Temporal and Spatial Design of Explanations in a Multimodal System

Florian Nothdurft[1](✉), Frank Honold[2], and Wolfgang Minker[1]

[1] Institute of Communications Engineering, Ulm University, Ulm, Germany
{florian.nothdurft,wolfgang.minker}@uni-ulm.de
[2] Institute of Media Informatics, Ulm University, Ulm, Germany
frank.honold@uni-ulm.de
https://www.uni-ulm.de/in/nt/research/ds.html
https://www.uni-ulm.de/in/mi.html

Abstract. Modern dialog systems are known to act user-specific. They apply individual decisions for content presentation and course adaptation. However, it is still an open research question how additional, but required explanations should be integrated best into a given dialog structure. Previous research focused on the improvement of user knowledge models and its fine-grained use in human-computer interaction, but does not directly address the temporal and spatial aspects of presentation when it comes to explanations. In this paper, we introduce different strategies for an ad-hoc integration of required explanations. We describe a user study, and show which parameters from the fields of user experience, personality, and cognitive load theory have what effects on the applied strategies. We expect that our findings can help to increase usability and decrease unwanted cognitive load.

Keywords: Dialog · Adaption · Multimodal · Explanation · HCI

1 Introduction

Modern dialog systems evolve from simple task solvers into intelligent assistants that are able to assist the user in a variety of challenging tasks. These are, for example, *Companion*-Systems, which are "continually available, co-operative, reliable and trustworthy assistants which adapt to a user's capabilities, preferences, requirements, and current needs" [14]. However, because of the increasing capabilities and functionalities of these systems, they also become increasingly complex to operate, and less intelligible for the user. One of the main reasons for this is that the interaction between human and dialog system may exceed the users' knowledge, or capabilities. Hence, such systems should adapt its content and course of interaction to the user's knowledge. One of the most important means of this undertaking are explanations. Explanations can be used, for example, to clarify concepts, provide information on how to perform a task, or justify decision-making. Therefore, they are vital and appropriate instruments

M. Kurosu (Ed.): HCI 2016, Part II, LNCS 9732, pp. 201–210, 2016.
DOI: 10.1007/978-3-319-39516-6_19

for adapting a dialog to the user. Previous research, e. g., in the field of *intelligent tutoring* [1], or *expert systems* [9], focused on the improvement of user knowledge models and its fine-grained use in human-computer interaction (HCI) (e. g., in [4,5,10]).

However, not only the modelling and appropriate selection of knowledge is important, but also how it is presented to the user. If knowledge needs to be imparted, several factors influence how pleasant and effective this will be for the user. Future cooperative *Companion*-Systems behave as interactive peers, which support their users in arbitrary decision making processes of their daily lives. Therefore, here we describe how temporal and spatial distances of providing explanations relative to a selection task in a cooperative decision-making process affect the user experience (UX). We aimed at gathering insights into how different users assess the different variants based on their individual sensation to help to derive layout criteria, select appropriate media types, and structure the dialog in future cooperative *Companion*-Systems. This vision of cooperative systems comes with two implications that are of interest in this paper.

2 Demo System and Scenario

Since the application domain of such systems is not specified but universal, their implementation cannot be realized in an all-embracing manner. Therefore, as the first implication, such systems will act as multimodal interpreters (almost like today's web browsers). They will render the desired user interface (UI) in a model-driven manner. That is why we apply a model-driven prototypical *Companion*-System system [3,7] for our study, which automatically generates a multimodal UI as described in [8].

The aspect of universal application leads to the second implication. Such systems shall be able to provide dynamically-generated explanations whenever they are of need [2,12]. Since complex issues can be explained more convenient with the use of pictures, we use multi-media explanations that consist of text and pictures. Based on that, we focus on the challenging situations, in which an extensive explanation in combination with the underlying selection may exceed the size of the screen. The realization of such an UI would either result in a wizard-like sequence of multiple screens (explanations plus selection) or in one, but scrollabel UI (see Fig. 1), hence varying the spatial and temporal distances between explanation and selection task. As baseline condition we also assessed UX during a selection task without explanations, to compare it to the various conditions.

