User Experience and Social Attribution for an Embodied Spoken Dialog System

Benjamin Weiss and Simon Willkomm

Quality and Usability Lab, TU Berlin, Germany
`Benjamin.Weiss@tu-berlin`,
home page: `http://qu.tu-berlin.de`

Abstract. A public information system with an Embodied Conversational Agent is evaluated in a laboratory setting concerning Social Actorship, Social Acceptance, perceived Control, Pragmatic Quality and Hedonic Qualities. Results show a positive experience for Pragmatic Quality and Control, but negative ratings for Social Acceptance. Differentiating these various aspects of User Experience has proven to be fruitful for this summative evaluation, especially considering the potential public situation of interaction.

Keywords: Embodied Conversational Agents, Social Actorship, Spoken Dialog System, User Experience

1 Introduction

Spoken dialog systems (SDS) can provide a natural and intuitive way of interacting due to an interface operated by voice. Embodied conversational agents (ECAs) also use spoken language to interact, but in addition exhibit at least an anthropomorphic interface, for example by visually modeling a human face. From a user point of view, embodiment can result in increased expectations on the capabilities of the ECA, assuming for example social skills and intelligence, which should be reflected in sophisticated (human-like) communication behavior. If such expectations are not met, user experience will be negative.

But the embodiment might also result directly in positive user experience (UX): The multimodal stimulation itself (typically audio-visual for non-robot embodiments) can be positive. It also might increase user attention and thus facilitate interaction with such a system. Additionally, embodiment enables designers to present an attractive interface for more than the acoustic modality. Concerning expectations, assumed social and cognitive capabilities attributed to an ECA will be beneficial when such expectations are not disappointed.

The main objective of this paper is to evaluate UX, social abilities in particular, of an embodied visitor guide.

This virtual visitor guide is a speech operated system with an audio-visual synthesis in the form of a lip-synchronous talking head. Its purpose is to inform visitors in a welcome and demonstration hall about research and development projects. Typical visitors received in this hall are student groups, prospective students, colleagues from industry on a company outing, professors and managers on a collaboration visit and at last, Berlin citizens and tourists on the annual "long night of science".

By enabling spoken conversation and showing literally a human like face, the virtual guide is supposed to motivate and support interaction and interest and provide an interacting mode which ...

- ... is complementary to visual information on posters,
- activates the visitor (as s/he has to talk to the guide instead of just read the posters available),
- and sends visitors to demonstrators and thus actives visitors to explore.

2 Attribution of social abilities

Researchers from various disciplines have used different approaches on their own definition of UX [1]. Consequently, the aim of defining a standardized definition of UX resulted in "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [2], which incorporates every aspect of perception and response concerning the usage of an interface.

The focus of User Experience is on any experience of users during interaction with a system. It is not limited to conscious (retrospective) reflections on the usability or usefulness of a given service operated with an interface, but concentrates on sub-conscious affective reactions of the interaction, which can, of course, be asked for in retrospection. This paradigm shift on the last decades aims at understanding the user better, especially event-driven affection ("Wow Effect", frustration) and sometimes confusing decisions concerning, e.g., user acceptance of certain devices based mainly on the big impact of aesthetics or Social Norm [3].

Although dimensions of UX are not fully understood [4], the separation of overall attractiveness (how positive or negative a user rates a device or interface) into one pragmatic (how usable or useful) and two hedonic qualities [5] seem to be quite established. These two hedonic dimensions are Identification – how much can a user identify with a device/interface – and Stimulation – how interesting, exciting is using this device/interface. A questionnaire assessing these dimensions is also already provided.

Still, other dimensions or more concrete aspects of UX are of interest, especially when dealing with embodied spoken interaction and with interaction in public spaces. Social aspects come into play for such interfaces and usage situations, e.g., the attribution of *social actorship* and the experience of *interacting in a social context.*

Whereas the former issue deals with assumed, expected or attributed competences towards the system (e.g., intelligence, intentionality, awareness), the latter issue deals with the user feelings concerning privacy, control, or social acceptance. This view is actually a little different from the definition developed within the EIT RIHA 12124 "Computers as Social Actors" (2012) that subsumes both aspects mentioned under the term "Social Actorship":

> *Social Actorship is the ability of the system to act in a social context, with an implicit or explicit goal. From the user perspective, Actorship is a characteristic of the system that makes the user perceiving it as a human actor to which s/he can direct their attention and have attention in return (This can be explained by the Mirror Concept: the system that sense something and acts in response). Although, some systems could be seen as just a mediating actor, like mobile phone and ICT in general, that fosters social interaction among people. In this case social Actorship is seen as the ability to influence and support the social life of people.*

