research questions, are centered around the experience of the use of the Nabaztag for my research and the test-subject’s recognition of the Nabaztag’s emotional states communicated through the beeps.

R1. Will R2D2-like appropriate beeping sounds help in establishing an emotional relationship between the elderly person and an ECA?

R2. Will a Nabaztag linked to a personal-server with home-brew software create a usable/believable relational-ECA?

R3. How will the subject perceive working with the ECA?

R4. How will the SE2PM principle applied to the Nabaztag using beeps to communicate work out?
The research is designed to allow qualitative results about how elderly persons experience life with a Nabaztag as ECA. Conclusions from researcher’s point of view and experience from the participants. Therefore a global online questionnaire, (impression in Appendix A), a case study and evaluations of this case study is done.

4. APPLIED METHODS

In this study, a Nabaztag embodied the role of the ECA. The idea was to have a Nabaztag respond to certain situations with certain ‘emotional beeps’. This response has to be very ad-hoc, and the initial idea was to let the communication go via a JNabServer1, an open source JAVA server for the Nabaztag. After having spent too many hours configuring this server and seeing an increased lag between the command and the actual beep, it was decided to use the official Violet servers. There have been problems with the Nabaztag servers in the past, but after a few tests, the Nabaztag showed its reaction time was within one second after tapping the ‘sound beep x’ command.

In advance of the scenario play, an online questionnaire was held to establish the most accurate beep-emotions fitting the expected reaction on a pre-designed scenario.

To embody the relational robot concept in the research, our ECA has a Text-to-Speech engine built in and introduced itself to the test-subject, told a joke and read some random news from the internet. Also, the participants could ask the ECA to predict the weather. The ECA also responded to a few of the user’s basic introductory questions (using the Wizard-of-Oz method).

To promote a strong belief that the ECA was aware of the conversations, his passing and capable of understanding and communicating with the participants.

4.1 Study design

The basic components of the research consisted of an online questionnaire, a case study and an interview to evaluate each case study. The use of an online questionnaire was chosen to reach a broad group of people. The case study followed by the interview was chosen because a case study is most likely to give detailed information about living with the Nabaztag.

During the online questionnaire and scenario play, the same set of beeps were used. However, in the ‘appropriate beeps’ group, a distinct sub selection of beeps was made using the classification from the online questionnaire. Highest scoring beeps for; hate, joy, happy, sad, disagree, attracted-to, scared, happy-for, and sad were used.

4.1.1 First stage - Pre-test study

For the first stage of the study, an online questionnaire with 38 beeps for the questionnaires was prepared with the aim to have as many people as possible classifying beeps with a label. (impression in Appendix A) To come up with a list of labels, a set of emotions from the ‘Rationalized 22’ [19] was picked. This is a widely used classification of emotions and the emotions are grouped in quadrants, appraisal of an event, fortune of others, appraisal of an agent’s actions, appraisal of an object. These quadrants could be useful in case identifying individual emotions proved to be not possible. While picking these labels for the beeps, it was kept in mind that the biggest possible group, Dutch people between 20 and 60, should be able to translate and notice a clear distinction between the different ‘emotions’. From each quadrant in the Rationalized 22 scale, a few emotions were picked. In case the labeling should not give enough distinct results, constraints on the analysis could be relaxed, and instead a comparison made of quadrants in order to still make a clear distinction between the emotions.

4.1.2 Second stage – Scenarios with the Nabaztag

In the second stage of the research, a scenario play with a Nabaztag and an elderly person was done. The scenarios were pre-written sets of events, dictating what should happen during the play. To each event, the Nabaztag responded with a beep and the behavior and reaction of the elder was analyzed. For example; “a loud noise”, “preparing tea, Nabaztag left alone”, “excitement during the game when picking a new tile from the pile”. In short, the play consisted of; introducing, drinking tea, playing a game, saying goodbye. The length of the play was between 1.5 and 2 hours.