In this scenario the user's task was to create individual strength training workouts using said prototype. In each strength training workout at least three different muscle groups had to be trained and exercises chosen accordingly. The user was guided through the process by the system, which provided a selection of exercises for training each specific muscle group necessary for the workout. The user was assisted during these selection tasks using the following conditions, which are explained in the following.

3 Methodology

For this evaluation, the temporal and spatial distance of explanations relative to a selection task were varied. The following conditions were implemented (see Fig. 1) and tested:

Joint Explanations beforehand (JE-B) showed collectively all respective explanations temporally prior to the upcoming selection. This means that *one* additional dialog step was created to present *all* explanations at once. As a result of the limited space the explanations were text-only.

Seperate Explanations beforehand (SE-B) showed all respective explanations separately prior to the upcoming selection. Thus, for *every* explanation, a *separate* dialog step was created that presented the explanation, as text and picture, temporally before the upcoming selection.

Joint Explanations during (JE-D) showed collectively *all* explanations plus the related selection within the same dialog, meaning that during the selection, the necessary explanations could be easily accessed and seen by the user. However, as a result of the limited space the explanations were text-only.

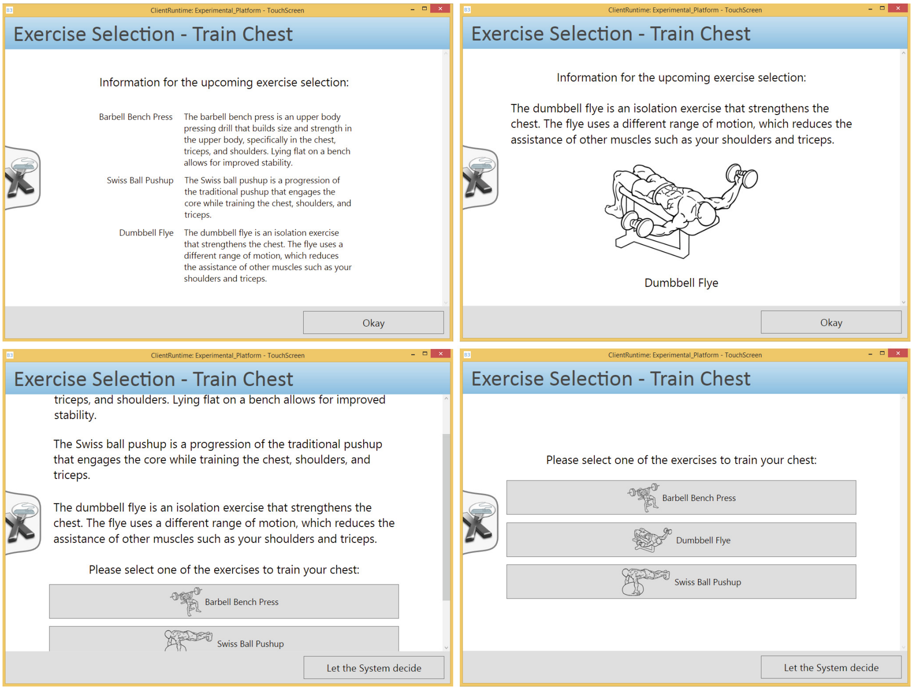
No Explanations (NE) acted as the baseline. In known fashion, the user was confronted with a selection with no prior help by additional explanations. However, the users could still manually request explanations via an additional explicit user interaction. In these cases, the selection dialog was hidden and the requested explanation was shown instead of the selection (as in SE-B). After a user-given confirmation, the former selection dialog was presented; again without any additional explanation.

These four settings allow to vary the temporal distance (i. e., before or during the selection), as well as the spatial distance (i. e., separately, jointly, or only on request) between the explanations and the related selection.

After cleaning the data (e. g., because of incomplete questionnaires) a total of 72 participants were used for the analysis. The participants were distributed through a random-function to the variants, resulting in 18 participants for the baseline condition NE, 28 for JE-D, 13 for SE-B, and another 13 for JE-B.

