Online Text Translation System with an Avatar Assisted Intelligent User Interface

Thijs Kloosterman, Minha Lee, Alex Olieman, Robert de Vries, Michael Wolbert
Research Methods 2012 (Group A)
University of Amsterdam

Abstract

This paper discusses the research that was performed in order to evaluate whether or not an avatar assisted user interface can be perceived to be intelligent by the users of an online text translation interface. Two versions of the same online text translation interface were developed: a version *with* an avatar and a version *without* an avatar. Participants were recruited through the researchers' network and were randomly assigned to one version of the translation interface. The results of the interface evaluation questionnaire, which the participants completed after the translation task, indicate that the interface with an avatar was perceived to be less intelligent than the interface without an avatar. Also, a strong relation was found between likeability and direct perception of intelligence of the interfaces. As for the influence of demographic factors, participants who declared to have an above-average level of computer skills perceived both versions of the interfaces to be less intelligent. Conclusively, our findings suggest that the existence of an avatar as a part of an interface does not necessarily increase users' perception of intelligence of an interface.

1. Introduction

As a result of globalization and technological advances, increased online connectivity has significantly improved the information supply for an individual. However, as with all globalization processes, natural language barriers arise. A logical method would be to have a language that is widely used, such as English, as a carrier for the majority of information found online. Yet, this would limit the access to information for those who do not speak the language and also would disregard the diversity of languages around the globe. So, the challenge is in how to present easy access to the abundance of information that is found online no matter what language the information is initially available in. The current state of machine translation offers good solutions to this challenge. One of the best known practical solutions is the instant translation that web browser Chrome from Google offers (Google Support, n.d.). This allows a user to translate a page into his or her natural tongue with just one click. The aim is to improve user interaction to assist them with their translation tasks. This is achieved by researching the differences in user perception of two interfaces, namely an interface with a digital avatar and an interface without a digital avatar.

The intention of this study was to test whether or not an avatar that assists in online text translation causes the users to perceive the interface as more intelligent compared to an interface that does not have an avatar for text translation. The definition of the concept

intelligent interface was derived from semi-structured interviews conducted with three experts in the field of Human-Computer Interaction (see Appendix for structure and transcripts). In the method section we elaborate on the operationalization of the concepts found in the interviews for the measurement of perceived intelligence. In the context of this paper, an intelligent interface is defined as 'interaction units designed to anticipate and adapt to the context of its human users within the scope of specific tasks, while providing an optimal user experience'. The specific task at hand is online text translation. To test this, two versions of the same website with a different interface to translate paragraphs were created. Both versions of the interface had four paragraphs. The 'no-avatar' version included a context menu for translation and the 'avatar' version had an instruction avatar which assisted in the translation. These versions both deal with the same task of translating a paragraph of the given text. Following from aforementioned factors, the research question of this paper is:

Do people consider an interface with an avatar as more intelligent than an interface without an avatar for the task of translating a part of a text on a website?

Regarding the above mentioned research question, we consider the following possible outcome. The hypothesis is clarified below.

The interface with an avatar is perceived to be more intelligent than the interface without an avatar.

We expect the above hypothesis to be the outcome of our research.

The above expectation is due to the fact that the usage of an avatar as an agent that assists users with specific tasks has been an approach that has been exploited by countless user interfaces (Campbell et al., 2012). Often, avatars are a part of the user interface in a form of animated virtual agents that interact with the users to assist in performing specific tasks. Avatars as user assistants for various systems have been met with varying degrees of success. Clippy, an animated virtual paper clip as an assistant for word-processing first appeared in Microsoft Office 1997. Since then, Clippy has disappeared as an omnipresent assistant, for it often came across as obtrusive to the user experience of word-processing (Gentilviso, 2010). Yet, many other interactive virtual agents have successfully contributed to positive user experiences, of which we chose Links the Cat that first appeared in Microsoft Office to be the representative avatar for our research.

Pedagogical Embodied Conversational Agents for many different virtual learning environments have been employed successfully by using tested pedagogical techniques (Dragone et al., 2006). Avatars as agents that assist in various tasks on a user interface level are effective because the users assume, consciously or subconsciously, that the system underneath the agent-based intelligent interfaces are intentionally intelligent (Friedman, 1995., Campbell et al., 2012). This supports the hypothesis that the interface with an avatar for text translation will also be perceived as more intelligent.

There are three main reasons why users perceive avatar-aided interfaces as more intelligent: avatars possess anthropomorphic features (Kim et al., 2004), they can make interfaces more expressive, and they interact in a personalized manner for each user (Dragone et al., 2012). These points lend credibility to the overall intelligence of the interface and the underlying system. Anthropomorphic features of an avatar create an experience that gives the users the impression that the avatar can behave autonomously by assisting users in real-time (Kim et al., 2004). The behavior of the avatar can thus make user interfaces expressive and the interactions with the avatar becomes seemingly personalized to the users.

The remainder of this paper consists of the following sections: method, results, conclusion, and discussion. The method section will describe in detail which choices are made during the development of the experiment, and how the research was designed to provide scientific conclusions. The results section will report on the data gathered from the participants as well as the statistical analyses that were performed. In the conclusion, a summary of the results will be presented as well as an answer to the research question. Finally, the discussion section will elaborate on the limitations of the conclusions along with the broader implications of the entire research and an indication of possible future work.

2. Methods

2.1 Participants

The population from which a sample will be taken consists of the population of computer users. The minimum amount of participants for our experiment was set at 30 per interface, resulting in 60 in total. We recruited 63 participants of which 33 evaluated the interface with an avatar and 30 evaluated the interface without an avatar. There was no preference for gender or age as long as the participants had access to the internet, but eventually there were more male (n=48) than female (n=15) participants. All participants were recruited using convenience sampling by using the personal network of the researchers. Sending the link of the experiment through email or sharing it via Facebook provided the necessary 60 participants. The use of personal networks shows in the distribution of the age of the participants. The largest age groups are 22, 23 and 24 (together more than 50% of all participants), which roughly corresponds with the age of the researchers. The remaining participants were between 18 and 51 years old.

Other descriptive characteristics were also obtained from the participants. When asked about their computer skills, 88.9% of the participants rated their own skills to be ‘above average’ or ‘very high’. Participants were also asked to indicate how often they used computer assisted translation. The results here were more divided; 41.3% stated ‘sometimes’ while 41.3% answered ‘regularly’ or ‘very often’.

2.2 Materials

The experiment will be conducted with two groups. For this, two versions of the same simple interface will be provided, one with an avatar and one without an avatar. The participants will be chosen through the self-selective sample method. They will be asked to take part in the research on a welcome page. Each participant will randomly be assigned by the system to either one of the two experimental groups. In the given interface, the participants will perform a simple translation task. After performing the translation task in the interface, each participant will be asked to fill in a questionnaire. Cronbach's alpha was used to test the internal consistency of the questionnaire (n=10): $\alpha = 0.858$. The closer Cronbach's alpha coefficient is to 1.0 the greater the internal consistency of the items in the scale. According to the rules of thumb of George & Mallery (2003), this result corresponds with a 'good' internal consistency.

2.2.1 Interfaces

The participants are asked to perform a translation task that consists of the same interaction steps for both experimental groups. The steps consist of:

1) Select a paragraph to translate;
2) Decide whether to translate the paragraph;
3) Choose a language to translate the paragraph into;
4) Repeat steps 1, 2, and 3 until three paragraphs have been translated.

The website from which participants should translate the paragraphs is also kept consistent between the two groups. The text is part of the Basque Wikipedia page about cider ("Sagardo," n.d.). Cider as a topic choice was the result of using the random article generation option in Wikipedia. The Basque language was chosen for its small number of speakers, and thus the low probability that our participants can read the provided text without translating it. This is preferable because all participants should feel a similar need for translating the text. The translations are produced with Google Translate, and are available in Croatian, Dutch, English, and French. They are hard-coded into the interface pages to ensure that translation always occurs with the same timing.

In the 'no-avatar' group, an explanatory text at the top of the page explains that paragraphs can be selected by clicking on them with the right mouse-button. If a participant does so, a context menu is given in which a language can be chosen (see Figure 1). Participants may decide to not translate the selected paragraph by clicking elsewhere on the page. If a language is chosen, the original paragraph is replaced by a translated version.