To participate in the research, two groups of two subjects were established. A group where the Nabaztag responded to different actions with ‘random beeps’, and a group where the Nabaztag responded to the different scenarios playing an appropriate beep to display his primitive emotional mind. For usage during these scenarios, a subset of the beeps from the initial questionnaire was picked. For example, beeps classified as ‘Joy – joyful’ or ‘Distress – shock’ in situations where the Nabaztag is happy with the attention or when the Nabaztag is frightened.

Proceeding this ‘scenario play’ with an ECA, the participants were asked to fill out the questionnaire as well. Classifying beeps into emotional categories. After the scenario play, the participants were asked to do the same thing. The goal of these two tests was to see with how much consistency in answers the group with the ‘appropriate’ and the group with the ‘random beeps’ answered the questions to the pre-test and post-test questionnaire. More similarity in answer in the ‘appropriate beeps’ group could indicate that the subjects learned to recognize the different emotional states from the Nabaztag’s beeps during the scenarios. Also, from the ‘appropriate beeps’ group, more conformity with the ‘classification’ created by the initial questionnaire was expected. As these elderly people should have learned to distinguish the Nabaztag’s different emotions during the play more effectively then the elders exposed to the ‘random beeps’.

4.1.3 Third stage - Evaluation

After the subjects finished the post-test questionnaire, the ‘performance’ of the Nabaztag and the experience of the subjects was evaluated. This was done using Dormann, Ruttkay and Noot’s framework for evaluating ECAs [20]. Asking the subjects questions like “How did the dimensions of the ECA impact your view on the Nabaztag’s Helpfulness?” and “How did the beeps contribute to the Nabaztag’s believability?”. For example, a 2 meter high walking crocodile could be perceived as less helpful. The subjects were free to give all their feedback and answers were written down for further evaluation.

Only after this had taken place, the subjects were told about the setup of the research. Initially, as little as possible about the goal and aim of the research was communicated. During the conversations with the participants nothing was mentioned about research and the evaluation of their response to beeps and ECA’s emotions. And especially to not influence the results of the questionnaires, it was not mentioned that the subjects had to submit the questionnaire twice and that their results would be compared.

4.2 Study participants

For the questionnaire preceding to the scenario play with the subject and the Nabaztag, the questionnaire was made publicly available on the website. To promote it, an email was sent to all

---

1. JNabServer project: http://www.cs.uta.fi/hci/spi/jnabserver/
currently known contacts of the researcher. However, this is a special, select group of people. To make sure the results were valuable and usable during the research, in the mail the recipient was asked to forward the mail to all their friends, family or colleagues, which should broaden the group. Which, from the reactions in response to the invitation, people apparently did. The last few questions in the questionnaire questioned people for some demographic data. This should alert when the questionnaire had reached too homogenous a group. To get even more valuable results, many email conversations with questions and extra feedback on the questionnaire were processed.

The questionnaire was submitted by 44 people. 16 within the age of 16-30, 10 persons in the range of 46-60 and 5 persons answered to be above this age. 28 of the respondents answered to be a male.. There was the option to pick ‘I rather not say’. An overview is printed in Table 2.

For the session with the subject and Nabaztag four elderly people (50+) with known background were selected. They had been rated on their expressionism and their affection with computers/electronics. None of them had contact or had previous experience with ECA or other robots.

4.3 Procedure

For usage during the four scenario plays, a script was prepared. Emotions were linked to the same events, and reactions were examined in more-or-less the same situations. As much as possible, the same actions were played, globally consisting out of introductions, tea-time, and a game of Rummikub. The length of the scenario play was between 1.5 and 2 hours.

