

Culture and Facial Expressions: A Case Study with a Speech Interface

Beant Dhillon, Rafal Kocielnik, Ioannis Politis, Marc Swerts, Dalila Szostak

► **To cite this version:**

Beant Dhillon, Rafal Kocielnik, Ioannis Politis, Marc Swerts, Dalila Szostak. Culture and Facial Expressions: A Case Study with a Speech Interface. Pedro Campos; Nicholas Graham; Joaquim Jorge; Nuno Nunes; Philippe Palanque; Marco Winckler. 13th International Conference on Human-Computer Interaction (INTERACT), Sep 2011, Lisbon, Portugal. Springer, Lecture Notes in Computer Science, LNCS-6947 (Part II), pp.392-404, 2011, Human-Computer Interaction – INTERACT 2011. <10.1007/978-3-642-23771-3_29>. <hal-01590881>

HAL Id: hal-01590881

<https://hal.inria.fr/hal-01590881>

Submitted on 20 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Culture and Facial Expressions: A Case Study with a Speech Interface

Beant Dhillon¹, Rafal Kocielnik¹, Ioannis Politis¹, Marc Swerts², Dalila Szostak¹

¹Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, Netherlands

²Tilburg University, Warandelaan 2, 5037 AB Tilburg, Netherlands

¹{B.K.Dhillon, R.D.Kocielnik, I.Politis, D.Szostak}@tue.nl

²M.G.J.Swerts@uvt.nl

Abstract. Recent research has established cultural background of the users to be an important factor affecting the perception of an interface's usability. However, the area of cultural customization of speech-based interfaces remains largely unexplored. The present study brings together research from emotion recognition, inter-cultural communication and speech-based interaction and aims at determining differences between expressiveness of participants from Greek and Dutch cultures, dealing with a speech interface customized for their culture. These two cultures differ in their tendency for Uncertainty Avoidance (UA), one of the five cultural dimensions defined by Hofstede. The results show that when encountering errors, members of the culture that ranks higher in the UA scale, i.e. Greeks, are more expressive than those that rank low, i.e. Dutch, especially when encountering errors in a low UA interface. Furthermore, members of the high UA culture prefer the high UA interface over the low UA one.

Keywords: Multicultural study, culture, cultural differences, uncertainty avoidance, expressiveness, speech interface.

1 Introduction

1.1 Culture

There are more than 150 ways in which culture has been defined over the years [22]. It is a broad concept that incorporates a wide range of features. Merriam-Webster online defines Culture as “customary beliefs, social forms, and material traits of a racial, religious, or social group.” Various new terms have been created by different authors [14,16,31] to break up culture into smaller components and study the differences. Fons Trompenaars, a Dutch management consultant [31], defines the various levels of culture, namely, National, Corporate and Professional. Hall [14] presented a four-factor model which distinguishes between cultures in terms of low vs. high context, multi-tasking vs. single tasking etc. Further, Hofstede [16] defines Culture in terms of five dimensions which include Masculinity vs. Femininity, Individualism vs. Collectivism among others. Though these cultural models may have been defined at different times by different researchers, at times, they tend to be

overlapping or refer to similar components of culture. In fact, Khaslavsky [20] demonstrated an instance of merging the existing cultural models by Hall [14], Trompenaars [31] and Hofstede [16] for applications in interface design. The present study is based on Hofstede's model thus, it relates to exploring differences between national cultures. The following section discusses some of the empirical studies carried out in context of culture and interface design.

1.2 Culture and Interface Design

Culture has been shown to affect perception [25], judgement [12] and decision making [5]. Moreover, several studies have shown how the cultural background of users affects their perceptions as well as preferences for interfaces. Researchers Cyr and Trevor-Smith [6] analyzed design elements such as the use of symbols and graphics, color preferences, and site features for 30 municipal websites in Germany, Japan, and the U.S to find out the effect of design elements in culture. They concluded that layout, symbols, navigation, and the use of color are amongst important design elements that vary across cultures and need to be taken into account for localization. Singh et al [29, 30] provided empirical evidence that in comparison to standardized websites, consumers from Germany, Spain, Netherlands, Switzerland, Italy, China, and India preferred culturally adapted websites. As described, there are several studies that emphasize the importance of cultural customization of graphical user interfaces to the cultural background of users [7, 27, 28] but research dealing with cultural customization of speech interfaces has not led to very conclusive results so far [24].

