A picture is worth a thousand words..Is it?

Deborah Oosting

University of Twente - Enschede

In partial fulfillment of the requirements for the bachelor of psychology at the University of Twente

Matthijs Noordzij & Martin Schmettow

2 -2 -2012

Abstract

All participants worked with the computer program Tribe Talk to create the same sentences with pictograms. The only difference was what the experimenter told them the program was for: to communicate with a household robot, for people with aphasia, or for a new phone application. The goal was to see whether differences in context would yield differences in usability comments, sentence completion time and electrodermal activity. All participants were able to learn to work with the computer program. Differences were found in usability comments, but not in sentence completion time or electrodermal activity. A learning effect was found for electrodernal activity.

A picture is worth a thousand words...Is it?

A lot of research has already been conducted on the subject of usability. Usability is used here for describing computer software programs, and can be explained as a combination of user friendliness of a computer program and its usefulness (Bevan, 1995). By measuring these factors, it is possible to tell how much effort it takes to learn to work with a certain program. However, according to Bevan (1995), user friendliness and usefulness of a program are not enough to describe usability, the context in which working with a computer program takes place is also important for defining this. People will give other comments about a program its usability when the context in which they have to use a program changes. For example, a software program for advanced Photoshop users would be completely worthless to someone who has never worked with Photoshop before, however an expert might really appreciate the program because he or she knows how to work with it (Bevan, 1995). Context then, exist of differences between users, differences in tasks, but also in different environment in which the tools will be used. One program does not have the same usability in different environments (Morris & Dillon, 1996).

Usability for a product is still mostly measured in ease of use of a program, and not on the context it is used in. According to Macleod (1994), context should be taken into consideration more when usability is defined. Context is an important factor for learning to work with a program, because it is useful for evaluating a computer program's usability thus improving it (Macleod, 1994). This is especially useful because working with computers is getting more and more important, not only for personal recreation but also for business and work related goals. People have to be more and more capable of working with computers in their everyday lives, and decent computer software, adjusted to their capability is required to be able to do this (for example, elder people might need easier software to work with). Even though context is more and more perceived as a usable factor for evaluating and revising software programs, the amount of research about how context influences usability is still limited, certainly if one takes the notice that context can have several meanings as stated before (Bevan & MacLeod, 1994; Macleod, 1994). Could the perceived usability be influenced purely because the same computer program is presented in different contexts?

Tribe Talk is a software application created by Symbolic Systems to create short sentences in English. It is a new application that was originally created to serve social media purposes. However, its usability has not been tested yet, nor has it a specific user group yet. Because of this, Tribe Talk could serve perfectly for testing the influence of context on usability, while at the same time finding a possible user group for this program as well. What would be possible user groups?

Tribe Talk has as its main purpose to create short and to the point sentences. It does so by restricting human speech to 'only' 1500 words, divided into several categories. There are only three possible sentence types that can be created with Tribe Talk (further explanation will be given in the methods section below). Besides that, Tribe Talk has correct spelling and verb inflection goes almost automatically. Given these facts, one could say that Tribe Talk puts a filter on human speech. There are at least two possible groups for whom these characteristics could work really well, and these are explained in the next paragraphs.

Lately, there has been quite a discussion considering human robot technology, especially about language programming. A lot of programming problems exist because of the ambiguity in human language. One word can have several meanings, and sentences can be interpreted in several ways because of punctuation and human intonation. When sentences or words can be interpreted in several ways, there are several possible commands for a robot to execute. If one takes for example the command '*boil the water in the plastic bowl'*. This can have the meaning to boil the water that's inside the plastic bowl, or to boil the water while it is in the plastic bowl. Executing the last option might end up having dangerous consequences,

both for the robot as for the person who commands the robot (Jusoh & Ma'azer al Fawareh, 2008)
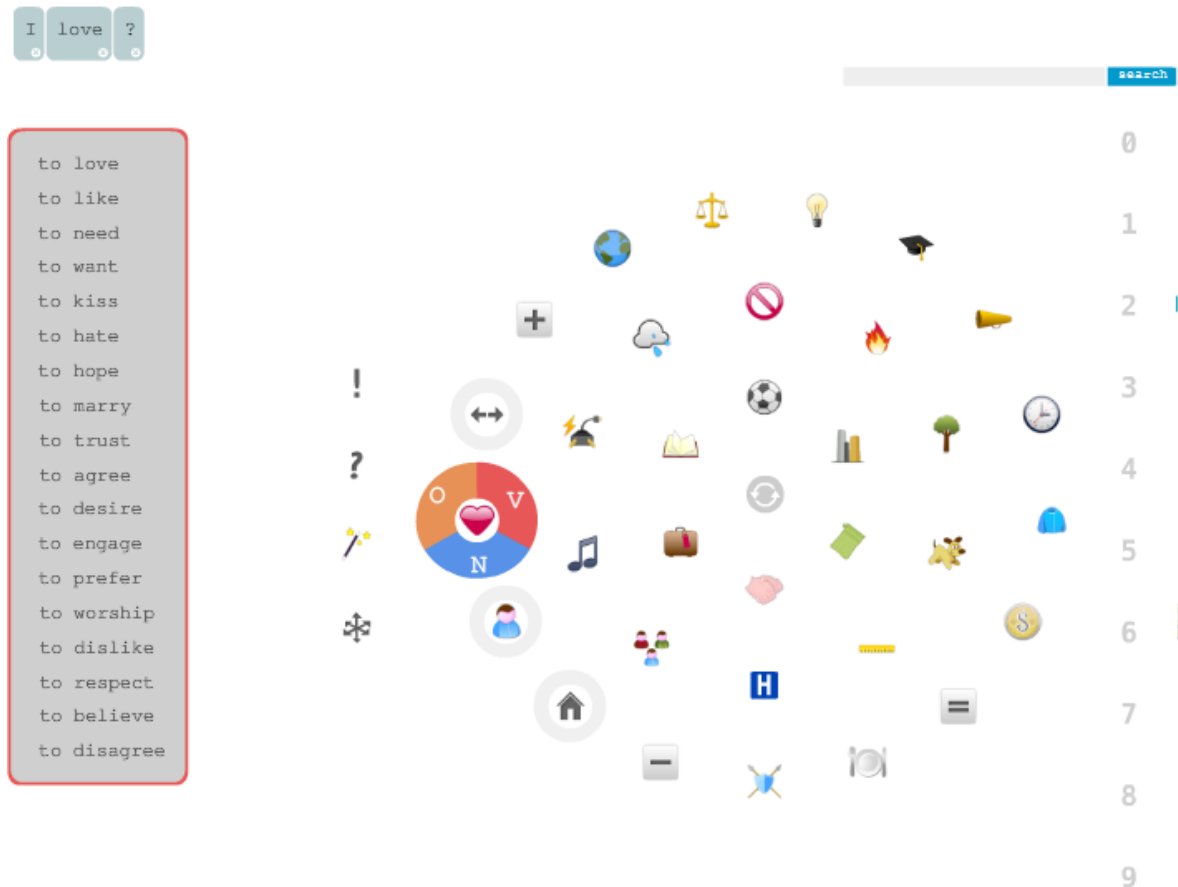
Figure 1. Tribe Talk: by clicking on the heart symbol a circle pops up around it. From this circle, one can choose whether one wants a noun (N), a verb (V) or a word from another category (O). If one clicks on a word, it appears in the left upper corner. This way, a diversity of sentences can be made. Further details will be provided in the methods section below.