During the sessions, notes were made about the participant’s behavior and reaction to the Nabaztag’s beeps (if any). These notes were used while evaluating the performance of the ECA afterwards. Some beeps totally neglected, misinterpreted, over-reacted, totally misfired, and information like that. Quotes give impressions of the reactions of the subjects to the ECA. In the meantime, a normal-as-possible conversation was held with the subject and a game of Rummikub was played. Because the storyline was prepared, the only action (to command the Nabaztag to sound his emote) needed to create the ‘emotional behavior’, was a simple tab on a PDA’s touch-screen interacting with an online dashboard. This dashboard was designed to be finger-friendly and ultra minimalistic, thus quick, in usage. Figure 2 gives an impression of the usability of this dashboard on a PDA.

The Nabaztag was linked to the internet using local wlan. With minimal distraction during the scenarios, these buttons could be tapped. This way the ECA could be operated with minimal disturbance on the process. During the scenario play quick notes were made about the participant’s behavior and response to certain situations, and the participants were quickly judged on their understanding and reaction to the beep.

The scenarios started with the Nabaztag and the participants introducing themselves to each other. The participants were told some basic information about the Nabaztag. The total control on the behavior of the Nabaztag was not mentioned. After the introduction, there was a natural proposal to go to the kitchen and prepare a cup of tea (with the Nabaztag responding Joyfully to the proposal). While preparing the tea out of sight of the Nabaztag, away from the Nabaztag, talking about something else then the research, the Nabaztag felt neglected (communicated by a beep). Back around the table with the Nabaztag, happily beeping to be in the picture again, a game of Rummikub was proposed. Rummikub is a tile-based game for up to four players. The game is invented in the 1930’s and these days targeted on elderly people. With the slogan ‘brings people together’ the game seemed fit for the studies. The Nabaztag itself didn’t take part in the game. Participants were told the Nabaztag was on the researcher’s side, ‘helping’ him and meanwhile judging the situation of the tiles and other things happening during the game. In total, two games of Rummikub were played. After the second game was played, a good-bye scene was orchestrated with a very disappointed/sad Nabaztag.

Figure 2. Impression of the minimalistic dashboard on a PDA

4.4 Measures and analysis

Initially, the questionnaire was designed to accept only one submission from each IP address. Quickly this was changed, because it caused problems when people tried to visit the questionnaire connecting to the internet via smaller networks. People responding their co-worker could not fill-out or housemate’s computer was blocked made me change the filter. In defense of people intentionally trying to sabotage the research; filling out the entire questionnaire (only then it was submitted to the database) takes enough time to prevent people from ‘filling it out for fun a few times’ or just filling the questionnaire out for the sake of sabotaging. When analyzing the results, there was manually scanned for falsified submissions and none were found. Date, time and IP address were recorded for this purpose.

The results of the initial questionnaire were used during the scenario play, where ‘appropriate’ emotions were needed. Out of all the beeps used in the questionnaire, a subsection was made with the highest scoring beeps with answers; hate, joy, happy, sad, disagree, attracted-to, scared, happy-for, and sad for usage during the scenario play.

In order to distil quantifiable results from the interview during the last stage of the research, the participants were asked to fill out a form about their experience connected to the various
5. QUANTITATIVE RESULTS

44 people filled out the questionnaire. The questionnaire consisted out of 41 questions. One beep was presented three times to check the consistency of the answers, most people scored this beep as ‘love – adore’ and ‘love – affection’. Two quite the same categories.

In the Table 1, a few results from the questionnaire are printed. Keep in mind that the questions in the questionnaire were designed, for the sole purpose of classifying the beeps in an emotional category, and then, for these beeps to be used during the scenario play, however that is not the only thing we can conclude from the answers. In the left column are a few facts and in the right column what was submitted.

People classified the most beeps as ‘joy- joyful’. The second most picked classification, ‘Disagreement’, has a more negative feeling to it. Meaning, people did not just pick funny, ‘happy’-sounding beeps. The ‘Disagreement’-label was added a few moments before the questionnaire went live. After some piloting, people responded really missing an ‘other’ category. This category was added and filled with ‘agreement’, ‘disagreement’ and ‘clueless’. ‘Clueless’ indicating that the agent tried to communicate it was clueless. A category like ‘other’ is always dangerous to spoil your results, but in the researchers judgment it was more important to not discourage people from filling out as many beeps as they could.