This definition also takes attribution of social abilities to a system/device/interface and the impact of such a system/device/interface on a user's social situation as two important aspects of UX. Therefore, a questionnaire was used to assess these aspects of UX, based on instruments and definitions available:

Attractiveness (ATT): The overall attractiveness of the system or interface after interaction. The difference to overall quality is the subjective aspect of attractiveness being not limited to pragmatic and general considerations, but including also hedonic subjectively experience aspects. (2 items [5])

Pragmatic Quality (PQ): The usability and usefulness of the system or interface. (4 items [5])

Hedonic Quality–Identity (HQI): The degree this system or interface fits to a user. This aspect is related to Social Acceptance. (2 items [5])

Hedonic Quality–Stimulation (HQS): The degree the interacting is positively stimulating. (2 items [5])

Social Acceptance (SA): User's social acceptance (according to [6]) subsumes how a user feels when interacting with a system regarding to the social situation, e.g. how uncomfortable or embarrassed in the light of potential other people or ones own norm. (5 items [7])

Social Actorship (SH): The degree the system exhibits social capabilities. (5 items [7])

Perceived Control (PC): The degree a user feels in control of the system and knows how to interact with it. (5 items [8])

The questionnaire provided by [8] is actually based on a model of technology acceptance described in [9].

3 Embodied conversational system

This ECA is embodied as a bald male person, based on the Thinking Head system [10]. The German text-to-speech system "OpenMary" [11] was chosen for the acoustic speech output and Sphinx as automatic speech recognition system [12]. The dialog was defined in VoiceXML running with Optimtalk [13]. The system itself is modular and uses events to let the modules communicate with each other.

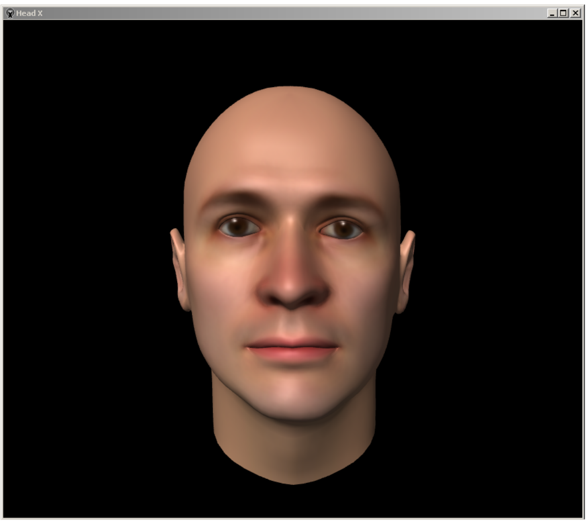

Fig. 1. Graphical representation of VirtualK.

The chosen visual appearance is a bald male talking head, determined in an informal pre-test with six participants in [14]. This embodiment also exhibits no photographic texture and represents the consensus, as it was considered most pleasant and least irritating (cf. Figure 1).

The ECA gives visual conversational feedback, i.e. a nod signals the processing of a user utterance and if the user is not recognized for 20 video frames, the ECA will close its eyes and stop/pause the conversation.

A webcam is used to detect a user within the interaction sphere of the system, and the ECA will open its eyes and initiate the dialog with general information, and by asking the user about the interest in one of four research fields (video, audio, smartphone apps, or mobile interfaces); however, only one out of two for the experiment conducted. The system provides project-related information either by project name or by suggesting a project based on the preferred topic (audio: music, communication; video: quality, mobile TV; apps: phone control and leisure time; mobile interfaces: security, cross-service). It is able to provide more project related information than the demonstrators and posters.

If a face is not recognized for 20 video frames, the ECA will again close his eyes.

For each project, there are two levels of information (and if available, using a demonstrator is offered): General description and additional information. After each block of information presented, the system asks whether it should proceed or not (see Figure 2 for a simplified scheme of the dialog).

