

A comparison of explicit and implicit proactive dialogue strategies for conversational recommendation

Matthias Kraus, Fabian Fischbach, Pascal Jansen, Wolfgang Minker

Angaben zur Veröffentlichung / Publication details:

Kraus, Matthias, Fabian Fischbach, Pascal Jansen, and Wolfgang Minker. 2020. "A comparison of explicit and implicit proactive dialogue strategies for conversational recommendation." In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC), 11-16 May 2020, Marseille, France*, edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, et al., 429–35. Paris: European Language Resources Association (ELRA). <https://aclanthology.org/2020.lrec-1.54>.

Nutzungsbedingungen / Terms of use:

CC BY-NC 4.0



A Comparison of Explicit and Implicit Proactive Dialogue Strategies for Conversational Recommendation

Matthias Kraus, Fabian Fischbach, Pascal Jansen, Wolfgang Minker

Institute of Communications Engineering

Ulm University, Ulm, Germany

{matthias.kraus, fabian.fischbach, pascal.jansen, wolfgang.minker}@uni-ulm.de

Abstract

Recommendation systems aim at facilitating information retrieval for users by taking into account their preferences. Based on previous user behaviour, such a system suggests items or provides information that a user might like or find useful. Nonetheless, how to provide suggestions is still an open question. Depending on the way a recommendation is communicated influences the user's perception of the system. This paper presents an empirical study on the effects of proactive dialogue strategies on user acceptance. Therefore, an explicit strategy based on user preferences provided directly by the user, and an implicit proactive strategy, using autonomously gathered information, are compared. The results show that proactive dialogue systems significantly affect the perception of human-computer interaction. Although no significant differences are found between implicit and explicit strategies, proactivity significantly influences the user experience compared to reactive system behaviour. The study contributes new insights to the human-agent interaction and the voice user interface design. Furthermore, interesting tendencies are discovered that motivate future work.

Keywords: Proactivity; Conversational Recommendation; Intelligent Assistant; Information Retrieval, Spoken Dialogue System

1. Introduction

Due to the recent success of commercial speech assistants (Amazon Alexa, Google Home, Apple Siri, ...), spoken dialogue systems get more and more accepted in our society and find access into our daily lives. For example, information retrieval, such as bus scheduling (Raux et al., 2006) or restaurant information (Young, 2006), is shifted to speech-based applications because speech forms the most intuitive and natural way to transmit information. However, getting access to information via speech is somewhat cumbersome, as people have a limited short-term memory and navigation through speech applications is still inconvenient compared to visual interfaces. Therefore, there exists strong interest in the research community to facilitate the process of information retrieval in dialogue systems by providing adaptiveness and recommendation functionalities. Systems that actively provide recommendations to the user can be labeled as *proactive* systems. According to Nothdurft et al. (2015) "Proactivity in technical systems is an autonomous, anticipatory system-initiated behaviour, with the purpose to act in advance of a future situation, rather than only reacting to it".

Recommendation systems fulfil these criteria by saving the user from information overload, anticipating future desires and supporting them with decision-making (Rafailidis, 2018). There exists a vast amount of literature on traditional and conversational recommendation systems, e.g. see the work of Ricci et al. (2011) and Christakopoulou et al. (2016) for an overview. One of the main research questions regarding these systems is how to elicit the interests and preferences of individuals. Contrary to most speech assistants or Information Retrieval (IR) systems, which typically use an explicit query prompted by the user, recommendation systems exploit the user's data in an implicit manner (Rafailidis, 2018).

In the scope of this work, it was investigated how proactive

suggestions based on these two approaches for gathering user information, affect the user's acceptance and perception of the interaction. Furthermore, it was examined whether proactive suggestions are desirable at all. Therefore, two types of proactive dialogue behaviour were addressed and compared to a baseline reactive condition in a small user study. A university-based restaurant information system was implemented as an Amazon Alexa application ('Skill') capable of providing proactive meal recommendations. Providing suggestions on food and nutrition is popular topic among recommendation system researchers. For example, both Freyne and Berkovsky (2010) and Elswiler and Harvey (2015) describe developments of intelligent food or meal planning applications for a healthy life style. However, their contributions focus on the feasibility of different recommender algorithms in this context, while our work takes a more interaction design approach.