In Table 2, answers to the demographical questions are printed. It indicates that the questionnaire did not reach a homogenous group, for example, not just male students between 16 and 30.

Judging the used set of beeping sounds, compared to the choices participants of the questionnaire made, the same results were shown as in Komatsu’s [18] research. People picked the more-or-less beeps with shorter duration and decreasing intonations were perceived as ‘positive mind’(love). And beeps with increasing intonation were perceived as a more ‘negative mind’ (distress).

Table 1. Results from the initial questionnaire

<table>
<thead>
<tr>
<th>Fact</th>
<th>Which</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st picked</td>
<td>Joy - joyful</td>
</tr>
<tr>
<td>2nd picked</td>
<td>Disagreement</td>
</tr>
<tr>
<td>3rd picked</td>
<td>Love - attracted-to</td>
</tr>
<tr>
<td>1st picked quadrant</td>
<td>Appraisal of an event</td>
</tr>
<tr>
<td>2nd picked quadrant</td>
<td>Appraisal of an object</td>
</tr>
<tr>
<td>3rd picked quadrant</td>
<td>Appraisal of possible future event</td>
</tr>
</tbody>
</table>

Table 2. Demographic data about people who filled out the initial questionnaire

<table>
<thead>
<tr>
<th>Age</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-15</td>
<td>2.3%</td>
</tr>
<tr>
<td>16-30</td>
<td>36.4%</td>
</tr>
</tbody>
</table>

Table 3. Results from the evaluation after the scenarios

<table>
<thead>
<tr>
<th>Participant</th>
<th>Help</th>
<th>Sat</th>
<th>E</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Random 1’</td>
<td>7</td>
<td>-1</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>‘Random 2’</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>‘Appropriate 1’</td>
<td>10</td>
<td>13</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>‘Appropriate 2’</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>23</td>
<td>15</td>
<td>55</td>
</tr>
</tbody>
</table>

The numbers in Table 3 are the results from the final user-experience-evaluation of the usage of the Nabaztag. The data is extracted from the form, the subjects were asked to fill out after they had the play with the ECA. The columns from left to right are; helpfulness, user’s satisfaction, believability, trust, engagement and total. The participants were asked to rate these factors from the viewpoints (personification, dimension, physical details, beeps, etc.) The participants rated these total 10 viewpoints with a score ranging from -2 to +2. The total scores in each column is displayed in Table 3. The participants exposed to the ‘appropriate’ beeps rated the performance of the ECA almost twice as high. Noticeable is the difference in score in the ‘believability’ category, the ECA with the ‘appropriate’ beeps scored five times higher!

Table 4. Results from the pre and post tests

<table>
<thead>
<tr>
<th>Participant</th>
<th>C. Pre</th>
<th>C. Post</th>
<th>equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Random 1’</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>‘Random 2’</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>‘Appropriate 1’</td>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>‘Appropriate 2’</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
The pre- and post-scenario tests were less conclusive. Where there was expected to see a clear difference in ‘answering the questions conform the classification’ (made by the big initial questionnaire in the start of the research). In Table 4, the results of the pre- and post-scenario tests are printed. The columns are: participant, conform classification in pre-scenario test, conform classification in post-scenario test, and answers from pre- and post-scenario tests are equal. (e.g. Participant ‘random 1’ scored in the pre-scenario test 2 answers conform classification, scored in the post-scenario test 2 answers conform classification and scored 4 times the exact same answer in both pre- and post-scenario test.) The appropriate group showed a higher score conform the classification made by the initial questionnaire in their post-scenario tests.

6. QUALITATIVE FEEDBACK

Evaluation after the scenarios and my notes during the play, resulted in some quotes as response on the ECA’s beeps:

Random:

He is left alone...

After the Nabaztag was left alone on the table, the participant preparing a cup of tea; He wants to continue the research.