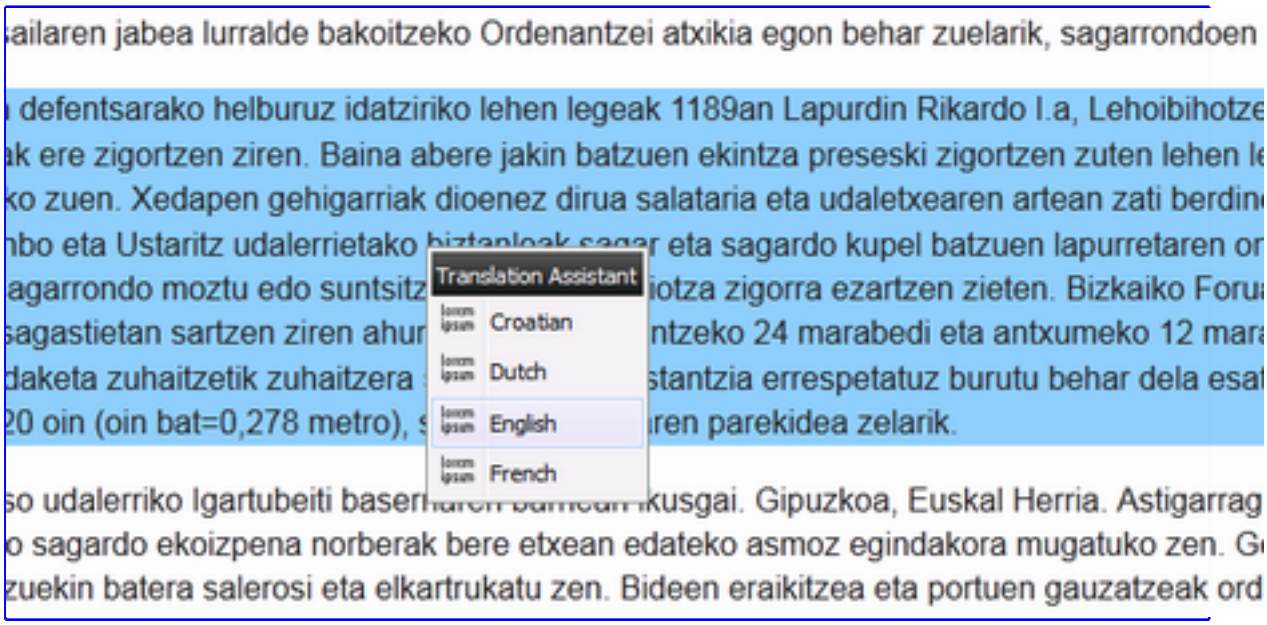


Figure 1. The language selection step in the 'no-avatar' interface.

In the 'avatar' group, all explanation is given by the avatar. We have used the existing avatar 'Links the cat', which was originally a Microsoft Agent, but is currently available to web developers in the Clippy.js JavaScript library (Smore, n.d.). Links was chosen instead of Clippy because it is less familiar and probably does not associate with Microsoft as strongly as Clippy would. Links first introduces herself as a translation assistant, and explains to the participant that paragraphs can be selected for translation by clicking on them. If a participant does so, the avatar moves to the point that was clicked, and proposes to translate the selected paragraph (see Figure 2). The choice to translate leads to the question "Into which language?", which is accompanied by the buttons for the four available languages. If a language is chosen, the original paragraph is replaced by a translated version.

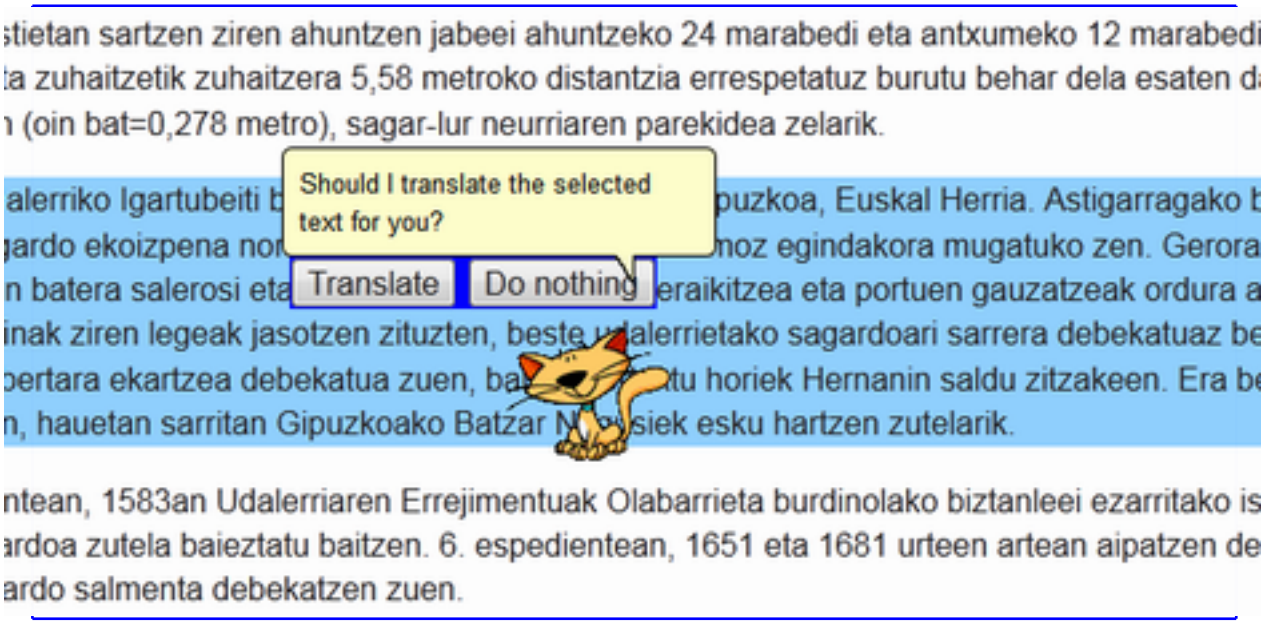


Figure 2. The 'decide whether to translate' step in the 'avatar' interface.

2.2.2 Questionnaire

The questionnaire consists of two parts. The first part provides demographic questions which will be used to elaborate on the representativeness of the sample. It consists of four questions regarding the participant's age, gender, self-reported computer skills, and frequency of usage of computer assisted translation services. We chose age and gender to provide a general basis for demographical analysis. The question regarding experience with computers has to do with the fact that our interface only exists within a digital environment and thus the knowledge about computers can influence participant's perception of our interface and his or her response to questions regarding the interface. We ask whether a participant has used computer assisted translation before because previous experience might influence the participant's answers.

The second part is designed to determine whether the participant perceived the interface as intelligent while interacting with it. The participants are asked (Q1 to Q9) to reflect on the following aspects about the interaction with the interface: the perceived flexibility, level of enjoyment, expected reactions of the interface, level of irritation, level of frustration and the ease of learning to use the interface. These indicators of perceived intelligence were chosen as a result of expert consensus retrieved from the semi-structured interviews on intelligent interfaces (see Appendix for structure and transcripts).

We referred to literature on user experience analysis for relevant questions that could be used in the questionnaire to assess these indicators. The model developed by (Hone & Graham, 2000) on measuring the Subjective Assessment of Speech System Interfaces (SASSI) was used as the main basis of the evaluation statements. The usability factors they considered were "System Response Accuracy", "Likeability", "Cognitive Demand", "Annoyance", "Habitability", and "Speed" of the system (Hone & Graham, 2000). The factors "Cognitive Demand", "Habitability", and "Speed" did not correspond to indicators of perceived intelligence, hence they were not included in the questionnaire. The remaining usability factors are briefly explained. "System Response Accuracy" measures how well the system seemed to understand the user's intentions and behaves in a manner that is expected by the user. "Likeability" deals with the level of enjoyment the user experiences whilst interacting with the system. The "Annoyance" factor is for assessing the level of irritation the user might feel during the interaction. These three factors, as developed and explicated by Hone and Graham, contribute to the assessment of intelligence of the translation interface with and without an avatar.

For the last question (Q10), participants are explicitly asked about their feelings towards the intelligence of the interface, because we relate the score of the indicators found in Q1 to Q9 to the direct perception of intelligence (Q10). All ten evaluation questions are given in the form of statements, in order to avoid suggestiveness. Participants rate the statements on a 5-point Likert scale with the labels: 'strongly disagree', 'disagree', 'neither disagree nor agree', 'agree', and 'strongly agree'.

The third and final part of the questionnaire consists of two feedback questions, where participants can provide textual feedback and leave their email addresses if they are interested in the results of the study. The former question can improve our understanding of the results

and provide additional input to our discussion. The latter question is to provide participants the option to receive the research results.

Summarized, our questionnaire contains the following specific questions:.

Part 1: Demographic information
D1: What is your age?
D2: What is your gender?
D3: How do you consider your skills with computers?
D4: How often do you use computer-assisted translation?