In other words: human speech is vague, while robots need direct and precise commands to execute their tasks properly (Wang, Jusoh, & Yang, 2006). Several studies have tried to solve the problem of ambiguity in robot language programming. Possible solutions were to create a system inside the robot that could transform human language in plain pre programmed commands. This could be done by creating a nodes system between categories of words. That way, it would be easier for a robot to "know" what the most logic interpretation of a ambiguous sentence would be at any given time (Jusoh & Ma'azer al Fawareh, 2008). According to Roy, Hsiao and Mavridis (2004), a solution to ambiguity would be to implement mental imagery in a robot. This could be done by programming a set of representations and procedures. Another solution could be to create categories of words which would make it easier for the robot to choose the meaning of a word in a certain sentence (Jusoh & Ma'azer al Fawareh, 2008). Yet another research claims that the solution would be by programming a set of representations, and let the robot complete possible missing pieces by using the internet. This way, several methodologies are used to create a language system (Beetz, Jain, Mösenlechner, & Tenorth, 2010).

The problem with a lot of the current research concerning human robot technology is that the emphasis is on the robot itself, and not the user. A lot of research claims that robots should be programmed in such a way that humans can communicate with them without any

language restrictions (Jusoh & Ma'azer al Fawareh, 2008). The realization is there that not everything can be programmed in a robot. The robot should be able to learn things itself, for example, rules of context (Iwahashi, 2007; Roy, Hsiao, & Mavridis, 2004). Yet, with the current pace of technology it's not possible to create and implement such a complicated learning system in a robot in a short term (Steels, 2003). This is problematic because a lot of robots are used for healthcare, which has the elderly as its main target group. According to *Statistics Netherlands* (CBS), there will be a higher percentage of elderly people among the population in the next 50 years (CBS, 2010). At the same time, there will be a shortage of nurses (Broadbent, Kuo, Lee, Rabindran, & Kerse, 2010). Robots would be able to do healthcare-related tasks, but there'll has to be a proper language system for them to do so.

Instead of trying to program a complete language learning system into a robot, one could also use Tribe Talk. Tribe Talk puts a filter on human speech. Sentences made with Tribe Talk are rather short, and the words used are already placed in categories. The focus would be on placing a filter on human speech this way, so a robot could be programmed just Instead of trying to program a complete language learning system into a robot, one could also use Tribe Talk. Tribe Talk puts a filter on human speech. Sentences made with Tribe Talk are rather short, and the words used are already placed in categories. The focus would be on placing a filter on human speech this way, so a robot could be programmed just to understand the categories and the words belonging to those categories in Tribe Talk, and not to learn an entire language system.

Another group that could do with less ambiguity in language is the group of people with aphasia. Aphasics have severe difficulties in producing and/or understanding language, which is caused by brain damage. A study from Brennan, Worrall and McKenna (2005) showed that aphasics can comprehend language much better if this is presented in an aphasic-friendly way. This means that aphasics comprehend language better if sentences are short, with easy words, and lots of white space between those words. Besides that, sentences should be written in active form rather than in passive. Pictures also have a positive effect on comprehension, but only in addition to other aphasia-friendly constructs. Another experiment showed that aphasics are able to learn to speak trough symbol cards. Even people with severe aphasia were able to learn at least the basic level of this communication medium (Gardner, Zurif, & Berry, 1976). Rabiee, Sloper and Beresford found in their research that children and young adults who don't use speech for communication like to use a symbol system to communicate with people around them. Tribe Talk is meant for producing short sentences, and those sentences are in active form. A lot of words in Tribe Talk are relatively short and easy: there are hardly any synonyms. Tribe Talk also uses pictograms to categorize words. Tribe Talk thus could be used for aphasics to communicate and express themselves, but also for relatives from aphasics to create sentences that are easier to understand and to react on for the person with aphasia.

A third possibility would be that Tribe Talk is usable for social media purposes. This was originally the intended goal for Tribe Talk. However, a small pilot test with 5 persons showed that while people think Tribe Talk is a fun way of communicating, they don't think Tribe Talk is usable for social media purposes. Arguments were that *Facebook* and *Twitter* are faster ways to update the world about one's life. Because a purpose for social media was the original intention of Tribe Talk, it will still be included as a possible user group, even if it is just for comparison: most people are relatively unfamiliar with communicating with either a robot or a person with aphasia, nor do most people have aphasia themselves. More people are, on the other hand, quite familiar with using social media sites or systems (e. g. *Facebook* or *Twitter*) to communicate with other people. Using the social media as a possible user group, it is possible to see whether the context being new would make a difference.

Whatever Tribe Talk might be used for, the program itself doesn't change, only the context in which it is placed does. As Bevan (1995) already stated, context makes a difference when people have to judge a product for its usability (Bevan, 1995).

Within this research, it would be important to define the impact of context on the perceived usability people have from a software program, while possibly finding a specific potential user group for the program Tribe Talk as well. Connecting to the literature above, the research questions are the following:

H1:   Do users detect different usability problems in Tribe Talk when tasks are the same, but context is different?

As noted earlier, context influences perceived usability of a product (Bevan, 1995). What is of particular interest here is whether people give different comments on Tribe Talk according to the context they know.

H2:   Is task completion time for 10 sentences faster/slower depending on a difference in context?

According to Nielsen (1993), usability can also be measured through speed of task completion time (Nielsen, 1993). However, there's more to this question than just measuring usability by time. Research by Nomura, Kanda and Suzuki (2006) showed that people will put less effort in communication if they dislike robots. Other research also showed that attitude is one of the highest predictors of robot communication, stating that people will be less interested, thus les engaged and slower, when their attitude towards robots is negative (Broadbent, et al., 2010; Gieselmann & Stenneken, 2006; Nomura, Kanda, & Suzuku, 2004, 2008). Also, people who do not like to communicate through social media sites will not take effort to do so, hence they will be slower. This question will not just measure usability simply by completion time, but also usability by people's attitude towards the context in which Tribe Talk is placed.

H3:   Are there differences in task engagement, as measured through galvanic skin response (GSR) between the possible user groups of robot, aphasia, and social media?

With galvanic skin response (GSR), one measures the change in skin resistance to electrical conduction. Activation of the sympathetic nervous system changes this level of resistance, which is generally seen as a way to measure arousal, related to both emotion and attention (Vertrugno, Liguori, & Cortelli, 2003). Research showed that if participants play a video game, their level of GSR increases when the video game becomes more difficult (Mohammad & Nishida, 2010). In other words, when they have to pay more attention to the task they're performing. However, an increase in GSR could also mean that the person being measured experiences frustration or stress (Lin, Omata, & Imamiya, 2005). Research showed that (task related) frustration could be identified (out of 6 emotions) 77% of the time using galvanic skin response (Mower, Feil-Seifer, & Matari´c, 2007). GSR results thus, should be interpreted with care. In this experiment, GSR will be used to measure engagement in a task. Because (as stated above) the robot context and the aphasics context will probably be new for people, that in itself could trigger a higher engagement in a task. However, attitude will also be measured with questionnaires to make sure engagement is based on attitude or just on the context itself being new.