1.3 Facial Expressions, Emotion and Speech Interfaces

Speech based interaction started with telephony but people quickly found telephonic conversations with machines (voicemail, booking systems) rather uncomfortable. Because of that reason, systems with human facades which try to mimic human interaction are becoming a preferred way of interaction [26]. One way of making the interaction more human-like is to develop systems that can interpret the facial expressions of humans to determine their emotional state and adapt their behavior and respond accordingly (Affective Computing). There are technological solutions available that can already provide a robust estimation of a person's emotional state by interpreting facial expressions [2], [11], [19]. Further, existing research suggests that interpretation of certain facial expressions (e.g. frustration) could help a system deal with errors more effectively [4]. Along the same line, the present study aimed to explore if members with different cultural backgrounds differ in their expressiveness towards speech-based interfaces adapted for culture. Based on the results, a few pointers are provided for cultural customization of a speech-based interface.

The following section describes the cultural dimension on which the interface design was based as well as the rationale for choosing Dutch and Greek culture. Since Hofstede's cultural model is one of the most well-known and widely applied models

in both the field of management as well as interface design [8, 21], it was used for the purposes of the current study.

1.4 Culture and Expressiveness

One of the cultural dimensions defined by Hofstede [16] that is related to frustration is Uncertainty Avoidance (UA). Uncertainty Avoidance Index (UAI) is used as a measure of UA, i.e., the higher the UAI of a country, the higher the UA tendencies of members of that culture. Cultures with high UA are associated with higher anxiety, stress, avoidance of ambiguity and tendency to be normative, as emphasized in rules. Hofstede [16] also suggests that cultures rating high on UAI are emotionally more expressive. As he reports, this might be because people from cultures with high UA experience higher stress and have less internalized emotional control. On the contrary, people living in cultures with low UA tend to experience lower stress and less conflict between norms and experience. Due to this “weak superego” and acceptance of deviation, people from these cultures tend to display less emotion [13]. According to Hofstede’s research, Greek culture has the highest UAI, i.e. 112, while Dutch has a considerably lower score in the same index, i.e. 53, (mean UAI=66.4, min=8, max=112). Since the UAI of Greeks and Dutch differ substantially, representatives of these two cultures were chosen as participants in the study.

2 Hypotheses

As described above, research suggests that emotional expressiveness differs across cultures. Additionally, guidelines have been developed for adapting interfaces to different cultures [23], based on the assumption that such an adaptation will lead to a more positive user experience. This study focused on determining whether there are differences in facial expressiveness of members of different cultures when they interact with an error prone interface. Further, the study tested whether the level of expressiveness differs when the interaction involves an interface which is adapted to one’s culture vs. one that is not. Thus, intensity of expression was the dependent variable. These expectations are justified by the researchers’ intuition that any deviation from the expected system behavior would elicit a reaction from the user and the reaction would, very likely, include a facial expression. This deviation was achieved by introducing errors in the system which resulted in an error-prone interaction. Pilot studies by independent evaluators showed that intensity of expression was the variable that most clearly reflected the perceived difference between conditions where there was an error encountered and conditions where there was no error. Two different versions of an interface that varied in the dimension of UA were designed. This enabled the researchers to assess differences in expressiveness during the interaction with each version. A more expressive response was expected for an interface that did not correspond to the cultural expectations of the user. Moreover, a general tendency towards higher expressiveness was expected in high UA cultures. The following hypotheses were formulated to be tested during the study:

- H1: When encountering errors, high UA cultures will overall be more expressive than low UA cultures.
- H2: When encountering errors in a low UA interface, high UA cultures will be more expressive than low UA cultures.
- H3: When encountering errors in a high UA interface, low UA cultures will be more expressive than high UA cultures.