Part 2: Evaluation statements
Q1: The Translation Assistant was aware of what I intended to do. (System Response Accuracy)
Q2: The interaction with the Translation Assistant was inflexible. (Annoyance)
Q3: I enjoyed interacting with the Translation Assistant. (Likeability)
Q4: The Translation Assistant reacted the way I expected. (System Response Accuracy)
Q5: The interaction with the Translation Assistant was irritating. (Annoyance)
Q6: I felt in control of the interaction with the Translation Assistant. (Likeability)
Q7: The Translation Assistant is pleasant. (Likeability)
Q8: The interaction with the Translation Assistant was frustrating. (Annoyance)
Q9: It is easy to learn to use the Translation Assistant. (Likeability)
Q10: I felt that the Translation Assistant was intelligent. (Direct Perception of Intelligence)

Part 3: Feedback
F1: Feel free to provide any remarks on the completed task and website you just visited.
F2: Leave your email if you are interested in the results of this study.

2.3 Procedure

The experiment and questionnaire are presented on a website that was specifically created for this research (Kloosterman et al., 2012a). This website is comprised of several pages, namely a welcome page, an avatar page, a no-avatar page, a questionnaire page and an ending page. The participant will first see the welcome page. The welcome page contains a short introduction to the research and presents the participants a time indication of 10 to 15 minutes for the experiment and questionnaire.

From the welcome page, the participant is able to navigate to the experiment page which features the actual translation task. The experiment page either has an avatar or does not have an avatar based on random allocation using a Python script to support independent sampling. Since we aim for 30 participants per group, we programmed the random allocation in such a way that when one group has over 30 participants, while the other group has less,

the participant is allocated to the latter group. We agree that this way of allocating is not purely random. However, there has to be a fair balance between the two groups, meaning that we cannot rely on not having a restriction on the random allocation itself. Our goal is that the proposed random allocation should not result in e.g. a 20:40 (1:2) distribution of participants over the two groups.

After the translation task is completed for one of the given translation pages, the participant is redirected to the questionnaire hosted in a Google Form (Kloosterman et al., 2012b). This form presents the questionnaire regarding our research which requires all questions to be answered except for the additional information (feedback and email address). After submitting the questionnaire, the data is send to a spreadsheet which is hosted on Google Docs and the participants are directed to the ending page where they are thanked for their collaboration and time. We chose Google Docs because of the available features for collaborating as well as the easy to use interface (for which Google is known) for both the researchers as well as the participants. Also, the data can be exported to statistical tools for analyses.

2.4 Data analysis

Since there are two groups with different participants in each group, a two-tailed independent t-test of the results will be performed. The goal is to test whether the evaluation scores obtained from the experimental group - the group *with* the avatar in the interface - are significantly higher than the scores obtained from the control group; the group *without* the avatar. If this is indeed the case, the conclusion can be reached that the interface with avatar is perceived as more intelligent regarding a translation task. Each evaluation question in the questionnaire will result in a numerical score between 1 and 5. The outcome of the questionnaire is the sum score of all the evaluation questions. This summed score will be used as the dependent variable of a t-test. We will use the SPSS software package to perform the t-test in which we will use a significance level of 95% (α=0.05), which is common in two-tailed t-tests.

3. Results

The data sample for this study contained no missing values and therefore all cases were usable. There was a minor issue with one respondent who claimed to be 99. This respondent was not removed in this particular case, because 99 is a valid age and the confirmatory analysis does not depend on the age variable.

The demographic questions and evaluation statements have been recoded using the recode functionality of SPSS. The recoding involved converting the Likert-scale labels from the respondents to a numeric value; ‘strongly disagree’ = 1, ‘disagree’ = 2, ‘neither agree nor disagree’ = 3, ‘agree’ = 4 and ‘strongly agree’ = 5. A similar procedure was used for the scaled demographic questions. Since the sum of all statements is considered as the input variable of the t-test, statements Q2, Q5 and Q8 had to be inverted to adhere to the direction of the statement in relation to its score on perceived intelligence.

Evaluation statements	Avatar mean	Avatar Std. dev	No-avatar mean	No-avatar Std. dev
Q1: The TA was aware of I intended to do	3.6667	0.73598	3.8333	0.79148
Q2: The interaction with TA was inflexible	2.9091	0.91391	3.2667	0.94443
Q3: I enjoyed interacting with TA	2.8485	1.06423	3.3333	0.75810
Q4: The TA reacted the way I expected	3.1818	1.15798	3.2667	1.20153
Q5: The interaction with the TA was irritating	2.9091	1.07132	3.7333	0.82768
Q6: I felt in control of the interaction with TA	3.3030	1.07485	3.7333	0.90719
Q7: The TA is pleasant	3.0000	1.00000	3.4667	0.81931
Q8: The interaction with the TA was frustrating	3.6667	0.88976	3.9000	0.88474
Q9: It is easy to learn to use the TA	4.0909	0.80482	4.1000	0.88474
Q10: I felt that the TA was intelligent	2.3333	0.98953	3.0333	0.85029
Total	**31.9091**	**6.75631**	**35.667**	**5.33261**

Table 1. means and std. deviations for Likert-scale (1 to 5) evaluation statements of both interfaces. Note that the mean and std. dev. for statements Q2, Q5 and Q8 are based on inverted values.

3.1 Confirmatory analysis: the difference between Avatar and No-Avatar

From the 63 participants, 33 were assigned to the website with an avatar and the remaining 30 were assigned to the no-avatar site. For the former group, the mean was 31.9091 and the standard deviation was 6.75631. For the latter group, the mean was 35.667 and the standard deviation was 5.33261 (see Table. 1). For both groups, no outliers were considered because no observed values were below or above three times the standard deviation from the mean (see Figure 3). For both groups, the overall standard deviation is low. Based on Levene's test for equality of variances, the significance of all cases for both groups resulted in $p = 0.051$. Since this figure is above 0.05, equal variances are assumed; the group with an avatar interface and the group with no-avatar interface have similar variances.

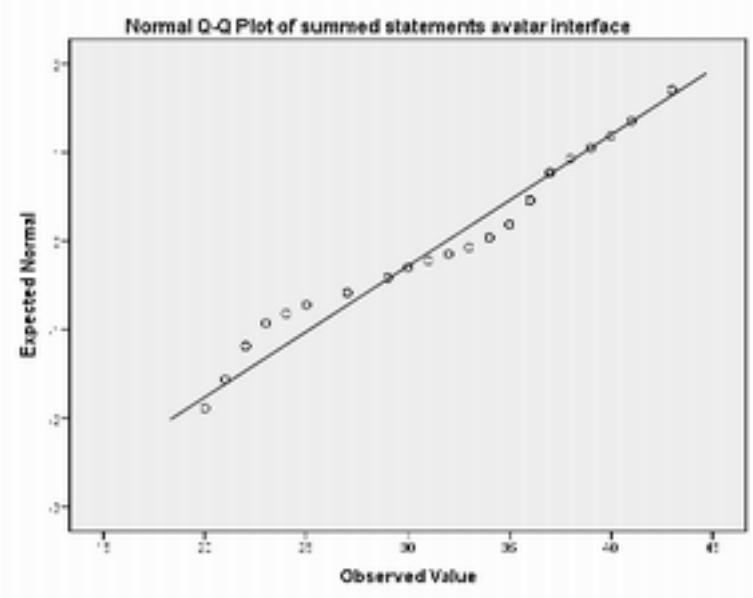
Then, to test whether one of the interfaces is considered as more intelligent than the other, a two-tailed t-test for Equality of Means was used, which resulted in $p = 0.018$ (2-tailed). The probability that the means of the two groups are equal is 0.018 and this is lower than the significance level of 0.05. Therefore, we can conclude that the current difference between the means (M=35.667 minus M=31.9091, see Table 1) is significant at 3.75758.

The perceived intelligence score for the interface with an avatar (M=31.9091, SD=6.75631) is lower than the perceived intelligence score for the interface without an avatar (M=35.667, SD=5.33261); $t(61) = -2.434$, $p < 0.05$.

At the outset of the paper, the hypothesis that was established was:

The interface with an avatar is perceived to be more intelligent than the interface without an avatar.

However, given the statistical results of the evaluation questions, the hypothesis has been rejected. Since a two-tailed test was used, the hypothesis cannot only be rejected, but also its negated version can be accepted: the interface with an avatar is perceived to be less intelligent that the interface without an avatar.