For measuring UX and other interesting aspects, different standardized and validated questionnaires were used. The *AttrakDiff* questionnaire [6] allows to assess dialog systems or software in general. Its items range from the limited view of usability, representing mostly pragmatic qualities, to the integration of scales measuring hedonic qualities, and the attractiveness in general. *Cognitive load* comprises of three types of load: intrinsic, extraneous, and germane. Therefore, we included an experimental questionnaire developed by Klepsch and Seufert [11] that measures all three types of cognitive loads separately. In addition, the analysis of the *big five personality traits* (Big 5) provides insights in broad dimensions of human personality, using the BFI-K [13].

We expect differences for the AttrakDiff questionnaire in general. The various conditions and the limitation of the content should have some effect on the



ClientRuntime: Experimental_Platform - TouchScreen

Exercise Selection - Train Chest

Information for the upcoming exercise selection:

Barbell Bench Press The barbell bench press is an upper body pressing drill that builds size and strength in the upper body, specifically in the chest, triceps, and shoulders. Lying flat on a bench allows for improved stability.

Swiss Ball Pushup The Swiss ball pushup is a progression of the traditional pushup that engages the core while training the chest, shoulders, and triceps.

Dumbbell Flye The dumbbell flye is an isolation exercise that strengthens the chest. The flye uses a different range of motion, which reduces the assistance of other muscles such as your shoulders and triceps.

Okay

ClientRuntime: Experimental_Platform - TouchScreen

Exercise Selection - Train Chest

Information for the upcoming exercise selection:

The dumbbell flye is an isolation exercise that strengthens the chest. The flye uses a different range of motion, which reduces the assistance of other muscles such as your shoulders and triceps.

Dumbbell Flye

Okay

ClientRuntime: Experimental_Platform - TouchScreen

Exercise Selection - Train Chest

triceps, and shoulders. Lying flat on a bench allows for improved stability.

The Swiss ball pushup is a progression of the traditional pushup that engages the core while training the chest, shoulders, and triceps.

The dumbbell flye is an isolation exercise that strengthens the chest. The flye uses a different range of motion, which reduces the assistance of other muscles such as your shoulders and triceps.

Please select one of the exercises to train your chest:

Barbell Bench Press

Swiss Ball Pushup

Let the System decide

ClientRuntime: Experimental_Platform - TouchScreen

Exercise Selection - Train Chest

Please select one of the exercises to train your chest:

Barbell Bench Press

Dumbbell Flye

Swiss Ball Pushup

Let the System decide

Fig. 1. The conditions. On the *top-left*, the explanations are presented jointly before the selection (JE-B); on the *top-right*, the SE-B condition showed all respective explanations separately prior to the upcoming selection; on the *bottom-left*, the explanations are presented jointly during the selection (JE-D); and on the *bottom-right*, the baseline condition, the user was confronted with a selection with no prior, proactively provided, help by additional explanations.

perceived attractiveness of the system; especially in terms of presented modalities when presenting collective explanations. We expect the joint explanations (JE-B/D) to perform worse than the separate explanations (SE-B) because the SE-B condition might leave sufficient space to also present a picture of the exercise. However, we expect the separate presentation to lead to a higher cognitive load, compared with presenting no proactive explanations at all, because the system behaviour is different than before. Providing explanations during the selection (JE-D) is expected to perform worst because of the sheer amount of content presented in one dialog step, including the limitation that not all content is visible without scrolling.

4 Results

***AttrakDiff*.** The AttrakDiff questionnaire is based on oppositional word pairs (see Fig. 2). There were statistically significant differences between the groups (despite the low number of participants for some conditions) as determined by one-way ANOVA for *ugly—attractive* ($F(3,68) = 5.714, p = .005$), and marginal significance for *unpleasant—pleasant* ($F(3,68) = 4.299, p = .071$).

Post hoc comparisons, using the Fisher LSD test, revealed for *ugly—attractive* that the JE-B condition ($M = 3.15, SD = 1.34$) performed significantly worse ($p = .028$) than providing no explanations (NE) ($M = 4.05, SD = 1.16$), significantly worse ($p = .000$) than SE-B ($M = 4.77, SD = .92$), and also significantly worse ($p = .040$) than providing explanations during the selection screen (JE-D) ($M = 3.92, SD = 1.01$). Providing explanations on separate screens prior to the selection (SE-B) was also rated significantly more attractive ($p = .026$) than explaining during the selection (JE-D), and marginally significant better ($p = .080$) than showing no additional explanations (NE), thus performing best overall. For the word pair *unpleasant—pleasant* the post hoc tests showed that providing separate explanations beforehand ($M = 4.84, SD = 1.21$) was perceived significantly better ($p = .010$) than providing them jointly beforehand (JE-B) ($M = 3.46, SD = 1.45$), as well as marginal significant better ($p = .061$) than providing the explanations jointly during the selection (JE-D) ($M = 4.00, SD = 1.27$).