3 Methods

3.1 Experiment Design

The study was set up as a 2X2 mixed factorial design with two independent variables: culture and type of interface. Culture was divided into Greek and Dutch and type of interface was divided into high UA and low UA. The dependent variables were frustration experienced by participants and facial expressiveness. UAI for each culture group was measured using the Values Survey Module 94 (VSM94) questionnaire, frustration was measured with the NASA Task Load Index (NASA-TLX) questionnaire and expressiveness was measured through a perception test.

There are some disadvantages of having a 2X2 mixed design, e.g. learning effects and fatigue. However, the advantage of the design is that a lower number of participants is required. Additionally, statistical power increases since participants' individual differences are kept constant between conditions.

3.2 Application Design

The study was conducted using a Wizard of Oz approach which involved interacting with a speech based Train Booking System. One of the experimenters controlled the system as a response to the users' utterances where the participants were asked to make a booking from London to Newcastle for a specified date and time. Two versions of the system were developed, one customized for a high UA culture and the other for a low UA one. The interfaces were developed using the guidelines available for cultural customization of Graphical User Interfaces [23]. Guidelines were adapted for use in a speech only interaction. Furthermore, during pilot testing, two independent reviewers were asked to interact with the interfaces and the interaction was modified according to the comments provided by reviewers. For the high UA interface, the following recommendations were applied:

- Guided interaction with confirmations at every step.
- Explicit cues to prevent users from feeling lost.
- Restricted amounts of data.
- Limited choices

The following is an example of the high UA interface used in the study:

System: Welcome to the Train Booker! Where do you want to depart from?

User: London.

S: Ok, you are departing from London. What is the destination? You can always go back or cancel.

U: Newcastle.

S: Ok then, what day do you want to travel between London and Newcastle?

U: Sunday.

S: And what time do you want to travel on Sunday?

U: 12 o'clock.

S: So you are traveling from London to Newcastle on Sunday around 12 o'clock, right? You can always go back or cancel.

U: Right.

...

For the low UA interface, the following recommendations were applied:

- Free interaction with fewer confirmations
- More flexible navigation
- Variety of choices

The following example demonstrates the low UA interface used in the study:

S: Welcome to the Train Booker! You can start with places or times, which do you prefer?

U: Places.

S: Between which places do you wish to travel?

U: From London to Newcastle.

S: What day and time do you want to travel? You can always go back or cancel.

U: Sunday at 12 o'clock.

...

Two types of errors, i.e. misrecognition and repetition, were introduced. In the first case, the system would respond as if it had misheard the participant, e.g.:

S: What class would you like to travel with?

U: Second class.

S: First class then?

...

In the second case, the system would ask the participant to repeat what he/she had last said, e.g.:

S: Do you wish to take a direct or an indirect train?

U: A direct one.

S: Sorry, can you repeat your train preference?

...

An equal number of misrecognition and repetition errors were induced in both high and low UA versions of dialogue, two of each type for each interface. As a result, each user would encounter, in total, 4 errors per interface.

3.3 Participants

Participants were recruited from the Eindhoven University of Technology. There were 16 participants, 8 Greek and 8 Dutch. All the participants selected for the study were university students as the VSM94 questionnaire is validated only for participants from similar professional backgrounds. Their age ranged from 18-30. The participants were asked to fill in the VSM94 designed by Hofstede [17] in order to measure the UAI and make sure their responses were in line with Hofstede's findings. Participants were offered coffee/tea with a snack as well as a small reward (5 Euro Coupon) for participation. Details on participants' demographics can be found in Table 1.

Table 1. Demographics for participant study.

| Nationality | UAI* | Average age | Average time abroad | Gender | |
|-------------|-------|-------------|---------------------|--------|--------|
| | | | | Male | Female |
| Greek | 105.0 | 24-29 | 1-3 years | 5 | 3 |
| Dutch | 40.7 | 20-25 | 0-1 years | 6 | 2 |

* UAI stands for Uncertainty Avoidance Index measured for this group

3.4 Questionnaires

Two questionnaires were utilized during the experiment: the Values Survey Module (VSM94) and the NASA Task Load Index (NASA-TLX).