He is happy he cannot join us in this tea-party?

Did you draw a Joker?

Happy Nabaztag made the participant believe I had playable series of tiles on my rack; Play the tiles from your rack!

You and the Nabaztag are cheating!

He is whistling like a bird!

He is laughing at you or… is he?

Is he happy for me or sad for you?

Appropriate

He is watching you… he is watching you… there he goes again!

Oh, what was he saying?

After a few minutes of no beeps; I have to ask him the weather again, that way he gets more involved in our game.

After a second ‘sad’ beep in a row as response to ‘cheating’; He is way to strict to the rules of the game.

You’re a good rabbit, you are very kind. [laughs] slijmbal!

When the Nabaztag unexpectedly started to play some music; He likes it here with us!

Discussing why the Nabaztag did such a great job as ECA; In all its simplicity, the Nabaztag is ‘fitting’.

Sometimes the beeps are a little inconclusive.

Beeps might be not enough total produced sound to become a real conversational partner.

After a series of ‘negative’ beeps; We have to make him happy again, I think he liked me asking him for the weather, I will do that again

All the participants anthropomorphized the agent to some degree, putting the Nabaztag in a category somewhere between a computer and a person. There were many comments about the agent-participant relationship and relational dynamics. The general trajectory seemed to be that the relationship and interaction seemed strange and unfamiliar at first, became more familiar and by the end, two of the ‘appropriate’ participant could just not stop trying to see how the Nabaztag reacted to their questions and actions. All the participants tried to have some sort of social dialog with the Nabaztag. But the scope of the research and the speed I could secretly command the Nabaztag to answer some questions limited this.

Left alone on the table, during the preparation of the tea, the participants ‘understood’ the Nabaztag was left alone. Thinking of a Nabaztag capable of having these ‘feelings’ tells us, participants were of the understanding, the Nabaztag was actually capable of having these.

Each time the participants were in conversation with me, about things not related to the research, the reaction of the beeps was very low. Concluding that the ECA and the researcher were not equal as ‘partners for conversation’. The same result showed when the participant had to make ‘important’ decisions, what tea-bag to grab, where to place that left-over tile. Appropriate emotions worked for the Nabaztag to get the attention from the subjects concentrating, but did not have such impact when they had a conversation with me.

The subjects were all pleasantly surprised when they heard the Nabaztag speak (using TTS) for the first time. They were amazed how understandable and ‘realistic’ the voice sounded. During the scenario play, while preparing tea and being out of line of sight from the Nabaztag, the Nabaztag felt neglected (communicated by a beep). This was recognized immediately by the participant. The subjects in the appropriate and random group both mentioned this, interpreting different beeps as ‘loneliness’. People started to fill in the mind of the Nabaztag after a few inconclusive beeps. Whether this is a good or a bad thing, this is often the case in human-to-human communication.

The perception of the ECA’s emotion of the excitement and disappointment of drawing tiles, playing out the first tiles from your rack, were all recognized by the subjects. Even using various random beeps, there was recognition of ‘happy-for’ and ‘excited’ and disappointment emotions, responding to tiles drawn from the pool. Trying to ‘read’ my freshly picked tiles, the subjects spent extra effort interpreting the Nabaztag’s emote.

7. DISCUSSION

People do not always have just one interpretation for each emotion. Some emotions are hard to distinguish if they have to put to words. The pre- and post-scenario test results would have been very different if for example ‘love - affection’ and ‘love - attracted-to’ and ‘distress – sad’ and ‘distress – unhappy’ would count as the same score. Many times, the difference between the pre- and post test was for example ‘satisfaction’ and ‘relief’. Another set of emotions, or another selection, might have given better results.