Methods

Participants

The study consisted of 30 participants, 10 in each condition. There were 20 female participants and 10 male, with an average age of 22.4 (ranging from 17 to 50). From these

participants, 17 were Dutch. The other 12 were German. Most of the participants studied psychology (18), two participants studied communication sciences, six did another study (not at the university of Twente), and four didn't do a study at the time of participation.

Apparatus

Tribe Talk is a software application created by Symbolic Systems, to create short sentences in English. This can be done by clicking on one of the pictograms. If one does so, a circle appears around the pictogram from which one can choose whether one wants to use a noun, verb, or a word that does not belong to either one of those categories ('other'). By clicking on a category, a whole list of words appears on the left side of the screen. One can now choose the desired word from the list. By clicking on it, it appears in the left upper corner of the screen. Verb inflection and plurals can be created as desired by clicking on words once they're in the left upper corner. Verb inflection, however, goes semi-automatically. There is a search function for finding words in the right upper corner.

Three different types of sentences can be made with Tribe Talk. These are assertions, questions and commands. For an assertion ('she is ok'), no other icon has to be clicked because it is the default sentence type in Tribe Talk. For creating a question ('who is he?'), one has to click the question mark, and for creating a command, also known as an imperative statement ('Give her a pizza!'), one has to click the exclamation mark.
There are plus minus 1500 words programmed in Tribe Talk. By putting words together as described above, and choosing a sentence type, one can create a variety of short sentences. Tribe Talk has correct spelling because all words are preprogrammed. Verb inflection goes almost automatically (with few exceptions, e.g. plural forms).

For 23 participants, the study took place at a lab at the university. Here, a 15 inch laptop with internet was used for working with Tribe Talk. Four participants also worked with this laptop, but at another location. For the remaining three participants the study took place at the participants house, on their own computer.

For measuring galvanic skin response, an Affectiva Q Sensor was used. This device measures electrodermal activity across the skin. It also measures body temperature and movement (for more information see also www.affectiva.com/q-sensor). Electrodermal activity, also known as Galvanic Skin response, was measured both in amount of skin conductance responses (*SCR*) per minute and amplitude of this *SCR* amount per minute. GSR data results were collected in a data file on a laptop. Task completion time was measured with a stopwatch.

For measuring overall usability of Tribe Talk, items from the Software Usability Measurement Inventory (SUMI) were used , a questionnaire with 50 items to rate software. SUMI is mentioned in the ISO 9241 standard as a recognized method of testing user satisfaction. There was also a short questionnaires consisting of items measuring participants' attitude toward social media (in the social media context), toward robots (human robot communication context), and towards aphasia/people with a handicap (aphasics context). Items for the robot questionnaire were be retrieved from the Negative Attitude Towards Robots Scale (NARS) and the Robots Anxiety Scale (RAS) (Nomura, et al., 2008).

Design

There were three conditions: 1) human robot communication condition 2) aphasics condition and 3) social media condition. Every person had to create the same sentences in the same order. Persons were assigned to a condition at random, until every condition had 10 participants.

Procedure

First, participants had to get wear the Q Sensor. Participants then got an explanation according to the condition (context) they were placed in. Participants in the robot condition heard that Tribe talk might be a new way to make communication with robots easier, because it is not yet possible to program an entire language system into a robot. They were told that that is important because there will be a lot of older people in the future, who'll need healthcare, which robots might be able to provide if human robot communication improves. Participants in the aphasics context heard that Tribe Talk might be good for training communication skills for people who have aphasia (psychology students will know what aphasia is because it belongs to the subjects they had to learn). They also heard that this might be a brand new way to communicate, not only for the people who have aphasia themselves, but also for relatives. Participants in the social media condition were told that Tribe Talk is a possible new candidate for a new phone application. This way, text messages won't be as confusing as they can be, and ambiguity through phone communication will be reduced.

All participants were told that the Q Sensor measured engagement to a task. After they heard their stories, participants got an informed consent form which they had to fill in. When they had done this, they got a questionnaire in which they had to fill in personal data (e.g. age, sex, nationality). After that, participants got to practice with Tribe Talk for five minutes. When they tried Tribe Talk for five minutes, they had to try and make 2 practice sentences.

Do you want me to walk the dog?
Please explain what you want!

After participants finished these two sentences, they got 10 real sentences to create on the screen. Between sentences 3 and 4 and the sentences 6 and 7 there was a two minute break. These breaks were included to give the participant a little rest and to find the level of galvanic skin response when the participant was not working on completing the sentences (baseline level).

She didn't smoke before
He is selfish
Please bring me my keys!
-2 minute break-
Offer them a drink from me!
Dial 112 fast!
I prefer pizza for dinner!
-2 minute break-
What is her name?
There isn't any music?
How much is that in kilograms?
Can you show me the way to the supermarket?

Completion time was measured for every sentence. If it took participants longer than four minutes to complete a sentence, they were told to skip that sentence and continue with the next one. This was done for multiple reasons: first because the user groups mentioned above would also have to be able to make sentences in a short amount of time (e.g. one has to be able to command a robot in less than four minutes for such a system to be convenient. Communication for aphasics would only work well if this would be faster than one sentence per four minutes, and social media in general is meant to be fast), but also because the experiment needed a time limit set to it (for this experiment a maximum of 60 minutes in total).

After completing the last sentence, participants were given another questionnaire they had to fill in. This questionnaire contained items from the SUMI, two open questions for positive and negative comments about working with Tribe Talk, and items concerning participants attitude toward either robots, aphasia, or social media according to the condition the participant was placed in. The participant was debriefed after filling in this questionnaire. If the participant liked, he or she could fill in an e-mail address on a list to receive the final report.

Data analysis
For all initial statistical tests performed, the independent variable was condition. The dependent variables were task completion time, amount of skin conductance responses (*SCR*s) per minute and amplitude of *SCR*s per minute.

For the second hypothesis, a repeated measures ANOVA was used first because a data plot showed that there is not a great difference in the means per condition, and a calculation showed that there is neither much difference in variances (see also Appendix 2, page 21 for this data plot) ANOVA is normally only used on data with a normal distribution, because it is really robust against variance and it assumes homogeneity of variance. From plotting the data per condition it was visible that this data had a gamma distribution. However, the plot also showed that there were no differences between conditions, nor are there big differences in variance (thus, homogeneity of variance). In this repeated measures ANOVA, the factor was time and the between-subjects factor was condition.

For the third hypothesis, both a one way ANOVA and a repeated measures ANOVA were used. In the repeated measures ANOVA, sentence served as a factor and condition as a between-subjects factor. Both amount of *SCR*s per minute and amplitude of these *SCR*s were taken into account. Skin conductance response (SCR) is the phasic increase in skin conductance and lays between 0.01 to 1.0 Micro Siemens (μS), measured 1-3 seconds after a stimulus onset. Minimum values mostly used are between 0.01-0.05 μS . The amount of *SCR*s is said to increase when a person experiences frustration, or has to pay more attention. The amplitude of *SCR*s measures the amount of increase in conductance (Dawson, Schell, & Filion, 1990). In this study, peaks lower than 0.01 μS were left out of the initial results, thus calculations. This way, data from four participants was left out of the calculations for the third hypothesis. Two of these participants were in the media condition, one in the robot condition and the other one in the aphasia condition. The data from another participant in the robot condition was omitted as well because the Qsensor data was incomplete for this participant.