The VSM94 [17] is a 26-item questionnaire developed by Geert Hofstede for comparing cultures, based on region or country, across 5 dimensions: Power Distance, Individualism, Masculinity, Uncertainty Avoidance, and Long-term Orientation. 20 questions (4 questions for each) are used for the dimensions, and the remaining six questions are demographic. Scores for each dimension are computed on the basis of four questions. Each question is scored on a five point scale and the index scores are derived using the mean scores of respondents from these questions.

The results are computed using a formula provided with the questionnaire. Score for each dimension is calculated using mean scores for related questions for a group of participants. For calculating UAI, Formula 1, as provided below, is used:

$$\text{UAI} = +25 * \text{mean Q13} + 20 * \text{mean Q16} - 50 * \text{mean Q18} - 15 \text{ mean Q19} + 120. \quad (1)$$

In the present study, the UAI score for Greek participants was found to be 105 and 40.7 for Dutch participants. A large difference between these scores supports the selection of Dutch and Greek participants based on differences in UAI.

The NASA-TLX is a tool that allows users to perform subjective workload assessments of tasks performed with human-machine systems [15]. The ratings are collected on six sub-scales: Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort and Frustration. The overall workload is computed from weighted averages of these ratings.

3.5 Procedure

Participants filled in the VSM94 questionnaire after which they were briefed about the experiment by one of the investigators. While interacting with the interface, the participants were video recorded using the built-in webcam of the computer. The average duration of the interaction was 2:03 minutes for the low UA interface (Std. Dev. 0:17) and 2:22 minutes for the high UA interface (Std Dev. 0:14). After having completed the task with the first interface, participants were asked to fill in the NASA-TLX questionnaire which was followed by the interaction with the second interface. The order of interfaces was balanced across subjects to counterbalance for learning effects. After the second interaction, participants were once more asked to fill in the NASA-TLX questionnaire followed by a debriefing session as well as a short interview session. The location and set up of the study was chosen in order to recreate a comfortable and friendly environment for participants. All the questionnaires were used in their entirety to preserve validity.

3.6 Perception Study

Perception study is a widely used method for evaluating facial expressiveness, in terms of type and/or intensity of expression [3], [9], [10]. It involves showing a series of videos or pictures to the participants and asking them what kind of expression is perceived by the viewer and/or rating how intense an expression is.

The videos recorded during the previous phase of the experiment were edited and instances where a participant encountered an error with the system were extracted. An equal number of videos from Dutch and Greek participants with misrecognition and repetition error were used. The situations considered as erroneous were only the ones where the users explicitly repaired the errors by repeating their command or correcting the system. At times, the participants did not correct the mistake made by the system, giving no evidence of realizing a misrecognition error. In such cases, their videos were not included in the perception test. Also, for each participant, one video with a neutral expression, i.e. a facial expression in a neutral, non-erroneous context, was included in the perception test. The neutral videos were selected from dialogue

fragments without errors, and without clearly perceivable muscle movements within a timeframe of 5 seconds, following best practice procedures outlines in related studies [3]. Neutral videos were used later as a baseline in terms of expressiveness. The results from the perception study validated this set-up.

The total number of clips presented during the perception test was 56. Each clip was about 2 seconds long. The participants of the perception study rated the intensity of the expression shown in each video on a scale of 1 to 5 (from neutral to very strong). All 17 participants for the perception study were employees of Eindhoven University of Technology (TU/e). In order to balance possible cultural differences in expressiveness perception, participants from at least 9 different cultures were selected. These participants did not have any information regarding the nationality of the participants in the videos or the context of the interaction. Furthermore, none of them had participated in the initial experiment.

4 Results

4.1 Quantitative

The ratings obtained from the perception study were averaged for each rater by video type and culture. This resulted in 6 average scores for each rater (2 cultures X 3 Video types): 3 videos of the Greek participants (high UA, low UA and Neutral), and the same for the videos of the Dutch participants. This procedure was followed for all 17 scorers. These scores were analyzed using a 2X3 repeated measurements ANOVA with 2 independent factors (nationality - 2 levels; video-type - 3 levels). The results display significant main effects for nationality $F(1,16)=27.109$, $p < .001$. Thus, hypothesis H1 is supported.