In the questionnaire a selection of answers had to picked from “the Rational 22”. Selecting too many would give useless results for usage during the second stage of the research. Selecting too few would give people not enough options and the feeling ‘the right answer is not there’. The final selection was based on emotions people were likely to use in the second stage on the research. Even the interpretations of people when reading for instance “Distress – shock” can vary. Some picture a situation where you are concentrated working and someone

---

2 Translated from Dutch to English.

3 During my research on subject, I used the Dutch ‘NL-Renate voice’
knocks on the door, other people picture their beloved ones sliced to pieces beside them when they wake up. The scenario play was not done in the same place in all cases. To guarantee usable results, scenarios were created to duplicate as many factors as possible, in the different cases. But during the research, it was noticed, even the weather can influence people’s perception on the beeps. Changing the environment altering the subjects experience and interpretation of the emotions. (while filling out the questionnaire).

8. FUTURE WORK
Communicating emotional states using beeps can be applied to other ECA in order to say something about having ECA communicating using beeps in general.

Further research can be done on the Nabaztag. Discover its ease-of-use in little research projects like this.

All the participants had a positive response to the ears of the Nabaztag. (These ears were not used for communication during the sounding of the beeps, to not influence the participant’s perception.) Research on the position of the ears, in combination with sounds or beeps could be done.

People showed interest in the difference of the ‘happy’ and ‘happy-for’ emot. (same for sad, and sad-for) This difference could help the understanding of the ECA by its user. Is the ECA commenting on events (sad-for) or is the ECA simply communicating its emotional state (sad, because he doesn’t get any attention). Being able to understand this difference is valuable when it is an ECA’s function to judge a user’s actions and give feedback (To educate or help the user). What emot is a response to user’s actions and what emot is just an update of the ECA’s emotional-status?

9. CONCLUSION
More and more social ECAs are developed to assist us with the different tasks in our daily live. In the future we might be dependent on the quality of conversations we have with these ECAs. One of the most important issues is how to express the ECAs minds to a user for communication between them. This issue is related to the expressed information and to the ECAs appearance. In his research, a Nabaztag proved to be a right embodiment for a beeping ECA. Using ‘appropriate’ emotions during a series of scenarios with a group participants, the participants trusted and were very much engaged with the ECA. The results were less positive when ‘random’ beeps were made to express the ECAs mind.

The answers to the research questions as stated in the beginning of the paper.
R1. Will R2D2-like appropriate beeping sounds help in establishing an emotional relationship between the elderly person and an ECA?
Yes, these beeping sounds will. From the data in Table 3, we can conclude ‘appropriate’ beeps are more engaging and trustworthy and believable as ‘emotions’ then ‘random’ beeps. However, from the pre-scenario and post-scenario tests, the same results were there, (printed in Table 4) but these results were less conclusive.

R2. Will a Nabaztag linked to a personal-server with home-brew software create a usable believable relational-ECA?
No, the home-brew server is currently too unstable and too slow to use for this purpose. However, when using the Violet server to control the Nabaztag, the Nabaztag performed well as a believable, trustworthy ECA.

R3. How will the subject perceive working with the ECA?
They liked it. Although the reactions were sometimes perceived as inconclusive, this did not really matter. “Isn’t this even more the case in real human-to-human communication” one participant answered.

R4. How will the SE2PM principle applied to the Nabaztag using beeps to communicate work out?
Upgrading to slightly more complex beeps than Komatsu used, and using an ECA with some relational features and more expressed behavior, the ECA was able to communicate its simple mind

ACKNOWLEDGMENTS
I thank all the participants who kindly spend some time on filling out my questionnaire. I especially thank the participants for my scenario test. Also I thank everyone who reviewed and therefore improved my paper. Last but not least I thank E.M.A.G. van Dijk and Heylen, D.K.J. for the research and their insightful comments. I thank the University of Twente to give me the opportunity to work with the Nabaztag and use it for my experiments. Nabaztag™ is a trademark of Violet™.

REFERENCES


APPENDIX A: IMPRESSION OF THE QUESTIONNAIRE

Figure 3. Impression of the questionnaire's introduction

Figure 4. Impression of the questionnaire's questions