Results

User comments
Participants wrote down positive as well as negative aspects they noticed while working with Tribe Talk. This resulted in a total of 93 comments, of which 50 were positive and 43 negative. To analyze these comments, the article of Leech and Onwuebuzie (2007) about qualitative data analysis was used (Leech & Onwuegbuzie, 2007). First, word count was used to find the words that were used the most. These words were '*fun*'(15 x), '*to find*'(15 x). '*searching/searching engine*' (17 x), '*category*' (13 x) and '*easy*' (10x). Based on these words being the most used, categories were set up to divide all comments. Five of these categories were for positive comments about Tribe Talk, the other five were for negative comments. The researcher and a cooperating master student of psychology placed every comment in one of these categories, reaching a Cohen's Kappa of .92. Comments with interrater disagreement were left out of the analysis. After doing so, there were 46 positive comments left and 40 negative comments left.

After this, a Chi-square test was done to see if there were significant differences in the number of positive and negative comments per condition. This effect was not found (χ^2 (2) = 0.422, $p = 0.810$) (See also appendix A on page 17 and further for the complete list of comments given by participants). However, because this is quite an explorative study not only about the influence of context on usability but also about the program Tribe Talk itself, a few points are worth being mentioned:

In the positive comments, 41.3% of all comments (19 out of 46) were about the program being new, or like a game. In the negative comments, 44.2% of all comments (19 out of 40) were about the words being illogically categorized to the symbols, or the symbols being too vague in general. Negative comments about the program being too slow came mostly from the people in the media condition (4 out of 8 comments regarding the speed of the program). These people stated that they would rather use the search engine because this was faster than looking up the words via pictograms. The other two comments were from the robot and the aphasia conditions. People in the media category paid more attention to what the pictograms themselves mean, while people in both robot and aphasia condition paid more attention to under which pictogram a word was located. This was noticeable because people in the media category complained more about the pictograms themselves being vague, while people in the media and the robot condition complained about the words being illogically placed under a pictogram.

Attitude

Participants were also asked some questions concerning their attitude towards the category they were placed in. This measure was used as a covariate for task completion time (see task completion time below). As expected, a striking difference was found between the condition being new to the participants. While almost every participant in the social media condition (9 out of 10) claimed to have at least one account on *Facebook, Twitter, Myspace* or *Hyves*, only 3 out of 10 participants in the robot condition said that they had ever seen a robot. From all participants in the aphasia condition, no one had ever met someone with aphasia.

Table 1
Means and standard deviations for each attitude subtest per condition

Subtest	Condition	*M*	*SD*	Max. score
General TribeTalk attitude	Robot	11.50	0.71	12
	Aphasia	10.50	2.37	12
	Media	10.10	1.17	12
Learnability	Robot	10.30	1.25	12
	Aphasia	9.70	2.06	12
	Media	10.60	1.58	12
Purpose of Tribe Talk	Robot	10.10	1.37	12
	Aphasia	9.50	1.84	12
	Media	7.90	1.20	12
Appropriate for condition	Robot	2.80	0.42	4
	Aphasia	2.90	0.32	4
	Media	2.50	0.53	4

Attitude was also measured for Tribe Talk. This attitude-test was divided into four subtests. These were 1) the overall attitude towards Tribe Talk in general 2) Attitude towards the learnability of Tribe Talk in general and 3) Attitude towards the purpose of Tribe Talk in

general (see table 1 below for the means and SD per condition). The highest obtainable score for these attitude tests was 12. A fourth item measured whether participants found this program appropriate for the condition they were put in. The maximum score for this item was 4. A higher score predicted a more positive attitude.

A one way ANOVA was performed to see whether these differences in means were significant. One effect was found for condition on the attitude test about the purpose of Tribe Talk ($F(2,27) = 5.791$, $p = 0.008$. A post-hoc Bonferroni test showed that this attitude was significantly more positive in the robot condition than in the social media condition ($CI = 0.49$ - 3.91). No effects were found in the other subtests(all p's > 0.05).

Task completion time

Calculated with a repeated measures analysis in SPSS, it was found that there are no significant differences in task completion time of all 10 sentences between the robot condition ($M = 77,4$. $SD = 19.2$), aphasia condition ($M = 66.25$, $SD = 24.5$) and the social media condition ($M = 67.5$, $SD = 15.2$) ($F (9,1) = 0.430$, p = 0.918). It also showed that there were neither differences in task completion time per sentence. (all p's > 0.05) Using age, study, gender, nationality and level of English as covariates did not show any effects either (all p's > 0.05). A final test yielded that there are no effects of attitude on task completion time. This was true for attitude towards human robot interaction, aphasia, and social media (all p's > 0.05). Another test showed that there were some time effects per sentence, but not condition (see table 2 below). Because these differences could be based on a lot of factors not accounted for in this study (e.g. sentence length, specific difficulties within sentences), they won't be mentioned here.

Table 2
Task completion time in seconds per sentence

Sentence	Minimum	Maximum	*M*	*SD*
1	24	165	56,10	34,764
2	12	69	24,63	12,893
3	38	240	103,77	45,636
4	31	132	70,33	23,241
5	32	157	74,10	34,380
6	39	131	61,00	22,133
7	12	120	41,53	21,323
8	19	105	49,90	21,667
9	38	152	95,23	34,109
10	41	240	127,83	56,007

Electrodermal activity
Amount of skin conductant responses (SCRs)

A one way ANOVA was used with condition as a factor and the average amount of *SCRs* per sentence as the dependent variable. This number was corrected for the baseline level of *SCRs* participants had when they weren't working on sentences (sentence *SCRs* – baseline *SCRs*). While the means of the data showed slight differences ($M_{robot} = 2.80$, $SD_{robot} = 2.41$, $M_{aphasia} = 1.51$, $SD_{aphasia} = 2.31$, $M_{media} = 1.85$, $SD_{media} = 2.78$), ANOVA showed there were no effects from condition on the amount of *SCRs* ($F(2, 26) = 0.666$, $p = 0.522$).
Amplitude of SCRs

The same correction for baseline made for measuring the amount of *SCRs* was also made here (sentence *SCR* amplitude – baseline *SCR* amplitude). The means per condition

showed that these were for both robot and social media condition below baseline SRCs level (M_{robot} = -0.0030, SD_{robot} = 0.0893 , M_{media} = -0.0022, SD_{media} = 0.02735) but the mean from the aphasia condition was above baseline($M_{aphasia}$ = 0.0552 , $SD_{aphasia}$= 0.05269).. However, a one way ANOVA made clear that there were no significant differences in amount of SCRs per condition ($F(2,26)$ = 3.233, p = 0.06).