Further, video type $F(2,15)=125.466$, $p<.001$ as well as an interaction between nationality and video type $F(2,15)=31.804$, $p<.001$ were found to be significant, both for low UA and high UA levels. To tease apart this interaction, a post hoc pair wise comparison (Bonferroni test) was conducted. This test revealed that the perception ratings for neutral video type differ significantly from the two other versions (low and high), but they don't differ among themselves. Furthermore, unlike high UA or low UA conditions, the neutral videos scored close to 1 (No expression) for both cultures. The mean score, on a scale of 1-neutral to 5-very strong, was 1.309 (Std. err: 0.070) for Greek participants and 1.322 (Std. err: 0.092) for Dutch participants. The neutral videos were indeed rated as having 'no expression', thus validating the set-up for the study. The bar graph (Fig. 1) presents the mean scores from perception test for each video with respect to each culture.

The results of pair wise comparison for the main effect of nationality and of the interaction between nationality and video-type can be found in Tables 2 and 3. A spliced analysis confirmed that the differences between UAI levels are significant (using separate t-test), both for Dutch and Greek participants separately ($P < .001$). Thus, the null hypothesis can be rejected and hypothesis H2 and H3 are supported.

Table 2. Pair wise comparison for expressiveness by nationality.

| Nationality | Mean | Std. Error | 95% Confidence Interval | |
|-------------|-------|------------|-------------------------|-------------|
| | | | Lower Bound | Upper Bound |
| Greek | 2.268 | 0.074 | 2.111 | 2.425 |
| Dutch | 2.077 | 0.079 | 1.909 | 2.245 |

Table 3. Interaction between variables nationality and context of interaction (video clips from Low UA interface, Neutral, High UA interface respectively).

| Nationality | Video Type | Mean | Std. Error | 95% Confidence Interval | |
|-------------|--------------------|-------|------------|-------------------------|-------------|
| | | | | Lower Bound | Upper Bound |
| Greek | Low UA interface* | 2.900 | 1.109 | 2.669 | 3.131 |
| | Neutral | 1.309 | 0.070 | 1.161 | 1.456 |
| | High UA interface* | 2.597 | 0.092 | 2.402 | 2.791 |
| Dutch | Low UA interface* | 2.230 | 0.094 | 2.031 | 2.429 |
| | Neutral | 1.322 | 0.092 | 1.126 | 1.518 |
| | High UA interface* | 2.678 | 0.088 | 2.491 | 2.865 |

* Low/High UA interface: video clip from interaction with low/high UA interface

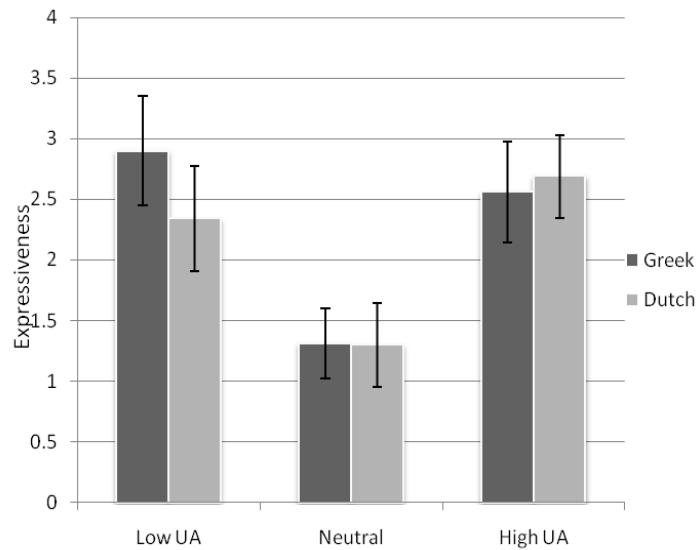


Fig. 1. Perception test results – expressiveness mean values for video excerpts from each interaction context (Low UA interface, Neutral, and High UA interface) and nationality.