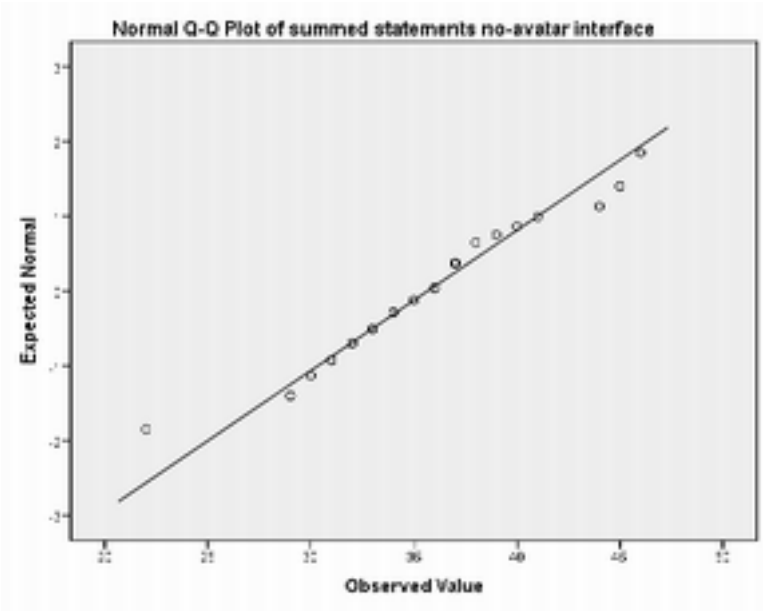


Figure 3. for both interfaces, the observed sum values of evaluation statements and distribution around the mean (expected normal = 0).

3.2 Exploratory analyses

In addition to testing the hypothesis, it is necessary to investigate the relationships between several elements of the questionnaire. Initially, it was interesting to know whether males and females had evaluated the interfaces differently. Another t-test was performed to determine whether gender was of decisive influence on perceived intelligence (the sum of all evaluation questions), but the difference was insignificant.

A linear regression test was used to find relationships between the non-nominal demographics of the participants and 'direct perception of intelligence' (the rating of Q10: "I felt that the Translation Assistant was intelligent."). Linear regression was also used to quantify the relationships between the usability factors and direct perception of intelligence.

A simple linear regression test determines whether or not there is a linear relationship between the variables X and Y, and estimates what this relationship is. This relationship can be expressed in the formula Y = a + b*X. Where Y is the dependent variable and X is the independent variable. The regression test reveals the intercept *a* and regression coefficient *b* for this formula. With this equation the outcome of Y for a certain X can be predicted (de Haan, 2011).

Linear regression may also be used to assess the strength of the relationships between multiple independent variables and a single dependent variable. To accomplish this, multiple linear regression is used, which can be expressed in the formula $Y = a + b*X_1 + c*X_2 + ... + i*X_p$. Multiple linear regression was used on two sets of independent variables: the usability factors and the demographics, excluding gender. In both cases, the dependent variable was the direct perception of intelligence. The analyses were conducted in SPSS with a stepwise regression method. The used selection criteria are: 'enter a variable into the model if the probability of F <= 0.050', and 'remove it if the probability of F >= 0.100'.

3.2.1 Multiple linear regression of demographics

One linear regression test has been performed to determine the influence of age, self-reported computer skills, and translation usage frequency on direct perception of intelligence. From this test it can be concluded that age and translation assisted usage frequency, do not have a significant relationship with direct perception of intelligence. So only self-reported computer skills show a significant relationship. The linear regression test indicates that there is a relation between $X_{\text{Computer Skills}}$ and Y. The intercept *a* has the value 4.511 and the regression coefficient *b* has the value of -0.43. This results in the linear equation $Y = -0.43X_{\text{Computer Skills}} + 4.511$. This model has an Adjusted R Square of 0.093, indicating that it has accounted for 9% of the variance in the ratings of direct perception of intelligence.

3.2.2 Multiple linear regression of usability factors

The usability factors "System Response Accuracy", "Likeability", and "Annoyance" were included in the questionnaire because they are expected to contribute to the overall perceived intelligence. However, it was still unclear how strongly these variables are related to the direct perception of intelligence. The relationship between usability factors and direct perception of intelligence were quantified with stepwise linear regression.

By applying the selection criteria for independent variables, one model was created. It only includes "Likeability" as strongly related to direct perception of intelligence. This model could not be made to fit the data better by adding other variables. In a form of an equation, the model is expressed as: $Y = -0.483 + 0.684X_{\text{Likeability}}$. This model has an Adjusted R Square of 0.459, indicating that it has accounted for 46% of the variance in the ratings of direct perception of intelligence.

4. Conclusion

This study researched the effect of a digital avatar on perceived intelligence of an online translation interface. The research question was: do people perceive an interface with an avatar as more intelligent than an interface without an avatar for the task of translating a part of a text on a website? From our hypothesis, we expected the interface with an avatar to be perceived as more intelligent. The results provide an answer to our research question; they indicated that the hypothesis should be rejected. The difference between the control and experimental group lead to the conclusion that the interface without the avatar was perceived to be more intelligent. Thus, the results suggest that the placement of an avatar as an assistant in an interface does not necessarily cause the users to perceive the interface as more intelligent.

The exploratory analyses indicated that participants' computer skills contributed significantly to the score of direct perception of intelligence with regard to both interfaces. In other words, the participants with higher self-reported computer skills perceived the interface that they were assigned to as less intelligent. Part of the analyses was to investigate possible relationships between a set of evaluation questions regarding the interface the participants interacted with. This analysis showed a significant relationship regarding the Likeability

factor of the interfaces. The Likeability factor is strongly related to the direct perception of intelligence (46%), while the self-reported computer skills of the participants contributed less strongly (9%) to the direct perception of intelligence. The effect of likeability provides a plausible explanation that affective aspects of an interface with an anthropomorphic avatar are of great influence to the evaluation of such interfaces.

5. Discussion

The results of this study did not match our expectations. We expected the interface *with* an avatar to be perceived as more intelligent than the interface *without* an avatar. However, the opposite was found. This section is devoted to discussing what might be the underlying cause for this, and proposes future research on this issue.

5.1 Participants' comments

Participants left comments after they participated in the experiment, which helped us to clarify their evaluations. In summary, the following remarks have been frequently made in the comments:

1) The avatar reminded participants of MS Clippy, with which they have had bad experiences.
2) The animations of the avatar, such as talking and moving, were too slow and the avatar was in the foreground more often than necessary.
3) There were not enough language options supported. Also, after performing a translation, the participants expected the possibility to translate the paragraph back to Basque or another language.
4) Some participants were confused by the animation of 'text replacement'. They prefered to have the text translated instantaneously.
5) It was unclear to participants whether judging the quality of the provided translations was part of the evaluation questionnaire.

After analyzing the comments, several reasons of why the research might not have lead to the expected results can be extracted. One simple reason is that the participants found the avatar annoying because of its slow speed and movement to the mouse pointer whenever a participant clicked on the text. Another cause might be that the participants regarded the task to be too simple to justify the presence of an avatar. Furthermore, an interesting aspect is that the participants showed their disappointment in the avatar. Their disappointment might be due to the belief that the avatar was intelligent because of its anthropomorphic features (Kim et al., 2004). So it could be that the avatar was indeed perceived as more intelligent initially, but that after not meeting the participants' expectations, it was judged more negatively (Nowak & Biocca, 2003).

5.2 Bias of experience

From the exploratory analysis, a weak but significant relation was found between the participants' self-reported computer skills and the direct perception of intelligence. Participants with a higher judgement of their computer skills perceived both types of interfaces as less intelligent. It might be the case that users with better computer skills had higher expectations of the interface, and this in turn had a negative influence on their evaluations.

As described in the method section, the sample contained a high percentage (88.9%) of participants with a self-reported computer skill score of 'above average'. Due to the non-probability convenience sampling method that was employed, the majority of the participants in the sample were experienced computer users (i.e. peers of HCI related programs at universities or IT practitioners). We argue that this has affected our findings because these experienced computer users study or work with different interfaces more frequently than moderately experienced users. Therefore, experienced users might have had higher and different expectations of interfaces. This limits the generalization of the findings to a more heterogenous population of computer users.

5.3 The influence of likeability

A strong relationship was found between the Likeability factor and the direct perception of intelligence. This relationship was characterized by participants that: enjoyed the interaction, felt in control of the interaction and found the interface pleasant and easy to learn to use. It is surprising to find such a strong relation between these variables because likeability and perceived intelligence are commonly treated as distinct concepts in Human-Computer Interaction research (Bartneck et al., 2009). One previous study in which these variables were evaluated for an interface with an avatar could be found, but there was no significant relation between likeability and perceived intelligence (Koda & Maes, 1996).

Likeability in Human-Computer Interaction has previously been compared to positive first impressions that one person may have of another person (Bartneck et al., 2009). These first impressions between people are referred to as 'Interpersonal Affect' in psychological literature. It has been found that interpersonal affect has a significant positive influence on performance evaluations (Robbins & DeNisi, 1994) in an experiment where 83 students rated the performance of three of their professors. Interpersonal affect is also a good predictor of whether a job applicant will be hired, and of whether a blind date will be a success (Berg & Piner as cited in Bartneck et al., 2009, pp. 75).