In this paper, it was differentiated between an explicit and an implicit proactive recommendation strategy. The explicit proactive strategy was based on user preferences that had been provided directly by the user, i.e. a user could put favourite dishes or restaurants on a favourites list. Contrary, the implicit proactive strategy used autonomously gathered information while interacting with the user, i.e. a meal or restaurant is put on the favourites list when the user asks for more details about them. The results provide interesting insights in the user's perception of proactive speech-based recommendation systems. This paper is intended to provide a contribution to new insights to the human-agent interaction and the voice user interface design.

The outline of this paper is as follows: Related work regarding conversational recommendation systems and proactive human-computer interaction (HCI) is presented in Section 2. Section 3 deals with the implementation of the dialogue strategies in context of the test scenario and describes the overall system design. Subsequently, the experimental setup and evaluation methods are described in detail.

In Section 5, the results of the study are presented. A discussion of found results is provided in Section 6. Finally, the paper is concluded in Section 7.

2. Related Work

2.1. Conversational Recommendation Systems

Generally, recommendation systems can be defined as personalised decision guides based on user preferences. For example, they can be used to provide assistance in restaurant search (Christakopoulou et al., 2016), online shopping (Linden et al., 2003), or to recommend movies (Dalton et al., 2018) by observing past user behaviour. There are basically two types of recommendation systems: Content-based recommender systems (Pazzani and Billsus, 2007) model the user by characteristics of the items he/she likes or dislikes. Alternatively, systems based on collaborative-filtering (Lu et al., 2015) assist users with making decisions by taking into account the opinions of other people who share similar interests.

Conversational recommendation systems (Sun and Zhang, 2018; Christakopoulou et al., 2018; Christakopoulou et al., 2016; Ikemoto et al., 2019) differ from traditional recommenders, as they include explicit user interaction for crafting their decisions. Christakopoulou et al. (2018) proposed a conversational strategy, in which the user has a more active role in the recommendation process. Here, the strategy was two-folded. First, the system asked the user a set of questions for learning their preferences and thereupon generated topics a user might be interested in. Ikemoto et al. (2019) described a similar strategy combining questions about a user preference and recommendations while refining their recommendation model using immediate user feedback.

Most previous work focuses on technical details, e.g. how to improve the recommendation accuracy (Christakopoulou and Banerjee, 2018; Ikemoto et al., 2019) or which features should be used to generate appropriate recommendation models (Sun and Zhang, 2018). In contrast, a more user-centred view based on natural speech interaction was taken in our approach. Solely speech-based recommendation systems are rare, as vocally delivered suggestions have disadvantages in comparison with visual recommendation (Yang and Lee, 2018). For example, users consume information more slowly and explore less. Hence, there is a need for better navigation mechanisms in speech-based recommendation systems.

In this paper, the design and implementation of an information retrieval system in combination with content-based recommendation functionality using only natural speech interaction is presented. The system was embedded in an industrial context using Amazon Alexa as a speech interface. In doing so, a realistic scenario for testing the system and its underlying interaction design could be created. The goal of our work was not the development of algorithms for modelling appropriate recommendations. Our approach considers a simplistic extraction of user preferences, while focusing primarily on the interaction. Contrary to other work, the acceptance of active suggestions based on manually provided versus automatically collected user data was compared.

This was supposed to give insights on how users perceive proactive recommendations from contemporary speech assistants. Especially since data security in spoken human-machine interaction gets more and more important as humans are cautious to provide personal information to speech-based home assistants (Brasser et al., 2018). For understanding the nature of proactivity in human-machine interaction, a short overview is presented in the following section.