Finally, the values from the NASA TLX Questionnaire were analyzed using multivariate ANOVA. Culture and system were the independent variables, and workload, frustration and frustration weights were the dependent variables. No significant main effects were found for this interaction.

4.2 Qualitative

According to the statements from users in the semi-structured interviews conducted after each experiment, five out of eight Greeks preferred the high UA interface, commenting that they “liked the step - by - step and uncomplicated structure and the clear confirmations provided”. Five out of eight Greeks disliked the low UA interface, describing it as “stressful and frustrating” and claiming that it presented them with “too few confirmations”. It is worth noting that none of the Greek participants preferred the low UA interface. Concerning Dutch participants, the preferences were equally distributed between the interfaces. Thus there are no strong conclusions to be made for Dutch people out of the interviews conducted. Only some of the participants expressed strong opinions about their preferences or dislikes when asked. Others, however, had no strong attitude towards either type of interface. Tables 4 and 5 summarize the main findings of this qualitative analysis.

Table 4. Preference towards a type of interface based on comments from interviews.

| Nationality | Total | Preference | | |
|-------------|-------|-------------------------|--------------------------|-------------|
| | | <i>Low UA interface</i> | <i>High UA interface</i> | <i>None</i> |
| Greek | 8 | 0 | 5 | 3 |
| Dutch | 8 | 3 | 3 | 2 |

Table 5. Dislike of a type of interface based on comments from interviews.

| Nationality | Total | Dislike | | |
|-------------|-------|-------------------------|--------------------------|-------------|
| | | <i>Low UA interface</i> | <i>High UA interface</i> | <i>None</i> |
| Greek | 8 | 5 | 0 | 3 |
| Dutch | 8 | 3 | 0 | 5 |

5 Discussion

Based on the aforementioned results, it can be concluded that the three alternative hypotheses, i.e. H1, H2 and H3, are all supported. When encountering errors, cultures that score high on the UA index are more expressive than cultures with low UA index. Also, when encountering errors in a low UA interface, high UA cultures are more expressive than low UA ones. This holds true for low UA cultures encountering errors in a high UA interface too, but the results are not as strong.

The results suggest that Greek participants, who belong to a high UA culture, prefer interfaces with more confirmations and a step-by-step guided navigation. They have stronger reactions when encountering errors during an interaction with an interface with fewer confirmations. This was confirmed by the exit interviews where five out of eight Greek participants commented that they disliked the low UA interface and five out of eight expressed a liking for the high UA interface.

The Dutch participants seemed to have slightly stronger reactions when they encountered errors in the high UA interface. However, as discovered during the interviews, three Dutch participants disliked the low UA interface and three of them liked it, thus no strong conclusions can be derived from this observation. Overall eight out of sixteen participants disliked the low UA interface, conveying that if no clear information regarding the culture is available, it is probably better to provide an interface designed for high UA.

These findings can be helpful in building systems which detect the expressiveness of a user and adapt their features, e.g. a navigation style that suits the needs of the user. If an interface could detect the UA (based on expressiveness) of its user, especially in context of errors, the interface could adapt its communication style to enable a less frustrating and more efficient interaction. An example could be the case of a virtual agent - based system for flight booking.

If the system detects a high level of frustration in the user, then, it could adapt itself to another context more suited to the user, e.g. adapt its navigation style to become more flexible. Since such a system takes into account the individual differences of the user, this adaptation could even be independent from the cultural background of the user and rather focus on the individual reactions of the user without binding the users into cultural stereotypes. Since the present study was conducted using a speech interface, care should be taken while extending them to other kinds of interfaces.

Another interesting finding was the fact that the results from the NASA TLX questionnaire provide minimal insights into participants' preferences for an interface. There was no match between the participant ratings for frustration in the NASA TLX questionnaire and their comments or the way their expressions were perceived in the perception test. Thus, this finding demonstrates possible limitations of quantitative methods when it comes to providing insights into human behavior, especially in relation to human computer interaction.