If similar effects occur between likeability and the perceived intelligence of an interface, users may unconsciously take an affective stance towards an interface within one or two minutes of using it. This affective stance will then influence the judgments about more cognitive aspects of evaluation, such as direct perception of intelligence. Bartneck et al. (2009) acknowledge that the concepts of user perception, such as likeability and intelligence, may be inherently related. They elaborate:*"A highly anthropomorphic and intelligent [avatar] is likely to be perceived to be more animate and possibly also more likeable."* (Bartneck et al., 2009, pp. 72-3). Based on the analogy with interpersonal affect, we argue that it is more probable that likeability influences perceived intelligence, instead of the other way around as Bartneck et al. suppose.

Our hypothesis is supported by previous findings that avatar-aided interfaces are perceived as more intelligent. The three main arguments are that anthropomorphic features (Kim et al., 2005) and expressive and personalized behaviours (Dragone et al., 2012) of an avatar may cause this perception. While the avatar used in this study had anthropomorphic features, its behaviour was likely too simple (e.g. reactive) to be experienced as expressive or personalized. This might explain our finding that an interface with an avatar is perceived as less intelligent. Therefore, we argue that this finding can only be generalized to avatars which lack expressive and personalized behaviour.

5.3 Future research

For more conclusive results, future research on perceived intelligence of avatars might consider defining which requirements digital avatars need for specific levels of anthropomorphism and how those requirements could contribute to the user experience. Such requirements should focus on different anthropomorphic features described by the literature, which then could be included in an experimental interface. A replication of this experiment might then give more conclusive results on how these anthropomorphic features relate to the likeability and perceived intelligence of a translation interface. Alternatively, a similar experiment could be conducted with the avatar that was used in this study as a control condition, while an avatar with the same appearance but with added affective behaviour could be used as the experimental condition. This would give more insight into the influence of affective behaviours on perceived intelligence. Moreover, these proposed areas of future research should focus on the comparison of avatar features instead of comparing an avatar with no avatar. We argue that the current comparison is too general and therefore limits our conclusion regarding the users' perception of intelligence of an avatar-assisted translation interface.

References

Bartneck, C., Kulic, D., Croft, E., & Zhogbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. In *International Journal of Social Robots* 1, pp. 71-81.

Campbell, A., Collier, R., Dragone, M., Gorgu, L, Holz, T., O'Grady, MJ., O'Hare, G.M.P., Sassu, A., & Stafford, J. (2012). Facilitating Ubiquitous interaction using Intelligent Agents. In Zacarias, M., de Oliveira, J. (eds.). *Human-Computer Interaction : The Agency Perspective*, Springer.

Dragone, M., Duffy, B.R., Holz, T., O'Hare, G.M.P., & Martin, A. (2006).
Mixed reality agents as museum guides. In *ABSHL 2006: Agent-Based Systems for Human Learning, AAMAS 2006 Workshop*, ACM Press.

Friedman, B. (1995). It's the computer's fault: reasoning about computers as moral agents. In *CHI 95 Conference Companion*, pp. 226–227.

Gentilviso, C. (2010). Worst Inventions: Clippy. In *The 50 Worst Inventions*. TIME Magazine. Retrieved November 15, 2012 from http://www.time.com/time/specials/packages/article/0,28804,1991915_1991909_1991755,00.html

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Google Support (n.d.). Translate a page to your language. In *Google Chrome Help.* Retrieved December 19, 2012, from https://support.google.com/chrome/bin/answer.py?hl=en&answer=173424.

de Haan, R.J. (2011). Lineaire regressie. In *Wiki Biostatistiek.* Retrieved December 19, 2012, from http://os1.amc.nl/mediawiki/index.php?title=Lineaire_regressie.

Hone, K.S., & Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). In *Natural Language Engineering* 6 (3–4), pp. 287–303.

Kim, J. H., Kim, Y. D., & Lee, K. H. (2004). The third generation of robotics: Ubiquitous robot. In *ICARA 2004: International Conference on Autonomous Robots and Agents*.

Kloosterman, T., Lee, M., Olieman, A., de Vries, R., & Wolbert, M. (2012a). Welcome. In *Interface research.* Retrieved December 19, 2012, from http://werockit.nl/rm-group-a/welcome.html.

Kloosterman, T., Lee, M., Olieman, A., de Vries, R., & Wolbert, M. (2012b). Evaluation survey. In *Interface research.* Retrieved December 19, 2012, from https://docs.google.com/spreadsheet/viewform?formkey=dHhJdUE1VkgwaFJPTkNadWVudzlKOHc6MQ.

Koda, T., & Maes, P. (1996). Agents with faces: The effect of personification. In *Robot and Human Communication,* pp. 189-194. IEEE.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. In *Presence: Teleoperators & Virtual Environments*, 12(5), pp. 481-494.

Robbins, T.L., & DeNisi, A.S. (1994). A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations. In *Journal of Applied Psychology* 79 (3), pp. 341-353.

Sagardo. (n.d.). In *Wikipedia.* Retrieved December 19, 2012, from http://eu.wikipedia.org/wiki/Sagardo.

Smore (n.d.). ClippyJS. In *A weekend project by Smore.* Retrieved December 19, 2012, from http://www.smore.com/clippy-js.

Appendix: semi-structured interviews

Interview structure

1) How would you define an intelligent interface?
2) What is the difference between an intelligent interface and intelligent system?
3) Which elements need to be present in an interface, in order for users to perceive an interface as intelligent?
4) Will users over time learn to view artificial intelligence as a distinct type of intelligence in interfaces?
5) How can users benefit from an intelligent interface?

Interviewee Liubo Borissov

Interviewer(s):	Interviewer Lee
Date and location:	8 November 2012, Skype conference call
Interviewee:	Liubo Borissov
Age:	37
Gender:	male
Education:	BSc in Mathematics and Physics at Caltech, MPS in Interactive Telecommunications at NYU , Ph.D in Physics at Columbia university
Profession:	Associate professor at Pratt Institute's Department of Digital Arts

Publications:

Physics:
Gabriela Barenboim, Liubo Borissov, Joseph Lykken and Alexei Yu. Smirnov (2002).“Neutrinos as the messengers of *CPT* violation” Journal of High Energy Physics Issue 10

Arts:
Maja Cerar, Liubo Borissov (2007). “Autopoiesis/Mimesis”. ACM SIGGRPH 2007 art gallery. 173.

Maia Marinelli, Jared Lamenzo, Liubo Borissov (2005). “Mocean”. Proceedings of the 2005 conference on New interfaces for musical expression, 272 - 272.

Exhibitions:
Liubo has exhibited at places like New Interfaces for Musical Expression, ICMC and SIGGRAPH conferences, the Lincoln Center Summer Festival, NYC and the Kennedy Center, Washington, DC

Most recent exhibitions:

- 2009 Interactive software design and consulting in collaboration with D'Strict for Samsung's Corby Launch Event, Milan, Italy
- 2008, 2009 Spark Festival of Electronic Music and Arts, Minneapolis, MN -- interactive multimedia performances in collaboration with violinist Maja Cerar.
- 2008 Multimedia director for Man and Nature, a concert event for the Fourth U.S. - China Strategic Economic Dialogue hosted by the Department of the Treasury. Created and performed a multimedia interpretation of Tan Dun's Water Music. Annapolis, MD.
- 2006-2008 Living Portrait, a video art and interactive installation commission by the New York Asian Women's Center presented at the Edward Hopper Studio, the Asian/ Pacific/American Institute Gallery and the NYU Kimmel Center. Collaboration with Heather Greer and Woody Pak.

Interviewee Gerard Alberts

Interviewer(s): Alex Olieman and Michael Wolbert
Date and location: 8 November 2012, Gerard's office at Science Park

Interviewee: Gerard Alberts
Age: 58
Gender: male
Education: Ph.D in Mathematics
Profession: Assistant Professor at the Universiteit van Amsterdam
Publications: *Appropriating America; Americanization in the history of European computing*, Gerard Alberts (ed), *Annals of the History of Computing*, Special Issue 32-2, April-June 2010

Interviewee Pablo Cesar

Interviewer(s): Interviewer Lee and Interviewer de Vries
Date and location: 8 November 2012, Pablo's office at CWI

Interviewee: Pablo Cesar
Age: 36
Gender: male
Education: Ph.D in Computer Science and Engineering at Helsinki University of Technology
Profession: Researcher at CWI (National Research Institute for Mathematics and Computer Science)
Publications:

- P. Cesar and W.T. Ooi, "Embracing Open Source Multimedia Software," *IEEE MultiMedia*, 19(2): 11-15, 2012

- J. Jansen, P. Cesar, D.C.A. Bulterman, T. Stevens, I. Kegel, and J. Issing, "Enabling Composition-Based Video-Conferencing for the Home," *IEEE Transactions on Multimedia*, 13(5): 869-881, 2011
- I. Vaishnavi, P. Cesar, D.C.A. Bulterman, O. Friedrich, S. Gunkel, and D. Geerts, "From IPTV to Synchronous Shared Experiences: Challenges in Design: Distributed Media Synchronization," *Elsevier Signal Processing: Image Communication*, 26(7): 370-377, 2011
- D. Williams, M.F. Ursu, J. Meenowa, P. Cesar, I. Kegel, and K. Bergstrom, "Video Mediated Social Interaction Between Groups: System Requirements and Technology Challenges," *Elsevier Telematics and Informatics*, 28(4): 251-270, 2011
- B. Gao, J. Jansen, P. Cesar, and D.C.A. Bulterman, "Beyond the Playlist: Seamless Playback of Structured Video Clips," *IEEE Transactions on Consumer Electronics*, 56(3): 1495-1501, 2010

Interview transcript Liubo Borissov

Interviewer: Based on your experience, could you tell me how you would define an intelligent interface? And it can be sort of based on literature and your own personal opinion as well.