Additionally, we found that for *cautious—bold*, providing no proactive explanations ($M = 3.611, SD = 1.09$) performed significantly worse ($p = .025$) than providing explanations separately beforehand (SE-B) ($M = 4.38, SD = .96$). For the word pair *discouraging—motivating*, no explanations ($M = 4.00, SD = 1.23$) performed as well significantly worse ($p = .029$) than the SE-B condition ($M = 4.92, SD = .95$).

Cognitive Load. For measured cognitive load, statistically significant results were found between the groups for *general cognitive load* ($F(3,97) = 7.979$, $p = .001$), assessed classically by one item, and for *difficulty* of the task ($F(3,97) = 7.419, p = .004$).

Post-hoc comparisons, to find the origin of significant differences, using Fisher's LSD, showed that the *general cognitive load* was the lowest when providing no explanations (NE) ($M = 1.61, SD = .63$). It was significantly lower ($p = .000$) than the JE-B condition ($M = 2.94, SD = 1.61$), significantly lower ($p = .003$) than the SE-B condition ($M = 2.70, SD = 1.38$), and marginal lower ($p = .072$) than providing the explanations during the selection (JE-D) ($M = 2.16, SD = 1.08$). For the perceived *difficulty* of the task the results are similar. The NE condition ($M = 1.57, SD = .80$) performed best, with a significantly lower perceived difficulty ($p = .002$) than separate explanations (SE-B) ($M = 2.80, SD = 1.39$), significantly lower ($p = .003$) than joint explanations beforehand (JE-B) ($M = 2.73, SD = 1.36$), and significantly lower ($p = .041$) than explaining during the selection (JE-D) ($M = 2.25, SD = 1.38$). However, for *germane load*, which is a good type of load, because it measures that the participants are willing to put effort into creating a schema, presenting no explanations (NE) ($M = 3.14, SD = 1.44$) performed significantly worse ($p = .046$) than joint explanations beforehand (JE-B) ($M = 4.20, SD = 1.38$), and marginal worse ($p = .079$) than separate explanations beforehand(SE-B) ($M = 4.07, SD = 1.49$).