5.1 Possible bias

The results of this study could have been affected by various factors such as cultural adaptation and the lab set up. Since culture was an important aspect of the study, it is essential to note that the study was conducted in the Netherlands and Greek participants living in this country could potentially have shown different behavior from Greeks living in Greece, although measuring the UAI might have taken care of this bias to some extent. While Hofstede's cultural model advises against making statements from responses from less than 20 participants from a given country, the researchers found a substantial difference between the two cultures even for 16 participants. Finally, testing in a lab environment with a Wizard of Oz approach could have introduced bias in comparison to usual context of interaction. Although the researchers are aware of the limitations of Wizard of Oz, it was still chosen as the best option for this type of research because it facilitated a controlled testing of specific hypothesis. Since it was required to control the number and type of errors that the participants experienced during the interaction, a script-based interaction was a good fit. We note that 14 out of the 16 participants stated they were convinced they

interacted with a real system, even though they were briefed that the interaction dialogue was being tested and not the system. The remaining two participants did not know that their expressiveness would be analyzed and their expressions didn't appear fake or staged either.

The sample size could be another point of concern in this study, but it should be noted that our initial work with a limited number of participants already displays observable differences. It can be expected that these promising effects will only become clearer when larger groups would be analyzed.

5.2 Conclusion and Future Work

Facial expressions are very important in the field of human-machine interaction, not only as an output modality (e.g. avatars), but also as an input modality. Automatic speech recognition is known to improve when systems interpret information from multiple modalities, but facial expressions as a modality have not been explored in depth in HCI. Facial expressiveness constitutes a big part of human communication and as language and gestures differ across cultures, so might facial expressiveness in the context of machine interactions. This study contributes to the area since it explores and finds out that these presumed differences are indeed true, for the dimension related to expressiveness and do not just refer to prejudices. Furthermore, this study demonstrates that such differences could be exploited to improve human-machine interactions by tuning interaction styles to specific groups of people. In this respect, the current study is both novel and relevant.

An important future step for this research could be to include a larger number of participants for both the test as well as the perception study. A larger number of participants would make the cultural difference statement stronger, regardless of the strong statistical values that the current research shows. Additionally, testing other cultural dimensions with similar interfaces that produce controlled errors, or testing the same dimensions but in a real interface that is naturally prone to errors could also be of interest for further research. Finally, testing Greek participants in their native country and comparing the results to those of the described study could provide richer insights in this field.

References

1. Ailon, G. Mirror, mirror on the wall: Culture's Consequences in a value test of its own design. *The Academy of Management Review*, 885-904. (2008)
2. Balomenos, T., Raouzaïou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S. *Emotion Analysis in Man-Machine Interaction Systems. Machine Learning for Multimodal Interaction. Springer-Verlag*, 318 – 328. (2004)
3. Barkhuysen, P., Kraemer, E., Swerts, M. Problem detection in human-machine interactions based on facial expressions of users, *Speech Communication*, 45 Elsevier, 343–359. (2005)
4. Barkhuysen, P., Kraemer, E., Swerts, M. *Audiovisual Perception of Communication Problems, Speech Prosody International Conference* (2004)