There are actually two versions tested, a typical one and a system with user-centered adaption concerning user recognition after a break, remembering interest for project suggestions, confirmation strategy dependent on no matches

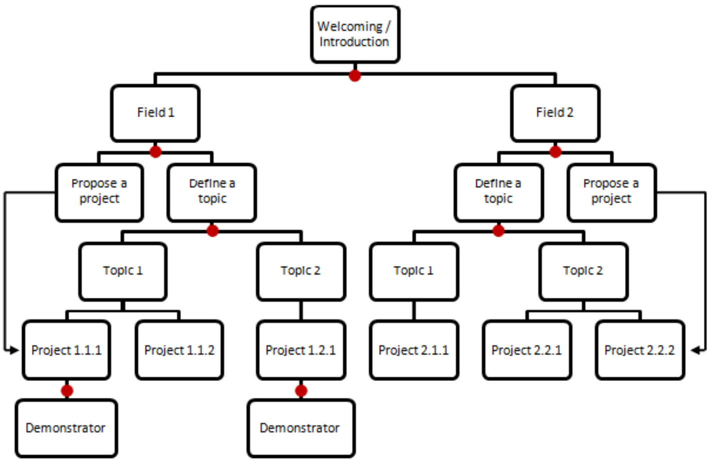

Fig. 2. The simplified dialog structure.

and confirmed false recognitions, and level of detail presented automatically. However, as there are no significant differences in the questionnaire data between both version, there will be no further description presented here.

4 Procedure

The aim of this evaluation is to assess User Experience and social capabilities attributed to the ECA in general and whether adaptive system components increase UX.

A laboratory experiment was chosen for this first evaluation regarding social aspects. The face recognition was set to a maximum by deleting previous users at the start of each experimental trial. For continuous duty, we lack information of the number of visitors a day, but it is expected to "forget" users after about four hours in order to successfully discriminate users. Also, the four research field were split into two categories, each comprising about half of the projects and demonstrators available to avoid boredom when trying out the system repeatedly.

A total of 30 test subjects took part in the experiment, gender balanced (14 female, 16 male), aged between 20 and 43 (average 26.4). All were paid for their contribution.

The initial experimental design was also planned for a comparison of the adaptive and non-adaptive version. Therefore, each user interacted two times with the system, each time with providing two of the fours research fields. The order of both research fields and order of adaptivity was balanced.

All users successively interacted with both versions of the SDS. They were asked to inform themselves about three to four projects and try out at least one demonstrator. Each individual experimental session took about one hour with roughly 15–20 min. for each interaction.

After each trial the test subjects answered a questionnaire comprising aspects of User Experience on 5-point scales (antonyms for the AttrakDiff [5] and a Likert scale for [7]) to subjectively test for a benefit of the adaptions. The AttrakDiff was also filled out in the beginning after a brief video to assess a user expectations. Also, the perceived ASR quality was assessed on one Likert scale after each interaction.

5 Results and Discussion

There are no differences between the adaptive and non-adaptive version on any of the questionnaire scales assessed, as well as for research field or position of adaptivity ($\alpha = .05$, repeated measures Anova). Therefore, the system is analyzed as one, averaging the rating for the adaptive and non-adaptive version for each user.Consequently, the analysis is concerned whether the ratings on the different scales is positive or negative in comparison to the center of the 5-point scale (see Table 1). The significant results are similar to those with the non averaged ratings (doubled number), anyway.

Table 1. Results for the t-tests on divergence from an average rating.

Subscale	t(df=29)	p-level
ATT	0.05	$p = .959$
PQ	2.70	$p < .05$*
HQI	0.79	$p = .437$
HQS	-1.77	$p < .861$
SH	-1.08	$p = .290$
SA	-2.42	$p < .05$*
PC	4.38	$p < .001$***

For three of the seven scales, there is a significant positive or negative derivation from the center of 3. See Figure 3 for the distribution of ratings (median and quartile). Positively rated are Pragmatic Quality and Perceived Control, whereas Social Acceptance is more negative than the center of the scale. PQ and PC are of course significantly different from SA.

The former two scales represent related constructs, at least based on their descriptions. A strong correlation, actually the highest except for ATT and HQI, strengthens this impression (PQ–PC, ATT–HQI: Pearson's $r = .79$, $p < .001$***).

The other constructs which are assumed to be related, are HQI and SA. But these two do not show similar results and neither the strongest correlation

($r = .63$, $p < .001$***), as both, HQI and SA correlate stronger with ATT ($r = .79$, $p < .001$*** and $r = .74$, $p < .001$***), and SA also with SH ($r = .66$, $p < .001$***).

In summary, the results can be interpret that interacting with this embodied system was quite positive from a usability point of view (PQ, PC), but also quite unpleasant regarding the social situation (SA). The latter scale, however, has to be considered as more important in the frame of this evaluation, as the usage situation is public and the embodiment was explicitly chosen to improve the User Experience. Of course, there is no comparison with a non-embodied version of this SDS, but as a conclusion, this system should be either improved concerning the negative aspects, or even replaced by a different interaction paradigm, e.g. a non-embodied touch-screen.