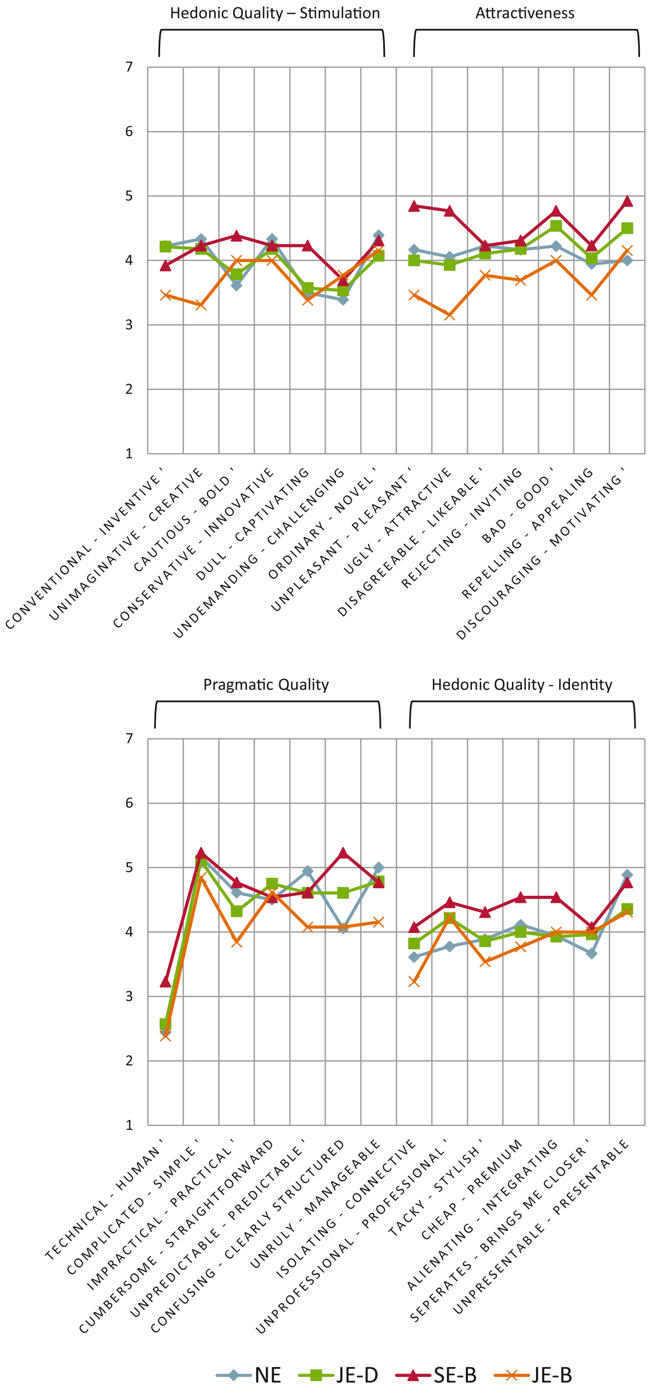


Fig. 2. Average means of the AttrakDiff comparing the explanation conditions on a seven-point Likert scale. The apostrophe (‘) indicates inverted scales, in the interest of readability.

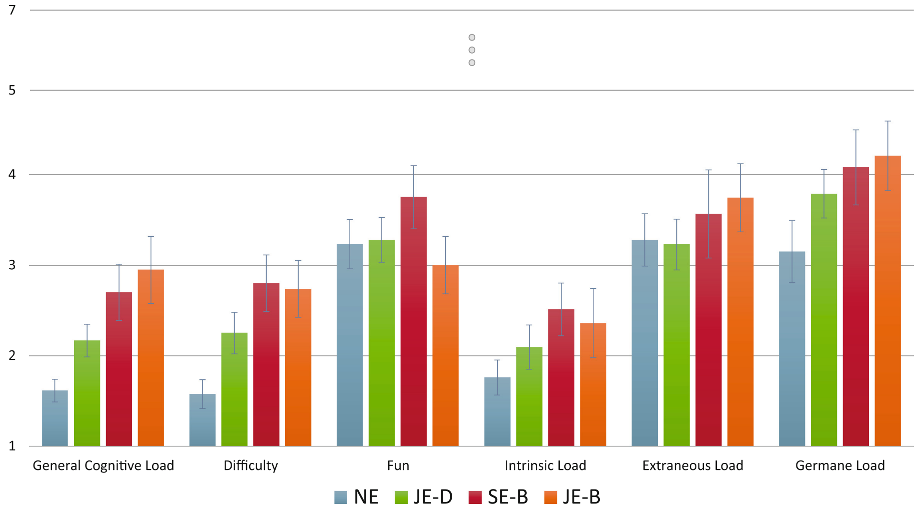


Fig. 3. Average means of the Cognitive Load comparing the explanation conditions on a seven-point Likert scale.

Big 5 Personality Traits. Analysing the Big 5 Personality Traits questionnaire, and its relationship with the other instruments, we found several condition-dependent correlations, using a Pearson correlation test.

For the condition of presenting the explanations jointly during the selection, a significant positive correlation between *Extraversion* and *unpleasant—pleasant* ($r = .400, n = 28, p = .035$) was found. For the Big 5 dimension of *Agreeableness* significant positive correlations with the word pairs *unpresentable—presentable* ($r = .458, n = 28, p = .014$), and *conventional—inventive* ($r = .466, n = 28, p = .012$) were indicated by the data.