5. Briely D.A., Morris,M., Simonson, I. Reasons as carriers of culture: dynamic vs. dispositional models of Cultural influence on decision making. *Journal of Consumer Research*, Vol. 27. (2000)
6. Cyr D, Trevor-Smith H. Localization of Web design: an empirical comparison of German, Japanese, and US Web site characteristics. *Journal of the American Society for Information Science and Technology*. Volume 55. Issue: 13. Page: 1199 – 1208. ISSN: 1532-2882. (2004)
7. Cyr D, Kindra GS, Dash S. Web site design, trust, satisfaction and e-loyalty: the Indian experience. - *Online Information Review*. Emerald Group Publishing Limited (2008)
8. Dhillon, B. Effects of cultural background of users on interface usability, VDM Verlag Muller, London,UK (2010)
9. Ekman, P., Friesen, W.V. A new pan-cultural facial expression of emotion, *Motivation and Emotion* 10, 2, Springer 159-168, (1986)
10. Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology* , 53 (4), 712-717, (1987)
11. Caridakis, G., Malatesta, L., Kessous, L., Amir, N., Raouzaoui, A., Karpouzis, K. Modeling naturalistic affective states via facial and vocal expressions recognition. *International Conference on Multimodal Interfaces (ICMI'06)*. Alberta, Canada, (2006)
12. Gardner W. L., Gabriel S, Lee A.Y. "I" Value Freedom, But "We" Value Relationships: Self-Construal Priming Mirrors Cultural Differences In Judgment. *Psychological Science* July 1, 1999 vol. 10 no. 4 321-326. (1999)
13. Gudykunst, W. a. Bridging Japanese/North American differences. Sage Publications, Inc., (1994)
14. Hall ,Edward. The silent language. Anchor Books (1990)
15. Hart, S. Nasa-task load index (nasa-tlx) : 20 years later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Human Factors and Ergonomics Society, 904-908, (2006)
16. Hofstede, G. Culture and organizations. *International Studies of Management & Organization*, JSTOR , 15-41, (1980)
17. Hofstede, G. Values Survey Module manual. University of Limburg, Maastricht, the Netherlands, (1994)
18. Hofstede, G.H., Hofstede, G. Culture's consequences: Comparing Values, Behaviours, Institutions, and Organisations Across Nations. Sage Publications, Inc., (2001)
19. Jarkiewicz, J., Kocielnik, R., Marasek, K. Anthropometric Facial Emotion Recognition. 13th International Conference on Human-Computer Interaction, Part II: Novel Interaction Methods and Techniques. Springer-Verlag, (2009)
20. Khaslavsky, J., Integrating culture into interface design. in CHI 98 conference summary on Human factors in computing systems, (Los Angeles, California, 1998), ACM Press, 365-366, (1998)
21. Kondratova, I., and Goldfarb, I. Proceedings of EdMedia 2005 - World Conference on Educational Multimedia, Hypermedia & Telecommunications. Montréal, Québec, Canada.. pp. 1255-1262. NRC 48237. (2005)
22. Kroeber, A. L. and C. Kluckhohn. Culture: A Critical Review of Concepts and Definitions. (1952)
23. Marcus, A. Cultural dimensions and global web user-interface design: What? So what? Now what? http://www.tri.sbc.com/hfweb/marcus/hfweb00_marcus.html
24. Matousek, V.; Mautner, P.; Moucek, R.; Tauser, K. (Eds.). Text, Speech and Dialogue. XIII, 452 p., Softcover. 4th International Conference, TSD 2001, Zelezná Ruda, Czech Republic, September 11-13, 2001. Proceedings. (2001)
25. Nisbett R. E. and Miyamoto Y. The influence of culture: holistic versus analytic perception. *TRENDS in Cognitive Sciences* Vol.9 No.10. (2005)

26. Pitt, I., Edwards A.D.N. Design of speech-based devices: a practical guide. 179 p., Hardcover ISBN: 978-1-85233-436-9. (2003)
27. Reinecke K. and Bernstein A. Culturally Adaptive Software: Moving Beyond Internationalization. In Proceedings of the 12th International Conference on Human-Computer Interaction, Beijing, China. Springer (2007)
28. Sheppard, C., & Scholtz, J. The effects of cultural markers on Web site use. Proceedings of the Fifth Conference on Human Factors & the Web, Gaithersburg, Maryland. (1999)
29. Singh N., Fassott G., Zhao H, Boughton P. D. A cross-cultural analysis of German, Chinese and Indian consumers' perception of website adaptation. Journal of Consumer Behaviour Volume 5 Issue 1, Pages 56 – 68. (2006)
30. Singh N., Pereira A. The culturally customized website – Customising websites for the global workplace. Elsevier - Butterworth Heinemann, MA, USA (2006)
31. Trompenaars, F., & Hampden-Turner, C. Riding the Waves of Culture: Understanding Cultural Diversity in Business (2nd edition). London: Nicholas Brealey. (1997)
32. Zeng, Z., Pantic, M., Roisman, G., Thomas, S., Huang. A. Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39-58. (2009)