Learning effects amount of SCRs

A repeated measures ANOVA test was used to find out whether there was an effect of condition on the amount of SCRs per sentence. The means showed slight differences between the robot condition (Mr = 3.519, SDr = 1.186), the aphasia condition (Ma = 4.083, SDa = 1.125) and the social media condition (Mm =3.519, SDm = 1.125). Mauchly's test showed that the sphericity assumption was violated ($W(44)$ = 0.07, $p < 0.001$). Because of this the Greenhouse-Geisser correction for degrees of freedom was used. The linear and cubic contrasts were also significant. ($F(1, 26)$ = 13.210, p = 0.001) and ($F(1, 26)$ = 11.233, p = 0.002.), respectively (See also graph 1 below). There were no differences per condition on the amount of SCRs per sentence ($F(26,2)$ = 0.064, p = 0.938). Neither was there an interaction effect from sentence and condition on the amount of SCRs per minute ($F($ 7.89, 102.51$)$ = 0.690, p = 0.697).

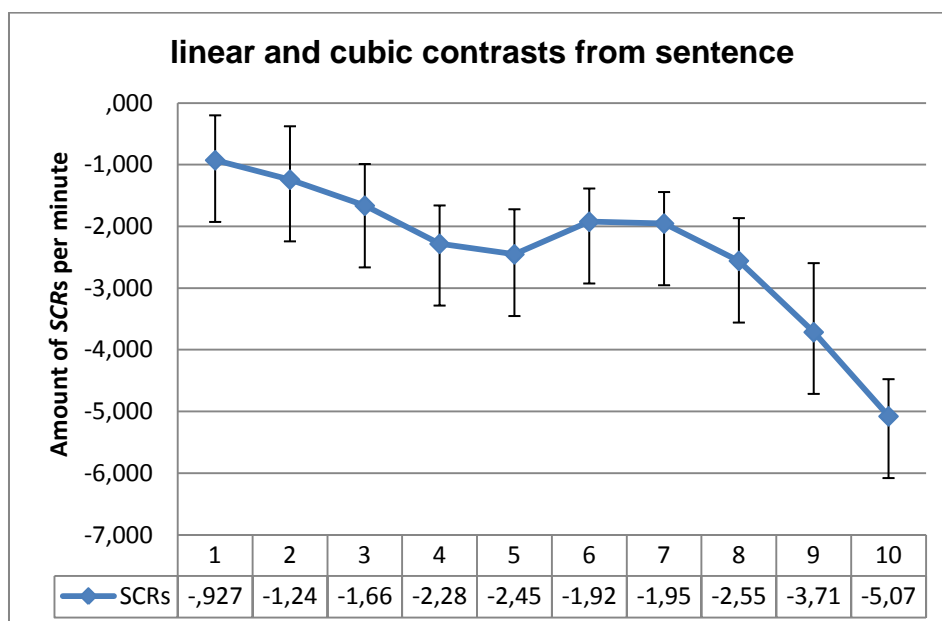| linear and cubic contrasts from sentence									
	1	2	3	4	5	6	7	8	9	10
SCRs	-,927	-1,24	-1,66	-2,28	-2,45	-1,92	-1,95	-2,55	-3,71	-5,07

Figure 2: Learning effects amplitude *SCR*s The amount of *SCR*s per minute is calculated as sentence *SCR*s - baseline *SCR*s (hence negative values)

The same repeated measures ANOVA was done to measure the effect of condition on the amplitude of the SCRs per sentence. the condition on the amplitude of the *SCRs* per sentence. There were slight differences in the means of *SCR* amplitude per condition (M_{robot} = 0.095, SD_{robot} = 0.021, $M_{aphasia}$ = 0.068, $SD_{aphasia}$ = 0.020, M_{media} = 0.060, SD_{media} = 0.020). However, there were no significant effects from sentence or condition, nor an interaction effect from sentence and condition on the amplitude of *SCR*s per minute. (all p's > 0.05).

Discussion

The expectation from this experiment was that differences would be found in usability between the three different contexts. While some of these different were found indeed, a lot of them were not. An attitude test about Tribe Talk showed that there was one significant difference in attitude between conditions; people in the robot condition had a more positive

attitude towards the purpose of Tribe Talk then people in the media condition. This seems rather strange because the overall attitude towards Tribe Talk was positive for all three conditions. However, a pilot test before the actual experiment already showed that people liked to work with Tribe Talk, but did not think it would be appropriate for social media purposes. This also shows in the comments given by participants in the media contexts stating that Tribe Talk is too slow for a decent social media purpose (further possible explanation is given below).

While there weren't any differences between the conditions in both amount or amplitude of *SCR*s, there turned out to be an effect of sentence in the amount of *SCR*s. This was combined with both a linear and a cubic contrast (as seen in figure 2 on page 11). It is very visible that the amount of *SCR*s declines with each sentence (except for sentences 6 and 7, but those sentences probably account for the cubic contrast). It is well known that presenting a stimulus to a several times causes habituation, and this effect also exists for electrodermal activity (Leonard & Winokur, 1963). This habituation effect, or learning effect, is visible in this experiment. It is also in agreement with Mohammad (2010), who's article states that the amount of *SCR*s increases when one has to pay more attention to a task (Mohammad & Nishida, 2010). If the amount of *SCR*s lowers with each sentence, this means one has to pay less attention when working with Tribe Talk for a longer time. This implicates that one can get better at working with Tribe Talk, even though there is no visible learning effect in task completion time. It's also important to note that this effect is independent of condition; One could learn to work with Tribe Talk no matter what condition one's placed in.

Any other expected differences were not found, or were not significant. The first expected difference would be in usability comments. While there were no significant differences here between the actual number of positive and the number of negative comments, a few things can be said about the participants in different conditions giving different kinds of usability comments: most comments about the program being too slow came from the media condition. Another difference was found in the comments about the pictograms themselves. People in the media condition mentioned the pictograms themselves being too vague, while people in both the robot and aphasia condition mentioned the categorizing of words more. Bevan (1995) explained that the context the user experiences, can influence what usability comments this user will give. This happened in this experiment as well. So why is it then that people in the media condition complained more about the program being slow than people in both the robot and the media condition?

The first explanation is that people in the robot and aphasia condition were told that this program is for a certain user group, and these user groups (older people in the robot condition, aphasics and their families in the aphasia condition) will be using this program for a longer period of time. People in the media condition, however, were told that Tribe Talk was mostly for reducing ambiguity in text messages. It was told that the elderly would probably benefit the most from this system. However, the condition of robot and aphasia could be seen as more life invading: that is, a household robot would be there every day, and someone with aphasia and his or her family would be using this system every day, while a text message does not have this importance. Not only do people adapt their learning strategies to context in general, but also specifically to learning context (Govender, 2009). So if someone learns to work with Tribe Talk with the intention that one will use this program over a longer amount of time (like in the robot and aphasia condition), one might be more patient to learn to work with it than someone who will only use this program for short text messages once in a while (social media condition). Also, social media has a purpose of being fast in general, so people will expect an application for social media to be fast (as said in the introduction, a pilot test already showed this). Because expectation is also a factor for learning, this will have influence on the user comments as well (Paechter, Maier, & Macher,

2010). The difference between the comments about the pictograms themselves being vague in the media condition versus the words being placed under the pictograms illogically in the robot and aphasia condition can be explained in the same way: people in the robot and aphasia condition might have focused more on were the words themselves are, because their user groups will use this program a lot more intensively than one would use this in a social media context. Another aspect is that participants were told that Tribe Talk would serve as a new application on a phone. Because phones have a tiny screen and social media sites in general use a lot of symbols, these symbols should be clear and understandable. A last explanation might be that participants in the social media condition knew what to expect because they were familiar with social media sites, while the robot and aphasia conditions were new to almost all participants. This might also explain why the subtest about attitude 'purpose of Tribe Talk' was lower for the media condition than for the robot condition (it was also lower than the aphasia condition, but this was a non-significant result).