For the average means of the participants receiving the explanations separately before the selection, positive correlations between *Extraversion* and the AttrakDiff dimension *Hedonic Quality—Identity* ($r = .699, n = 13, p = .008$), with the dimension of *Hedonic Quality—Stimulation* ($r = .605, n = 13, p = .029$), and with the dimension of *Attractiveness* ($r = .586, n = 13, p = .035$) were found in the data. Besides positive correlations between *Extraversion* and some of the corresponding word pairs of the significant correlating dimensions, also a positive correlation was found with the word pair *technical—human* ($r = .563, n = 13, p = .045$).

For the Big 5 dimension of *Agreeableness* a negative correlation was found with the AttrakDiff dimension *Hedonic Quality—Identity*, $r = -.566, n = 13, p = .044$. This negative correlation originates from the negative correlation in the word pairs *unprofessional—professional* ($r = -.633, n = 13, p = .020$) and *tacky—stylish* ($r = -.645, n = 13, p = .017$). Additionally, also a negative correlation was found with *ugly—attractive*, $r = -.606, n = 13, p = .028$.

For the Big 5 dimension of *Openess to Experience* there were positive correlations with *conventional—inventive* ($r = .582, n = 13, p = .037$), and *conservative—innovative* ($r = .625, n = 13, p = .022$).

When presenting the explanations jointly in one dialog before the selection, there were negative correlations between *Neuroticism* and the AttrakDif dimensions *Hedonic Quality—Identity* ($r = -.742, n = 13, p = .004$) and *Attractiveness* ($r = -.607, n = 13, p = .028$). Additionally, negative correlations were found between *Neuroticism* and the word pairs *unpredictable—predictable* ($r = -.769, n = 13, p = .002$), and *confusing—clearly structured* ($r = -.701, n = 13, p = .008$), both belonging to the *Pragmatic Quality* of the system.

5 Discussion

The data indicates that providing explanations separately before presenting the selection itself (SE-B) performs best. This is perceived particularly more attractive than providing them jointly in advance, but also better than jointly during decision-making. Naturally, this can be attributed to the fact that separate explanations allow for a graphical representation of the exercise, unlike joint explanation conditions. Therefore, we do not think that separating explanations is always the recommended method, but it can be used in cases where the presentation form is limited by the amount of content. The graph in Fig. 2 represents the average mean values of the word pairs. It shows that the condition that presents the explanations separately performs mostly best or as well as the others. However, we think that at least some of these effects can be attributed to the automatically generated layout of the presentation. Comparing the conditions of joint explanations, which only differ in their temporal distance to the selection, and not the modalities, the presentation during the selection performs better. This variation is perceived as more practicable, manageable, connective, or motivating. We think that this can be attributed to the fact that a direct connection between the explanations and decision-making, in the form of the selection, can be made.

The cognitive load (see Fig. 3) is, as expected, the lowest when no additional explanation is provided because this is the system behaviour with which the participants are familiar. However, this also results in the lowest germane load compared with the other conditions. The general cognitive load is modest for the condition that presents explanations during the selection, performing second best. Because most of the other dimensions measure cognitive load (i. e., difficulty, fun, intrinsic load, extraneous load), they are also at least second best (to no explanation); from this perspective, presenting explanations during the selection seems to be the best option when explanations are needed.

In addition, Personality traits seem to have an impact on system perception, as well. We found several correlations between the Big 5 Personality Traits and other measurements. However, it is always important to mention that correlation does not imply causation. When presenting explanations separately in advance, more extraverted persons tend to affiliate higher hedonic qualities and attractiveness to the system compared with not so extraverted individuals, because for all of these dimensions, a significant positive correlation between *Extraversion* and the respective dimensions can be found. Extraverted persons, who are sociable, outgoing, and positive, seem to enjoy this variation with included graphical representations. In addition, for the dimension of *Agreeableness*, which relates on the positive side to more cooperative, trustful, and compassionate people, there is a positive correlation for the dimension of *Hedonic Quality—Identity*. Hence, people with higher agreeableness tend to have a higher perceived identity with the system in this condition, attributed by characteristics such as *connective* or *integrating*. The correlations indicate that presenting the explanations separately in advance seems to be more suitable for extraverted persons.

Contrary to that, we found especially strong negative correlations for the Big 5 dimension of *Neuroticism* in the condition of presenting the explanations jointly in advance. There are negative correlations between *Neuroticism* and *Hedonic Quality—Identity* and *Attractiveness*. Hence, participants with a higher neuroticism score generally perceive the system as less attractive and could identify less with the system. For the other conditions, no correlations to *Neuroticism* could be found, leaving potentially the difference in using graphical representations, and not the joint presentation, as one of the probable reasons for these correlations.