2.2. Proactive Human-Computer Interaction

The term proactivity originates from the domain of occupational and organisational psychology (Grant and Ashford, 2008; Parker et al., 2006). According to its definition, proactive behaviour is about taking control, anticipating and preventing problematic situations instead of only reacting to them. The definition leads to three fundamental elements of proactive systems (Peng et al., 2019): anticipation, action initiation, and target of impact. Anticipation describes the system's awareness of the user's future intention and goals. Therefore, a proactive system needs to be able to recognise and understand the user's verbal, e.g. natural language understanding (He and Young, 2003), as well non-verbal cues, e.g. gaze (Huang et al., 2015a) and body orientation (Huang et al., 2015b).

The manner in which a proactive system should initiate its actions is a frequently discussed topic in HCI and mostly related to the intervention style (e.g. see Wagner et al. (2019)) or a system's level of autonomy (Baraglia et al., 2016; Peng et al., 2019; Zhang et al., 2015). Often the system's autonomy is categorised into 10 levels (Beer et al., 2014; Sheridan and Verplank, 1978; Rau et al., 2013). The levels can be roughly abstracted to low- (computer offers no assistance), medium- (computer offers a suggestion), and high-level (computer executes a suggestion) proactivity. Recommendation systems typically fall under the category of medium-level proactivity, as normally one or a ranked set of suggestions is made. Recent studies showed that different levels of proactivity strongly affect the user's perceived helpfulness and their behaviour in dealing with the system (Rau et al., 2013; Peng et al., 2019).

Therefore, it is essential to study the effects of different types of proactive suggestions for building accepted recommendation systems. In this paper, two proactive - explicit and implicit - are compared to a reactive recommendation strategy. As the intention of the user was obvious before the interaction started (obtain information about current meal options), the proactive actions were triggered after system invocation. The target of impact was to assist the user in finding preferred meals.

3. System Description and Dialogue Design

For creation of a test scenario, we implemented an information retrieval system¹ assisting the user in planning their daily university-restaurant (canteen) visit. Our university has several restaurants, e.g. "Bistro", "Burger Bar", and "Canteen A". Each restaurant offers meals in different food categories, for example "Burger" or "Vegetarian".