The second expected difference was in task completion time per condition. No significant differences were found here. This means nothing can be said about one condition being liked better than the other based on task completion time. However, this study also aimed at testing Tribe Talk usability in general. Nielsen (2003) stated that usability can be measured through task completion time. If there are no differences in task completion time, it does not mean that the program itself is not learnable, only that it does not matter in what context a participant gets to work with Tribe Talk (Nielsen, 1993). Also, the means of the attitude test towards condition showed that attitude was positive for all three conditions, so as the attitude test towards Tribe Talk in general. According to Nomura *et al* (2006), a negative attitude predicts a lower effort in communication, but that was not the case here. The results of task completion time don't differ per condition, but they do show how fast people are at using Tribe Talk when they use it with an attitude that is positive in general (Nomura, Kanda, & Suzuku, 2006).

The third expected difference was in electrodermal acticity. This was measured in two ways: the first was measuring the amount of *SCR*s per minute, the second measure concerned the amplitude of these *SCR*s. There were no differences in amount of *SCR*s nor amplitude between conditions. Both baseline amount of *SCR* and task *SCR* were measured. In the robot and the media condition, the average task *SCR* was below baseline. Lin, Omata, Imamiya and Hu (2005) claim that an increase in *SCR*s is a sign of stress or frustration (Lin, et al., 2005). It was expected that because both the robot and aphasia condition were new to participants, this would trigger a higher amount of *SCR*s than in the media condition. This was not the case. Attitude towards Tribe Talk was positive for all three conditions, so was attitude towards the conditions themselves. Also, the amount of task *SCR*s was rather low so participants were not frustrated or stressed while working with Tribe Talk.

Shortcomings in this research

The research itself has some shortcomings. First, the sample of participants was not very large: 10 participants per condition is a very small sample, and future research would be better if there would be more participants in each condition. This is also the case because some of the results measured by the Qsensor were not usable (because the minimum amplitude was below 0.01 μS, as explained in the methods section above), which gave us even less data to work with. Second, it was impossible to measure a learning effect in the sentences by task completion time, because not every sentence was the same length and/or difficulty. In a next research, sentences should be more alike to be able to measure a possible learning effect. Third, while performing the experiment, the experimenter was in the same room as the participant, because the task completion time for each sentence had to be written down. While this was the same for every participant (so there were no differences between the

experimenter being in the same room as the participant or not), it could well be the case that this had an effect on how participants rated Tribe Talk (Baarda, 2009; Falk & Heckman, 2009). In a next research, it might be an option to use a computer program, or to place a camera. A fourth issue was with giving the participant an explanation about Tribe Talk. Because there were problems with the website of Tribe Talk, the manual was put offline, and the explanation was given verbally. While every participant got the same explanation, a paper version or an online version on how to work with Tribe Talk would have been better, because it leaves no room for differences in explanation between participants, thus no differences in results because of this. A fifth and final concern is that our sample of participants consisted of healthy, young students, while our user groups were the elderly and aphasics in two out of three conditions. In a next research it would be better to use a sample that is closer related to the user group, because this might raise other usability problems that are closer related to the user group (e.g. someone with aphasia might have issues with other aspects of Tribe Talk than a regular student has). This will also be important because one's age influences how well one can learn to remember the symbols (Freudenthal, 2001).

Recommendations for improving Tribe Talk

While the major goal of this study was to find out whether context influences usability, the program Tribe Talk was never thoroughly tested for usability nor for possible user groups. After the experiment participants wrote down recommendations for improving Tribe Talk for better usability. The experimenter also took notes at moments participants experienced difficulties with Tribe Talk. A few possible recommendations for improvement are the following: the first comment that was given a lot (in all three conditions) was about some of the symbols being too vague to recognize what words would be placed under them. Another comment here was that words weren't logically categorized under pictograms (e.g. it was mentioned 12 times how strange it is that the word 'key' goes with the music symbol, but there's no 'key' in the list of words going with the house pictogram). A good idea would be to do another research to find out which symbols are seen as vague, and how many words are seen as illogical categorized, and how this would be more logic thus more user-friendly. For the word categorization, one option would be to put those words under another, more logical symbol (e.g. 'key' goes with the house symbol, and not with the music symbol), another option would be to put the word under both symbols (e.g. the word 'key' both under the music symbol and the house symbol). Another issue addressed by the participants was that the searching engine is rather small. Research was conducted on a 15 inch laptop for most of the participants, and they were not able to see what they were typing. The searching engine turned out to be important for initial learning to work with Tribe Talk, so it would be an improvement to at least be able to see what you are typing when you use it. A last comment is that, when you click a symbol, the words in the list that appears on the left side of the screen aren't alphabetically ordered. This makes it hard to find a word because it is easy to overlook it. Another research could be done to see whether one can find the words faster if they are alphabetically categorized.

Conclusion

A few things have to be said about the implications of this research. While a lot of differences were expected to be found considering the influence of contexts on usability, only a few have been found. Maybe this means context, as measured in this experiment, is not as much of an important factor in usability as expected. What is important, however, is whether it is possible in the first place to learn to work with a software program in general, and if people like to work with it, no matter what context one is in. Differences in context are not only about differences in environment in which the tool will be used, as measured here (the

user being practically always a student). They are also about differences between users themselves, and differences in tasks(Morris & Dillon, 1996). This experiment was too small to test the importance of these factors altogether. It is very well possible that in a next experiment, environment might make more of a difference when the chosen participants truly represent the user groups of a possible environments (e.g. elderly and aphasics instead of students), and if tasks are changed according to these environments. Future research will definitely have to take that into account

Literature

Baarda, D. B. d. G., M.P.M & Teunissen, J. (2009). Observeren *Basisboek kwalitatief onderzoek* (Vol. 2). Groningen/Houten: Noordhoff Uitgevers bv.

Beetz, M., Jain, D., Mösenlechner, L., & Tenorth, M. (2010). Towards performing everyday manipulation activities. *Robotics and autonomous systems, 58*, 1085-1096. doi: doi:10.1016/j.robot.2010.05.007

Bevan, N. (1995). *Usability is quality of use.* Paper presented at the 6th International conference on human computer interaction, Yokahama, Japan.

Bevan, N., & MacLeod, M. (1994). Usability measurement in context. *Behavior and information technology, 13*, 132-145. doi: 10.1080/01449299408914592

Broadbent, E., Kuo, I. H., Lee, Y. I., Rabindran, J., & Kerse, N., Stafford, R. & Macdonald, B.A. (2010). Attitudes and reactions to a healthcare robot. *Telemedicine and e-health, 16*(5), 608-613. doi: 0.1089/tmj.2009.0171

CBS. (2010). Tempo vergrijzing loopt op Retrieved May 17th, 2011, from http://www.cbs.nl/NR/rdonlyres/BB0BFB7A-6357-4D2E-92DF-FA706E4EE6E1/0/pb10n083.pdf

Dawson, M. E., Schell, A. M., & Filion, D. L. (1990). The electrodermal system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of Psychophysiology: Physical, social, and inferential elements* (pp. 295-324): Cambridge Press.