6 Conclusion

The results show that providing explanations separately in advance makes sense when the amount of content would impair the presentation form. However, if a convenient method for presenting the explanation content on the same dialog is possible without impairing the modality choice, for example, this is the best option. We were able to show that both temporal and spatial distances of the presentation of explanations relative to decision-making (i. e., the selection) influence user experience. If these strategies shall be applied in a purely model-driven UI process, additional attributes have to be added, marking these respective items as explanation. Such an attribute can be used to influence the automatic temporal and spatial layout processes, in order to achieve the desired effects. In addition, by analysing the existent correlations, we show that extraverted participants seem to profit from the presentation of graphics, whereas neurotic persons seem to suffer from a low quality of explanation dialogs because for the jointly, only-textual, presentation, neuroticism correlated negatively with the perceived system attractiveness and hedonic system qualities (especially the identification dimension).

Acknowledgement. This work was supported by the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" which is funded by the German Research Foundation (DFG).

References

1. Anderson, J., Boyle, C., Reiser, B.: Intelligent tutoring systems. Science **228**, 456–462 (1985)
2. Behnke, G., Ponomaryov, D., Schiller, M., Bercher, P., Nothdurft, F., Glimm, B., Biundo, S.: Coherence across components in cognitive systems - one ontology to rule them all. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2015). AAAI Press (2015)
3. Bercher, P., Richter, F., Hörnle, T., Geier, T., Höller, D., Behnke, G., Nothdurft, F., Honold, F., Minker, W., Weber, M., Biundo, S.: A planning-based assistance system for setting up a home theater. In: Proceedings of the 29th National Conference on Artificial Intelligence (AAAI 2015), pp. 4264–4265. AAAI Press (2015)
4. Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 3–53. Springer, Heidelberg (2007). http://dl.acm.org/citation.cfm?id=1768197.1768199
5. Fischer, G.: User modeling in human–computer interaction. User Model. User-Adap. Inter. **11**(1–2), 65–86 (2001)
6. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus, G., Ziegler, J. (eds.) Mensch & Computer 2003: Interaktion in Bewegung, pp. 187–196. B. G. Teubner, Stuttgart (2003)
7. Honold, F., Bercher, P., Richter, F., Nothdurft, F., Geier, T., Barth, R., Hoernle, T., Schüssel, F., Reuter, S., Rau, M., Bertrand, G., Seegebarth, B., Kurzok, P., Schattenberg, B., Minker, W., Weber, M., Biundo, S.: Companion-technology: Towards user- and situation-adaptive functionality of technical systems. In: 10th International Conference on Intelligent Environments (IE 2014), pp. 378–381. IEEE (2014)
8. Honold, F., Schüssel, F., Weber, M.: Adaptive probabilistic fission for multimodal systems. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI 2012, pp. 222–231. ACM, New York, 26–30 November 2012
9. Jackson, P.: Introduction to Expert Systems, 2nd edn. Addison-Wesley Longman Publishing Co., Inc., Boston (1990)
10. Jokinen, K., Kanto, K.: User expertise modelling and adaptivity in a speech-based e-mail system. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p. 87. Association for Computational Linguistics (2004)
11. Klepsch, M., F.W., Seufert, T.: Differentiated measurement of cognitive load: Possible or not? (2015, in press)
12. Nothdurft, F., Minker, W.: Using multimodal resources for explanation approaches in technical systems. In: Proceedings of the 8th Conference on International Language Resources and Evaluation (LREC 2012), pp. 411–415 (2012)
13. Rammstedt, B., John, O.P.: Short version of the 'Big Five Inventory' (BFI-K). Diagnostica : Zeitschrift für psychologische Diagnostik und differentielle Psychologie **4**, 195–206 (2005)
14. Wendemuth, A., Biundo, S.: A companion technology for cognitive technical systems. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R., Müller, V.C. (eds.) COST 2012. LNCS, vol. 7403, pp. 89–103. Springer, Heidelberg (2012)