¹ Link to service: <https://amzn.to/2O4Pisj>

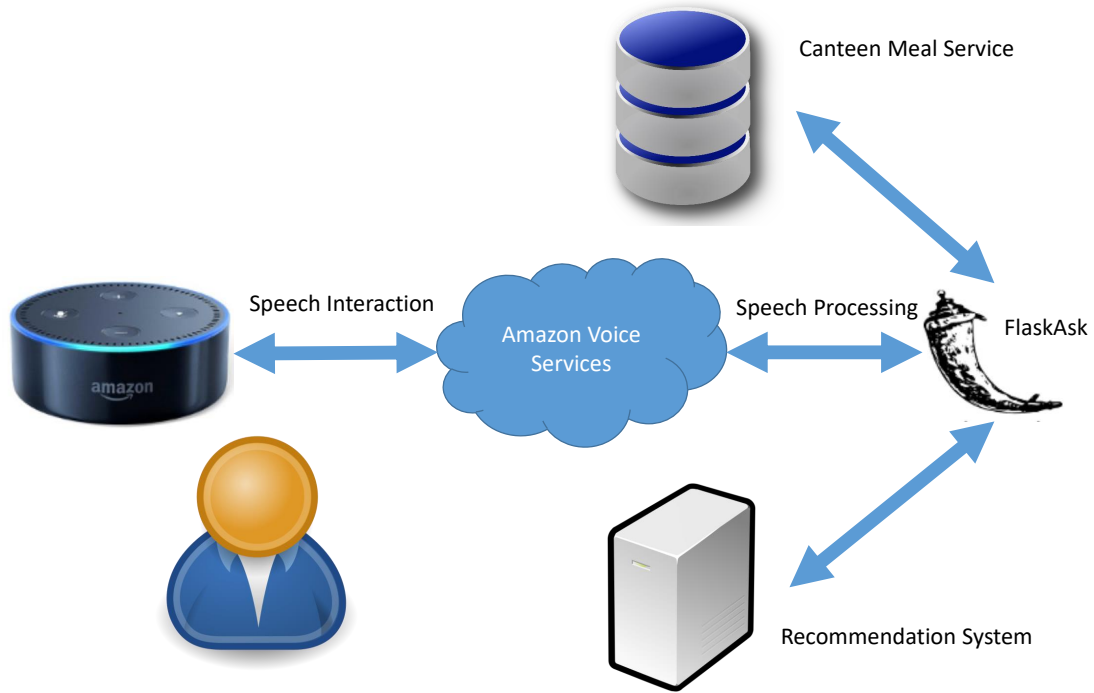


Figure 1: Overview of the combined information retrieval and proactive recommendation system. A user interacted with Amazon’s Alexa speech interface. Amazon Voice Services transformed the spoken words into text. Afterwards, a FlaskAsk-Python application handled the dialogue management, i.e. retrieved meal or restaurant information from the canteen’s database, and selected suggestions from the recommendation system.

The implementation of the system was based on Amazon’s Alexa and its cloud-based speech recognition.

For communicating with the Amazon Voice Services, the Python FlaskAsk ² framework was used. All application data required for the system’s functionalities, in particular for proactive meal suggestions, was stored in a database. This database contained information about six restaurants, their meal plans (ca. 70 different kinds of meals), and meal-specific information like food category or price. A meal plan for one of the restaurants consisted of several food categories, which are repeated over all weekdays. For example, herbal cream soup and potato soup could be assigned to the food-category ‘soup’. While the categories remained the same over the days of the week, the offered dish of a category changed daily. An overview of the system’s architecture is depicted in Figure 1.

Generally, the information retrieval system had two main functions: returning a list of meals and returning detailed information about one single meal. Both existed in several variants. A list of meals was returned by Alexa when users had asked for meals in a specific restaurant, meals from a specific category, when they had searched for meals by name or when they had asked for vegetarian meals. In all cases, users could ask for meal lists of the current day or for another weekday within the next seven days: e.g. ‘What can I eat in the bistro on Tuesday?’; ‘What are tomorrow’s meals in the category pizza?’; ‘Can you find any burger today?’.

The corresponding answer of Alexa either contained only a list of meal names when asking for meals in a specific

restaurant or a list of meal names and the related name of the restaurant in all other cases. After having received a list of meals, users could ask for more detailed information about every meal which was in the list before. More precisely, users could request information about restaurants, food categories, descriptions, and pricing information of meals: e.g. ‘What do you know about vegetable soup?’; ‘How much is pizza salame?’; ‘Where can I find this?’.

For providing proactive system behaviour, two recommendation variants were implemented: *explicit*, and *implicit* proactive assistance. Both extended the described basic functionalities by collecting user preferences and suggesting appropriate meals to users. Hence, users were provided with personalised active assistance with the purpose of retrieving information more convenient and time efficient. The *reactive strategy* did not provide recommendations.

The *explicit* proactive dialogue strategy recommended meals by utilising user preferences that were managed by the user directly. Here, personal *favourites* lists were applied that could be directly manipulated. These lists contained preferred restaurants, food categories, and meals. Users were able to explicitly add or remove content by saying for example, ‘Put a cheeseburger into my favourites.’ or ‘I don’t like this anymore.’. The system then recommended one meal to the user depending on these favourite lists. An example of such system behaviour is presented in Table 1.