Falk, A., & Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science, 326*(5952), 535-538. doi: 10.1126/science.1168244

Freudenthal, D. (2001). Age differences in the performance of information retrieval tasks. *Behaviour & Information Technology, 20*(1), 9-22. doi: 10.1080/01449290110049745

Gardner, H., Zurif, E. B., & Berry, T. B., E. (1976). Visual communication in aphasia. *Neuropsychologia, 14*(3), 275-292. doi: 10.1016/0028-3932(76)90023-3

Gieselmann, P., & Stenneken, P. (2006). Communication with robots: evidence from a web-based experiment om human-computer interaction. *IEEE Computer Society*, 118-121.

Govender, I. (2009). The learning context: Influence on learning to program. *Computers and Education, 53*(4), 1218-1230. doi: 10.1016/j.compedu.2009.06.005

Iwahashi, N. (2007). Robots that learn language: A developmental approach to situated human-robot conversations. In N. Sarkar (Ed.), *Human-robot interaction* (pp. 95-118). Vienna, Austria: Itech Education and Publishing.

Jusoh, S., & Ma'azer al Fawareh, H. (2008, May 27-29, 2008). *An intelligent interface for a housekeeping robot.* Paper presented at the The 5th International Symposium on Mechatronics and it's Applications.

Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis toold: A call for data analysis triangulation. *School Psychology Quarterly, 22*(4), 557-584. doi: 10.1037/1045-3830.22.4.557

Leonard, C., & Winokur, G. (1963). Conditioning versus sensitization in the galvanic skin response. *Journal of Comparative and Physiological Psychology, 56*(1), 169-170.

Lin, T., Omata, M., & Imamiya, A. H., W. (2005, November 23-25). *Do physiological data relate to traditional usability indexes?* Paper presented at the OZCHI, Canberra, Australia.

Macleod, M. (1994). Usability in Context: Improving Quality of Use: Elsevier.

Mohammad, Y., & Nishida, T. (2010). Using physiological signals to detect natural interactive behavior. *Appl Intell, 33*, 79-92. doi: 10.1007/s10489-010-0241-4

Morris, M. G., & Dillon, A. (1996). The Importance of Usability in the Establishment of Organizational Software Standards for End User Computing.

Mower, E., Feil-Seifer, D. J., & Matari´c, M. J. N., S. (2007). *Investigating implicit cues for user state estimation in human-robot interaction using physiological measures.* Paper presented at the 17th EEEE International conferennce on robot & human interactive communication, Jeju, Korea.

Nielsen, J. (1993). Usability Engineering

Nomura, T., Kanda, T., & Suzuku, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI & Sox., 20*, 138-150. doi: 10.1007/s00146-005-0012-7

Nomura, T., Kanda, T., & Suzuku, T. K., K. (2004, 20-22 Sept. 2004). *Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots.* Paper presented at the 13th IEEE International Workshop on Robot and Human Interactive communication, Kurashiki, Okayama Japan.

Nomura, T., Kanda, T., & Suzuku, T. K., K. (2008). Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics, 24*(2), 442-451. doi: 10.1109/TRO.2007.914004

Paechter, M., Maier, B., & Macher, D. (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & Education, 54*(1), 222-229. doi: 10.1016/j.compedu.2009.08.005

Roy, D., Hsiao, K., & Mavridis, N. (2004). Mental imagery for a conversational robot. *IEEE transaction on systems, man, and cybernetics-part B: cybernetics, 34*(3), 1374-1383. doi: 10.1109/TSMCB.2004.823327

Steels, L. (2003). Evolving grounded communication for robots *TRENDS in Cognitive Sciences* (Vol. 7, pp. 308-312).

Vertrugno, R., Liguori, R., & Cortelli, P. M., P. (2003). Sympathic skin response: basic mechanisms and clinical applications. *Clin Auton Res, 13*, 256-270. doi: 10.1007/s10286-003-0107-5

Wang, F., Jusoh, S., & Yang, S. X. (2006). A collaborative behavior-based approach for handling ambiguity, uncertainty, and vangueness in robot natural language interfaces. *Engineering Applications of Articifial Intelligence*(19), 939-951. doi: 10.1016/j.engappai.2006.02.003

A PICTURE IS WORTH A THOUSAND WORDS

Appendix 1: Hypothesis 1, full comments given by participants.

Normal text: robot condition
Cursive text: aphasia condition
Underlined text: media condition
Gray text: interrater disagreement

pro:
I) Nieuwigheid/spelachtigheid van het programma
1.Het is een andere vorm van communicatie op de computer
2.Interessant, nog nooit eerder mee gewerkt
5.Leuke uitdaging om het woord te vinden
6.Leuk om te kijken waar zinnen precies staan, normaal denk je daar niet over na
7.Grappig om op deze manier zinnen samen te stellen
8. Puzzelen
1. and that it is actually possible to communicate with a computer program this way
4.Het is een nieuwe en aparte manier van communiceren.
5. en het is leuk als het is gelukt om de zin helemaal te vormen.
6. Eenmaal bezig is het een leuk spelletje. Gebruik van pictogrammen vond ik leuk.
7.Het is net een puzzel ☺
8. Ik vond het leuk om via de verschillende categorieën zinnen samen te stellen, omdat ik nieuwe associaties met betrekking tot de woorden heb geleerd.
9. Het is leuk als je een woord kan vinden met de pictogrammen ipv met de zoekmachine.
10. Ik vond het spelen wel leuk, maar de pictogrammen moeilijk om te onthouden.
2. Ik vond het leuk om een nieuw communicatiemiddel te gebruiken:
4. Het is een keer wat anders. Pictogrammen zijn altijd leuk.
6. De oefening.
7. Het is een beetje een uitdaging bij het eerste gebruik de woorden te vinden, en daardoor ook juist interessant.
8. je kunt op een speelse manier "zinnen bouwen".
9. het is een andere manier van "communiceren". Het is leuk om te ontdekken hoe je zinnen kunt vormen en waar je woorden kunt vinden.
10. I like the idea of writing something with pictures.
II) Intelligentie van het programma
3) Het programma maakt echt correcte zinnen, en denkt met je mee.
1. I like the amount of words in Tribe Talk
8.Het programma werkt heel systematisch en logisch, daarom doet het ook plezier er mee te werken.
3. de opbouw is goed georganiseerd
7. De woorden onder ?!* en het stafje zijn goed te vinden.
10. It is good that if you make a mistake, you can correct if afterwards.
III) Is het programma goed te leren/geschikt voor de doelgroep?
4.Leuk om te werken met verschillende categorieën.
4. Makkelijker met de categorieën die aan de linkerkant staan.
5.Dat het later sneller gaat omdat je de categorieën beter kent geeft een prettig gevoel, alsof je het programma snapt.
10. Als je het door hebt is het niet moeilijk om te gebruiken.
2. Bovendien kun je altijd nog ergens nieuwe woorden vinden.
3. en heeft mij niet het gevoel gegeven dat ik beperkt werd in het formuleren van de zinnen.
Volgens mij kun je alles met dit programma zeggen en jezelf goed duidelijk maken.

4. Als je je er goed in verdiept is het makkelijk en goed te gebruiken.

2. het was interessant om te zien dat het steeds wat makkelijker werd en sneller ging om de zinnen te formuleren.