Liubo: My sense is that because technology is just so prevalent in everyday use we start with a certain expectation of response. Actually, the technology we have used for now for a few decades has kind of conditioned us to certain expectations. And so, part of the intelligent interface that I think most people understand is when you start for example say in google, you start typing and google starts telling you things that maybe you don't already know. But things that you might know. I think that the main thing that I would define as a kind of proto-intelligence is anticipation and expectation. And a lot of that you could probably argue whether that's really intelligent or not, but it definitely has to do with kind of analyzing how people behave. So it's almost... the system learns from you and the better it knows you as the user the more it tries to respond to you in a way that it thinks that you want. And so it kind of learns you and gives you back what you give it. So that's what we call intelligence in terms of interfaces or any kind of technological human interaction. But beyond that, what last year, no it was just a few months ago, I was watching the Watson Jeopardy show, (http://www.youtube.com/watch?v=qpKoIfTukrA), like the semi-intelligent computer. And that's me is the cutting edge of where intelligence is right now. It's not intelligence in the free-will sense of the word, but it's intelligent enough so that if you start with a self-contained system like that machine, it can actually extrapolate facts, understand questions, and kind of respond to you in a way where it might not fool you enough to pass the Turing test, but as a machine, it really makes for a conversation, for a natural conversation between you and the technology.

Interviewer: OK so, then will users over time learn to view artificial intelligence to be a distinct type of intelligence than interfaces?

Liubo: Yeah, I mean, when I first started looking into... When I was a kid, really actually my goal was to be a computer scientist because I was so excited about this and in the end, that was the dream of true artificial intelligence where you just have this autonomous being. But what I think has started happening is that technology just gets smarter, but not intelligent and that's the original sense of the word. But it becomes so gradual and prevalent that at some point it does not matter. And so I think the question of whether we can achieve hard AI is, this truly autonomous system, becomes less and less practically relevant because the systems that we end up using, and you know again, they learn more and more about us so they appear to be effectively intelligent at least for the purposes that we want them. And moreover, I think inevitably technology just changes us, like as human and as society and how we actually relates to it. So in addition to technology learning from us, we are actually learning and developing our own vocabulary to how to interact with anything. And so in that sense, I feel like we are converging this idea of computer intelligence, or machine intelligence, and human intelligence; instead of us setting the bar of machines have to get to this point and have to pass the Turing test or something like that. More and more I feel like that it's actually the two sides are kind of connecting slowly. And it might take years. It might never even reach some kind of a Utopian state or has that where machines become smarter than humans and take over the world. But I think for the way we relate to technology, actually we are kind of getting closer and closer. Plus we spend more time interacting with technology than with humans. Or using technology as some kind of a mediate.

Interviewer: OK. So then, overall, how you and I, just as users, can benefit from an intelligent interface?

Liubo: So, the first thing about building an intelligence as opposed to say, let's take the flip side, the unintelligent interface I would say. The computer, say windows type interface, not the operating system, but the what-you-see-is-what-you-get thing, where you have to be trained to use the system. And eventually you become proficient. I would say if this is the opposite, or the flip side of an intelligent interface, I feel the first thing that happens when you have an intelligent interface, the bar is lowered or the threshold is lowered for who can access it. So, after that, once you open this up, and by design, the system learns more and more from its users, so exponentially it gets much better in terms of where it can get with the system that adapts to the users as opposed to a system that's designed by a designer and does not change. And so to me that's the big, main benefit from it. As somebody who enjoys programming and going into the system, almost inherently I would prefer to actually have that low access to the system and be able to understand it, change it, and modify it. But that's not in a way, I separate that activity in my mind from something like calling up on Skype and talking to you, right. So I want my everyday user systems to be smart enough so that I don't have to think about them, I don't have to click too many times. Ideally, I'd want to be kind of an an extension of my being. I think hat's where mobile devices are moving in that direction. On the other hand, I think there's a benefit to a system that is a system on its own that allows you to build other system with it. Personally, I don't see the benefit of an intelligent interface there. That's because in some ways, it's limiting. Once you start querying the system about

itself, then an intelligent interface, I think the way that we both have been talking about it, kind of prevents it from doing that. It's kind of this extra layer that it just stops to come back. So you know, on a practical side, there's this now, people are talking about Windows 8 that just came out. If you start thinking in the terms that we are talking about, Windows 8 is an attempt to make the system more intelligent in terms of its users. So it simplifies the interface, and kind of tries to anticipate what things you care more, and so on and so forth. And I think that will possibly, kind of make more headway in expanding the user base and kind of more people feeling comfortable with this interface. And when I talk to programmers or people who design software, video games, et cetra, they really don't like the system for the same reason actually, that it actually prevents them from access certain things. But I think that there's going to be room for both. In the end, the evolution of technology is probably going to go to systems that are much closer to, and more anticipating of how humans behave. And maybe programming at some point becomes irrelevant because if the systems become so intelligent that I can just say in natural language something, build me a program that does this, maybe we can get to that point. I don't know how and how long it can take to do that.

Interviewer: OK. I will just ask one more question just as a follow. I am just curious to know how you think the difference between intelligent system and intelligent interface is.

Liubo: Mhm. I think that's a good question. I think in retrospect, now that I think about it, you are probably interested in the interface than the system itself right? That's just how question started.

Interviewer: Yeah it's a vague area yeah.

Liubo: Yeah, let me think about... Well, I think the intelligent interface... One way to think about technology is as an extension of the human, or the extension of the controller. Let's say that the humans are controllers. Meaning that in general tools are extensions of our beings. It's just one of the paradigms. You can think of a fork as an extension of your hand, glasses are extensions of your eyes. So we are kind of projecting outwards our senses and the way we can manipulate technology, that those become our tools. Now, when our tools become smarter that allows us to project further. Like I can have more power of controlling the environment. Now the interface, when machines and systems are concerned, have to function as one. It still is a tool for us. It has to extend our gestures our sense, etc, but it also has to interact with the system that's underlying it. So yes, you could have a very dumb system with a very intelligent interface that extends the human gestures and the human control in a very natural way so we think that it's intelligent because we feel that it understands us, but we know in the end we just get directions for how to get from point A to B, which is not that smart. It's a useful thing, but it doesn't necessarily require intelligence. And so in that sense, trying to just constrain this to interfaces only. Yeah, I think the main role of intelligent interfaces would be to really make us as humans feel in control and in power to change the environment, navigate the environment, which now is not anymore just the physical space, but the intersection between the physical space and the technology

space. So that point, yes. I think the intelligent interfaces going to be the norm more than the exception of this because that's what we want as humans.

Interviewer: OK. Yeah, I think those are the questions that I wanted to ask. Do you have anything that you want to add?

Liubo: Hmm, no. I think this is good.

Interview transcripts Gerard Alberts

Interviewer: Zou u een definitie kunnen geven van een intelligente interface? Hoe zou u dat concept afbakenen?

Gerard: Ik zou naar de alledaagse notie gaan van: stoort het ons dat het een machine is, en geen mens? Als je de vraag framed door te vragen naar ‘wat is intelligent’, dwing je mensen al in een bepaald hokje.

Interviewer: U gebruikte eerder de term ‘slim’?

Gerard: Ja, meer alledaagse termen. Want intelligent associëren we met iq, voor de rest is het afgeleid van wat de geeks en nerds ons zeggen. Daar zit dus in je vraagstelling de veronderstelling dat je die dingen gemakkelijk kunt vergelijken. Ik zou dus zoeken naar een definitie in alledaagse woorden.

Interviewer: En dan met name voor het concept intelligentie?

Gerard: In het algemeen ga ik voor omgang met de machine. Dus gaan we gemakkelijk om met een machine? Game developers zeggen dan ‘immersive’. Ik vind het problematisch om te definiëren, omdat je met je definitie je vraagstelling zo framed dat je wat je eigenlijk weten wilt, niet meer te weten komt. Namelijk, komen mensen er prettig mee overweg? Zien ze het niet meer als bedreiging. En dat, als je het hebt over een robot die iets moet doen, onze . Je hebt dus een begrip je eigen gemaakt, intelligent. Vanuit jou definitie ga je vragen, en dat moet je loslaten, buiten het frame treden. En dan gaan interviewen. Bv, wanneer kunt u zich herinneren dat er een computer aan de telefoon was en dat u zich niet ergerde?