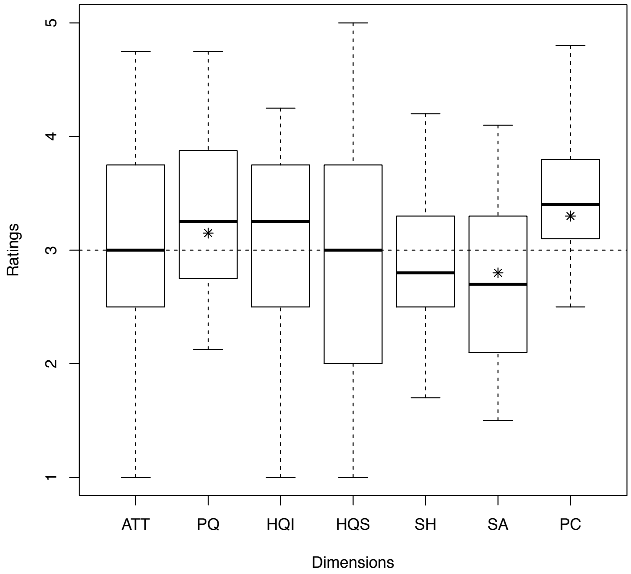

Fig. 3. Questionnaire results for all seven scales. Stars indicate significant divergence from the center (dotted line).

There is only one scale differing for gender: Perceived Control is higher for male users ($F(1, 27) = 4.33$; $p < .0.05$). The related PQ is not significantly different for gender ($F(1, 27) = 1.98$; $p = .17$). As there is no female face tested as well, it cannot be concluded if this result originates from the gender of the ECA or from other sources, e.g., technical affinity. Still, it would be interesting if female users find it especially easy to interact with a male face in this technological domain.

6 Conclusion

The ECA used in a spoken dialog information system was evaluated in a laboratory setting concerning various aspects of User Experience. Results indicate a negative experience concerning Social Acceptance, but a positive experience regarding Pragmatic Quality and Perceived Control. A relationship was found for the last two scales, which are also related in description. The various scales have proven to be useful for summative evaluation of this Embodied Conversational Agent in order to obtain a detailed feedback from users. The two scales with significant negative results have to be taken are more severe than the two positive ones when considering the public interaction situation.

References

1. Hassenzahl, M.: User experience (UX): Towards an experiential perspective on product quality. In: Proc. International Conference of the Association Francophone d'Interaction Homme-Machine, New York, USA, ACM (2008) 11–15
2. ISO 9241-210: Ergonomics of human system interaction – Part 210: Human-centred design for interactive systems (formerly known as 13407). International Organization for Standardization (ISO), Switzerland (2010)
3. Bevan, N.: What is the difference between the purpose of usability and user experience evaluation methods? In: Proc. UXEM'09 Workshop, INTERACT. (2009)
4. Scapin, D., Senach, B., Trousse, B., Pallot, M.: User experience: Buzzword or new paradigm? In: 5th International Conference on Advances in Computer-Human Interactions (ACHI), Valencia. (2012) 336–341
5. Hassenzahl, M., Monk, A.: The inference of perceived usability from beauty. Human-Computer Interaction **25**(3) (2010) 235–260
6. Montero, C., Alexander, J., Marschall, M., Subramanian, S.: Would you do that? – understanding social acceptance of gestual interfaces. In: MobileHCI. (2010)
7. Heerink, M., Kröse, B., Wielinga, B., Evers, V.: Assessing acceptance of assistive social agent technology by older adults: the almere model. International Journal of Social Robotics **2** (2010) 361–375
8. Venkatesh, V.: Determinants of perceives ease of use: Integrating control, intrinsic motivation, and emotion into the Technology Acceptance Model. Information Systems Research **11** (2000) 342–365
9. Venkatesh, V., Morris, M., Davis, G., Davis, F.: User acceptance of information technology: Towards a unified view. MIS Quarterly **27** (2003) 425–478
10. Luerssen, M., Lewis, T.: Head X: Tailorable audiovisual synthesis for ecas. In: Interacting with Intelligent Virtual Characters Workshop (IIVC). (2009)
11. Schröder, M., Trouvain, J.: The German text-to-speech synthesis system mary: A tool for research, development and teaching. International Journal of Speech Technology **6** (2003) 365–377
12. Carnegie Mellon University: Sphinx speech recognition
13. OptimSys s.r.o.: Voice browser
14. Weiss, B., Tönges, R.: Automatic adaption of spoken dialog systems for public and working environments. In: International Conference on Interfaces and Human Computer Interaction (IHCI), Lisbon. (2012) 284–288