5. Het is gemakkelijk.

9. Dat je er steeds beter in wordt naarmate je er langer mee werkt.

2.het was prettig om hele verschillende zinnen te construeren en met de tijd te leren hoe het werkt.

3. Men begrijpt het programma heel snel

5. het is leuk het gevoel te krijgen dat je het door hebt.

8. Je komt er redelijk snel achter hoe je dingen moet doen.

10. When you work with it for a little while, you learn real quickly where the different words are (under which symbols). The more you do it, the faster you get, and can communicate.

IV) Aantrekkelijkheid programma uiterlijkerwijze

4. De symbolen zijn heel mooi.

9. Het ziet er leuk uit

10. Het programma ziet er vrolijk uit, veel gebruik van kleur en dergelijke.

3.Het is best leuk vormgegeven,

3. en het ziet er goed uit.

5. De pictogrammen spreken aan

V) Overige

9), wekt bij mij nieuwsgierigheid op: ik wil graag weten waarom het dan juist zo verdeeld is.

1. Redelijk, het is vooral wennen en zoeken in het begin, dus kan je er niet echt van genieten.

con:
I)Indeling/categorisering/logica scherm/pictogrammen/woordindeling

1.Woordjes zonder groepje (zoals 'please') moeilijk te vinden en te onthouden, omdat het pictogram vrij algemeen is

3. De categorieën zijn onduidelijk soms, en ook niet erg universeel. Cijfers aan de zijkant onduidelijk, en 'search' is gewoon onleesbaar.

4.Soms waren er woorden die volgens mij bij elkaar hoorden (zoals much en any) maar niet bij elkaar stonden.

5.Sommige woorden stonden niet in de categorie hoe ze bedoeld waren (bijv key bij muziek ipv bij het huis), sommige categorieën overlappen daarnaast ook.

6. Sommige woorden, vooral werkwoorden, waren lastig te vinden (bijv can, zin 10)

7.Soms sluit de categorie niet aan bij mijn verwachting. 'Key' bijvoorbeeld, zoek ik sneller bij het huisje dan bij de muzieknoot.

9. Soms is de onderverdeling een beetje vreemd

10.Sommige woorden staan niet op een "logische" plek.

1.To find the correct form of a verb was sometimes not easy

2. Soms was het nog echt moeilijk om de juiste woorden via pictogrammen te vinden

5. Daarnaast staan de woorden in de rij ook niet op alfabetische volgorde, dus ik moest iets langer zoeken.

8. Sommige woorden waren moeilijk in een categorie toe te ordenen, zoals 'key' of 'supermarket'. Verder kan ik sommige werkwoorden zoals 'dial' moeilijk vinden, omdat ik meteen aan een telefoon had gedacht, en niet aan een klok

9. In het begin is het lastig om woorden te vinden, en aan het eind vond ik dat nog steeds bij woorden zoals 'the', 'to', etc.

10. Onduidelijkheid om te pictogrammen te onthouden,

10.Op het scherm staan sommige categorieën niet logisch bij elkaar. Bijvoorbeeld de categorieën "-", "+"en "=", die had ik op het scherm bij elkaar gezet. Misschien waren

rijtjes categorieën van dezelfde soort handiger geweest dan alles in een cirkel. Ik snap het waarom daarvan niet zo goed, want het maakt het gebruik niet makkelijke
1. Het vinden van sommige woorden is lastig, je vergeet soms de juiste vorm aan te klikken omdat je dan op het woord bovenin moet klikken terwijl je over het algemeen in het midden van het scherm bezig bent.
5. De pictogrammen zijn soms moeilijk te linken aan bepaalde woorden. Ik had soms een heel ander pictogram in gedachten dan degene die het daadwerkelijk was.
7. Bij veel woorden is het niet duidelijk onder welk pictogram zij zijn,   tenminste aan het begin.
8. het was een beetje irritant dat de woorden niet op alfabetische volgorde geordend zijn.
10. Also, there are some symbols which I kept mistaking. The four on the left were the easiest to learn

II)Problemen met het gebruik zoekmachine

8.Alle woorden die je zoekt staan in enkelvoud. Dat vergeet je tijdens het zoeken, en dan denk je dat het woord niet bestaat.
2. en een keer lukte het helemaal niet het juiste woord te vinden, hoewel ik deze wel in de zoekfunctie had ingevuld.
5. Ik vergat dat je de woorden in enkelvoud moet zetten (in de zoekmachine
7. Onduidelijk af en toe. Woorden worden niet altijd toegelaten ("can" deed moeilijk)
10.werkwoordsvormen moeilijk in te vullen.
2. Soms was het niet zo makkelijk bepaalde woorden met de zoekmachine te vinden, dus je moet wel weten in welke categorie je het woord ook kunt vinden (can)
10. When you search a word, I don't think it's good that you have to   type 'to' for verbs, because this takes extra time.

III)Is het programma makkelijk te leren/geschikt voor de doelgroep?

2.Te gecompliceerd om zinnen te vormen met dit programma, ik zou het zelf niet gebruiken.
10. Het zou makkelijker zijn als de woorden in de lijst op alfabetische volgorde staan, voor je het geleerd hebt duurt het wel even
1.For good communication, it is better to know all the symbol by heart, which takes some time to learn.
3. In eerste instantie is het programma niet eenvoudig om te doen De zoekfunctie is wel heel makkelijk, maar ik weet niet of mensen met afasie dit programma kunnen gebruiken.
3.Gaandeweg werd het programma wel makkelijker om te gebruiken, en ik denk dat op de lange termijn mensen alle plekken van alle woorden wel uit hun hoofd kunnen leren met dit programma.
4. In het begin duurt het even voordat je door hebt hoe het werkt. Daardoor gebruik je in eerste instantie veel de zoekmachine.
3. Ik werk liever met de zoekfunctie dan met de pictogrammen.
4. Zeker in het begin is het lastig als je de pictogrammen nog niet kent. Dan moet je als nog elk woord eerst in typen.
6. Merk dat ik vlugger typ dan dat ik kan zoeken naar de pictogrammen.
9. Ik denk dat dit systeem op de lange termijn zeer bruikbaar kan zijn. Als je er voor het eerst mee aan de slag gaat zal het meer tijd in beslag nemen om zinnen te vormen, maar naarmate je er meer mee gewerkt hebt wordt het steeds makkelijker en gaat het steeds sneller.
10. Further, it is not fast enough in my opinion. It would be faster to simply type the sentences.

IV)Problemen binnen het experiment zelf

3.Daarnaast was er een probleem met zin 10 (can)
6. Tijdsdruk in het experiment! Met meer tijd om zelf te prutsen kon ik het sneller leren, denk ik.

A PICTURE IS WORTH A THOUSAND WORDS

V)Overig

9.en duurt het wat lang.

7. De beelden en vele woorden die je ziet nemen de concentratie weg van de tekst die je eigenlijk wilt schrijven

10.) I don't think it is good that when you make a mistake, you can't set a word in between two others. You then always have to get rid of the whole sentence until the point which you want to change.

A PICTURE IS WORTH A THOUSAND WORDS

Appendix 2: hypothesis 2, deviation of time residuals per condition

1: deviation of time-residuals

1 = robot, 2 = aphasia, 3 = social media