Interviewer: Ik ben het met u eens, toch zijn we op zoek naar een definitie, misschien meer vanuit het perspectief vanuit de makers van intelligente machines. Hoe zou zo iemand dat definiëren.

Gerard: Dan kun je gewoon 40 jaar literatuur over hci bestuderen. Ik kan wel meedenken vanuit de experts, wat hun definitie zou zijn van intelligente interfaces, maar ik zie meteen voor me hoe die probleemstelling van wat is interactie tussen mens en machine en hoe vat je dat, die heeft zelf een geschieden zou oud als computers. Eerst geluid, dan beeldschermen, jaren 60.

Vanaf 1968 heb je een scherm waar je echt mee op kunt reageren. Daarna heeft xerox, echt onderzoek gedaan naar hoe kan je nou handig omgaan met de computer, muis bedacht, grafische interfaces ontwikkelt etc. Er zijn wel mensen die dan al bezig zijn met artificiële intelligentie, maar wanneer intelligentie in dit verband, weet ik eigenlijk niet wanneer dat kwam, jaren 90 denk ik.

Interviewer: het woord intelligentie is ook vaak voor het woord system geplakt.

Gerard: Ja, maar hier heb je bv met gedistribueerd input. In de jaren 80, de uni had 1 computer, met terminals waar je je programma kon uitvoeren, en dat waren domme terminals, met ponskaarten, of je kon typen, maar dat was remote access. Indien er lokaal iets gebeurde, was dat als slim. Dan zat er een processor in de terminal.

Interviewer: dat was een definitie die er toen voor gebruikt werd?

Gerard: dat was het spoor dat wij, hoe kunnen wij communiceren met onze machine, bij mijn weten is daar het woord intelligent in verband met interface ingekomen. Maar opnieuw, dat is dus voor de geavanceerde gebruiker merkt dat, de ontwerper communiceert met die gebruiker die kan dat woord gebruiken, maar voor de buitenwereld betekent het niets. Maar nu praten we dus over een algemeen publiek dat een interactie moet hebben met computers. En als we dan een terminal hebben en een gat in de muur om geld uit te halen, nou die doet het behoorlijk goed. In de zin dat er nog heel weinig onvrede bij de klant optreed. Maar als je nou de ontwerpers vraagt, is dat nou intelligentie? Dan zeggen ze nee, maar het is eigenlijk heel intelligent, want dat ding veroorzaakt geen ergernis. Een heel ander voorbeeld; mijn moeder is 87, en die vindt nu dat ze online moet bankieren. Waarom? Niet omdat ze ontevreden was met papier, maar de service van mensen die het op papier doen, wordt teruggeschroefd naar nul. Dus dan maar online. En we gaan iedere week een stapje verder. Eerst inloggen met een wachtwoord. Als dat verkeerd gaat, moet je weer een week wachten op nieuw wachtwoord, na 3 weken pas bij de eerste transactie aangekomen. En dan weet ze niet of die fout of dat nou aan haar gebrek aan vingervlugheid ligt, of dat het het plugje van het toetsenbord was, of de website, daar komt ze niet achter, het is gewoon misgegaan. Maar nu zitten we in het systeem, al een paar weken zover dat ze succesvol kan inloggen. Dan maak je een transactie, dan vul je het scherm in, en dan gebeurt het. Vanuit systeem ontwerp perspectief het summum van intelligentie, de klant typt iets in en nu gaan we hem erop wijzen dat hij iets verkeerd heeft ingetypt, dat de data niet klopt, of ergens iets kan toevoegen. Van achter het scherm gezien is dat het summum van intelligentie, maar dat is totaal onhanteerbaar voor iemand zoals mijn moeder. De klant is op het verkeerde been gezet.

Interviewer: als ik u goed begrijp, beschrijft u de interface waar uw moeder gebruik van gemaakt, en uiteindelijk het systeem dat intelligent moet

Gerard: vanuit het oogpunt van de systeemontwikkelaars is het intelligent, maar voor mijn moeder is zo'n reactie van de machine zó confronterend, daar kan ze niet mee uit

de voeten. Ik ben er aan gewend, jullie generatie weet niet beter. Daar is aan de hand dat de aansluiting van een goed doordacht en interactief systeem op het alledaagse leven is niet gelukt. Het stoort. Ik merk ook aan mezelf, dat ik als een soort tweede natuur, als een vanzelfsprekendheid ook die dingen doe om op het volgende schermpje te komen. Van oh, het schermpje geeft aan dat de naam niet klopt met het rekeningnummer, en daar reageer ik op, dan kom ik verder.

Interviewer: Is die vanzelfsprekendheid die u nu beschrijft, is dat intelligentie? Maakt dat een interface intelligent?

Gerard: Dat wij dat als vanzelfsprekend gaan ervaren, ja precies dat. Wij zijn gedisciplineerd om daar aan te passen. En dit gebeurt op alle niveaus, niet alleen bij mijn moeder, maar ook tussen mij een Maarten Marx. Hij moet mij iets overdragen, van de opleidingscommissie, wat hij deed moet ik nu gaan doen, en dan hebben wij een hele email correspondentie over wat ik wil weten, en hij begrijpt niet wat ik wil weten omdat het zo vanzelfsprekend is geworden voor hem om het zo te doen. Dat is dan over het thema vanzelfsprekendheid.

Interviewer: Maar dat gaat dus nog wel uit van een interface die alsnog voorspelbaar is? Dat hebben wij dan misschien doorzien hoe die interactie doorbroken kan worden.
Andere mensen zeggen dat aanpassingvermogen een belangrijk kenmerk is van de intelligentie van een interface, maar wat ik van u begrijp, is dat mensen het veel handiger vinden als het veel voorspelbare is, als het meer een machine is.

Gerard: Nee, dat zeg ik niet. Ik bedoel te zeggen, wat wij als voorspelbaar ervaren, dat evolueert ook. Ik moet ook telkens van nieuwe telefoons ontdekken wat ik ermee kan doen.
Het is een misverstand dat de aanpassing aan de kant van het systeem alleen maar ligt, het blijft interactie. En als je het niet ziet als interactie, maar als eenzijdig optimaal aanpassen, zit je altijd mis.

Interviewer: We hebben het net over eigenschappen van interfaces gehad. Wat denkt u dat belangrijke eigenschappen zijn van die interfaces, om door mensen als intelligent te worden gezien. Kunt u meerdere eigenschappen noemen?

Gerard: Ja, dat het niet-storend moet zijn. Storen is heel zintuiglijk. Goede communicatie gaat via onze zintuigen, alleen via de ogen werkt niet, er moet een goede spreiding zijn van het aanvoelen van verschillende zintuigen.

Interviewer: Geldt dat dan ook voor de interface zelf?

Gerard: Ja, interactie met de machine moet gezien worden als een lijfelijk proces, en als je het ziet als intelligent, is het een cerebraal proces, alleen via onze hersenen. Maar wij communiceren met onze zintuigen, lijfelijk. Dat is wat wij doen. Daarom vind ik de vraag naar intelligent dat die ons op het verkeerde spoor zet, omdat die ons aanspreekt op onze

hersens, maar dat gaat via zintuigen.

Interviewer: Ok, zullen mensen de slimheid van een apparaat altijd anders ervaren dan de communicatie met een persoon, zullen ze het als een soort van machine intelligentie gaan zien?

Gerard: Ja

Interviewer: Dus niet zo dat het apparaat ook een huid, en oren heeft, in die termen?

Gerard: Ja, al die metaforen zullen we gaan gebruiken. En natuurlijk zullen mensen ook gaan neuken met machines; doen ze misschien al. Dat kunnen we allemaal verwachten. Maar het zal altijd ervaren worden als dat het geen mens is, dat weten we donders goed. Aan de ene kant kun je zeggen dat we bezig zijn om die kloof te overbruggen, maar tegelijkertijd weten dat we die kloof niet kunnen overbruggen, dus ik denk dat we die kloof beter kunnen laten staan. In immersie is het niet storend dat het artificieel is. Bv de cave, een 3d omgeving, met je bril stap je erin en je ervaart dat als de omgeving waarin je staat. Een van de knapste ervaringen daarbij was dat je je in een bos kon waren, waarbij de bomen waren gemaakt van kindertekeningen, evenals de blaadjes. Dat stoorde op geen enkele manier de ervaring van in een 3d ruimte zijn. Je hoeft op dat vlak dus helemaal niet die kunstmatigheid weg te poetsen, die kunstmatigheid stoort ons niet, juist eerder de weggedrukte kunstmatigheid, de schijn van natuurlijk is het die ons stoort.

Interviewer: Omdat we het verschil toch wel doorhebben?

Gerard: Nee, we willen daar wel graag in opgaan, in een ki omgeving, we willen ook graag die lijfelijke band met de omgeving hebben. Maar de kunstmatigheid daarvan weerhoudt ons niet. Dat een computer bepaalde knoppen vereist is niet wat ons stoort, wij vinden het geen probleem om ons aan te passen. Voor mij het woord intelligent bij interfaces, dus eigenlijk het concreet geworden misverstand, dat dat niet een mens is, maar dat het als mens zou moeten dienen stoort ons.

Interviewer: Denkt u dat mensen uiteindelijk wel baat kunnen hebben bij dit soort interfaces?

Gerard: Ja natuurlijk, maar doe niet alsof het mensen zijn! Bv een koffiemachine, daar druk je een paar knopjes in en er komt koffie uit. En reclame heeft ons vertelt dat er zo'n schuimlaagje op hoort, nou dan maken we een machine die zo'n schuimlaagje erop doet. Top. Dat wordt niet beter op het moment dat die machine tegen ons gaat praten. Net zo min dat het beter wordt als er een juffrouw op die machine staat.

Interviewer: Ok, wilt u misschien nog iets meegeven, of eraan toevoegen?

Gerard: Ik zie met groot plezier jullie onderzoek aan. Laat de machine de machine zijn, en je framed je vraagstelling als een nerd vraagstelling door te vragen naar intelligente systemen, en het woord intelligentie betekent daar iets voor het design.

Interview transcripts Pablo Cesar

Interviewer: Hi Pablo

Pablo: Hello there

Interviewer: Our first question about intelligent interfaces is: How would you define an intelligent interface?

Pablo: I will define an intelligent interface as an interface that can adapt to the user context and to the user needs. Are you looking for intelligent interfaces or intelligent processes?

Interviewer: That would be the next question because we can think about the differences between intelligent interfaces and intelligent systems. What is like the difference between those two?

Pablo: Right, so the interface would be, whatever the user uses to access information for example. Where the system would be all the processes running behind services. So as I said before, an intelligent interface for me is the one that actually adapts to whatever are the user needs at one specific moment or in the context. Where then you will have intelligent services behind all of that, making sure that for example the appropriate information is coming to the user.

Interviewer: So the interface is thus actually in front of the system? As a shield?

Pablo: Right, The interface is the part of the system that is actually in front of the user. So this way the user has to interact with any computer devices or any other services.
But I have to say that it depends a lot actually of which kind of service you have in mind.
Because intelligence is really depending on the actual service that your system provides. The library system. A user wants to get videos. So the kind of intelligence system interface there are going to be completely different from the ones that for example in automate a home.

Interviewer: Yes, then you have to be contextually different in that aspect. I think that they left that area purposely vague, in that we have to figure it out.

Pablo: But you have an idea of where you want to go, in what direction?
Interviewer: At this moment we do not actually know, they have not been telling us.

Interviewer: I was assuming that it was going to be more based on user experiences with the computer. If you would like to strictly focus on that one,

because otherwise it gets to broad. Because if we are talking about intelligent interfaces of like a home video system for instance.

Interviewer: They really did not like specifically say this, but it could be everything but maybe the computer is the most interesting.

Pablo: I meant more, what is the the application area it which this intelligent system is going to be used for? Is it going to be used for communication? Is it going to be used in health? Is it going to be used for teaching?

Interviewer: Actually that is interesting, because the definition actually differs in different situations.

Interviewer: I think it is going be more difficult if it is like supposed to a system communication, like one-on-one communication and this is a Skype conversation. That interface, to me, is more difficult to define than one user interacting with one system. Because they are not using the system as a channel to get to another human being, so if it is like automated health, like symptoms, like detection or something. I think there it is easier for the builders to guess what the humans might try to interact with, than like we have to guess the computer response and to the other user if we are doing the communication channel.

Pablo: The thing is that, normally in this intelligent systems, if you do not have a clearly defined use case, then it is kind of impossible to define what is intelligence. Because intelligence has to be targeted towards a goal. We can not expect a machine that is intelligent. Because that does not mean anything. The same with an interface, interface intelligence, it depends what is the goal of that interface. So it can actually do things in order to have the user..

Interviewer: .. It depends on how much it meets the needs of the user actually if you can call it intelligent or not?

Pablo: Right

Interviewer: If you don't mind generalize things. We want to be more specific as well, but we don't know, they haven't given us the future purposes, right, so we just have to be a little more general. But so then for any interface, what would need to be present for the users to acknowledge that this is an intelligent interface?

Pablo: Right. So for example an interface that allows for personalization, I will say that that is an intelligent interface. So if whatever I see in that interface is different from whatever you have seen in that interface, then there will be some intelligence in there. Because obviously it is adapting to whatever I want to do. And if this is a video system then we are talking about recommender systems and that of course is suppose to be intelligence between recommender

systems.

Interviewer: And when is the system intelligent?

Pablo: In recommender systems itself, I will assume that the part that I see, is showing to me in a way that actually is helpful for me. And that could be... if, I mean in the car, I have a completely different situation than if I'm leaving my living room, showing to me that information is going to be different in one place, it might be only audio in the places, audio plus video plus..., and that would be behind a system that actually is doing my profiling.. searching for movies that I might be interested in or not, mapping that profile to the movies I'm interested or not... so actually helps the system to actually recommend me things that I will like to see.

Interviewer: Would you consider Netflix an intelligent system?

Pablo: I will consider TiVo as an intelligent system, but of course the intelligence of TiVo is that they record absolutely everything. And Netflix, I am very sure in that they have recommender systems behind them. I think that they are mostly about offering you mainstream content. They have iPlayer and in there they do have some intelligence in recommender systems, and they really use for example what your friends like or don't like.

Interviewer: Alright, oke, so we are going to the next question.

Interviewer: So, if it is an intelligent system, will the users over time learn to view artificial intelligence as a distinct type of intelligence?

Pablo: Wow, well I think users they don't even need to know that artificial intelligence exists so. I mean...

Interviewer: Because otherwise if they know like... if they are not aware of it anymore, you can call that maybe really intelligence?

Pablo: Right, it has to be absolutely seamless. Whether processes are happening in the back. Whatever is happening it has to be completely seamless to the user, the user doesn't need to see any of that, doesn’t even need to notice. I mean if I'm feeling that the person is intelligent , if I have my focus on recognizing how intelligent you are.. I mean the communication is completely lost, because we are not focusing on the real task that you are about to bring.
Interviewer: I think that that is a really nice definition.

Interviewer: Alright, so then, how can users overall benefit from intelligent interfaces suppose to none intelligent interfaces?

Pablo: I would say that if indeed intelligent interfaces adapt to my situation, then they are going to make my life easier and then I would say the user is going to be much more happier.

Pablo: Or was that not the direction of the question?

Interviewer: That was, I was just thinking, how realistic is that expectation in our daily life.

Pablo: I think that is very realistic in people's real life, that intelligent systems they are extremely restricted to specific situations. So if we don't try to solve all the problems, but we try to solve only very specific problems. I mean intelligence in a machine is basically what you program, it is nothing more, so you have to be restricted to a number of use cases, situations, contexts and you will radically fail many times so a lot of experimentation needs to happen.

Interviewer: Do you know an example of a system that is used to be beneficial, but is actually not working in practice? So a counter example of what should be actually beneficial, but is not?

Pablo: So I use to study artificial intelligence from the past, quite a lot, because it was part of my studies back in the Spain actually. And in there we will have all the expert systems. That was intelligent systems that you will basically... create environments that they solve problems by themselves... and I never really seen actual solutions coming out of there.

Interviewer: Never?

Pablo: Not really. I mean artificial intelligence that served people. It was so broad, so trying to basically replicate what we believe is human intelligence, instead of really understanding that machine intelligence is something else. That it never went really anywhere. Nowadays you see systems implemented in search and industries, and they are specifically for one task. And they work very well, they are really helpful, they improve safety. And they are really working, but of course they are much, much restricted. It is not like I have a machine and the machine can communicate with me and it can answer all my questions. My question is now, am I going to be rich next year?

Interviewer: So intelligence should be applied for very specific tasks, then it can work good?
Pablo: That is my own experience,. but it should also be extremely well tested in these situations. Thus the problem of course of intelligence is that my context now is different from my context in two minutes. Which means that you have to try so many different alternatives on things, that you can not say it works, just because now it works at a particular moment. Three hours later the system in a different situation can crash suddenly.

Interviewer: Great. I do not have any more questions. Do you have anything to add?
Pablo: Anything that will be helpful for you is studies of course.