Contrary, the *implicit* proactive dialogue strategy recommended meals by utilising automatically-gathered user preferences. For this, the user’s previous behaviour was considered. To illustrate this strategy, consider the examples in Table 1. Every time users searched for a meal, a food

² <https://github.com/johnwheeler/flask-ask>

Common Interaction	Interaction Including Recommendation
U: Alexa, start Mensa-Planner. S: Hi, I am the Mensa-Planner. I assist you finding your favourite meals. How can I help you today? U: What can I eat in the Burger-Bar today? S: The Burger-Bar serves cheeseburger, chickenburger, U: How much is a cheeseburger? S: The price for cheeseburger is around 2 Euro. U: Is a cheeseburger available in a menu? S: Yes, a cheeseburger is available in a menu with french fries. U: Ok, put Cheeseburger on my list. S: I put Cheeseburger on your favourites list.	U: Alexa, start Mensa-Planner. S: Hi, I am the Mensa-Planner. I assist you finding your favourite meals. S: Today, I recommend you a cheeseburger in the Burger-Bar. U: Ah okay, nice. Thank you Mensa Planner. S: See you soon!

Table 1: Examples dialogues representing the common (*reactive*) interaction flow (S: system; U: user). Left: typical dialogue between alexa and user. The green-coloured text represents the functionality of the *explicit* proactive strategy. Right: interaction including a recommendation by the system. The cyan-coloured text represents the typical suggestion utterance for both (*explicit* and *implicit*) proactive strategies. Note that original interactions were conducted in German.

category or a restaurant, the search name was added to a corresponding recommendation list with the value of one. In the presented example, 'What can I eat in the Burger-Bar today?' would have added "Burger-Bar" to list of favoured restaurants. When having asked for detailed information of a meal, its name would have been added to the list of favorite meals. Additionally, its category and restaurant would have been added to respective lists as well. If an entry with the same value already existed, its value would have been increased by one. As the user asked for detailed information about cheeseburgers in our example ('How much is a cheeseburger?' and 'Is a cheeseburger available in a menu?'), the values of "cheeseburger", "Burger-Bar", and "Burger" would have been increased respectively. The system then recommended meals depending on the highest rated entries of the favourites lists, e.g. 'Today, I recommend you a cheeseburger in the Burger-Bar'.

4. Evaluation

In our experiment, the system's acceptance (scale developed by Van Der Laan et al. (1997)) and the user experience with the system were assessed as dependent variables. For measuring the user experience, the well-known Subjective Assessment of Speech System Interfaces (SASSI) questionnaire (Hone and Graham, 2000) was employed. This questionnaire contains the sub-scales response accuracy, likeability, cognitive demand, habitability, as well as user satisfaction. In addition, the subject's motivation to interact with the system was measured using a scale developed by McAuley et al. (1989). All scales were translated into German, as the experiments were conducted using German participants. Additionally, the scales were slightly modified for content and study context. Furthermore, the user experience was assessed with a 7-point Likert scales ranging from 1 = "totally disagree" to 7 = "totally agree" except for the acceptance assessment which used contrary adjective pairs on a 7-point Likert scale.

Developed study design consisted of three independent between-subject conditions to which participants were randomly assigned. Conditions were as follows:

Reactive: subjects used the basic version of the implemented information retrieval system that provided no recommendations at all. This was used as baseline condition.

Proactive (explicit): the system provided recommendations at the beginning of each interaction based on the previously described *explicit* proactive dialogue strategy, where subjects managed their preferences themselves.

Proactive (implicit): the system provided recommendations at the beginning of each interaction based on the previously described *implicit* proactive dialogue strategy relying on automatic measures of previous user behavior.

As a cover story the subjects were told that they would have to test the new Alexa menu assistant of the university for usability. Furthermore, subjects were instructed to interact with the Alexa Skill for about five minutes daily over a period of 5 weekdays and to evaluate the application in a questionnaire at the end of the study. For the duration of the evaluation, they were provided with an Alexa Echo Dot on loan. Additionally, participants received detailed online study instructions. These instructions contained example utterances as well as an explicit task description.

First, subjects had to make meal queries (at least one) choosing from three request types: query of meals offered by specific restaurants, query of meals of a specific food category, or an explicit search for meals. Subsequently, they had to make requests (also at least one) for details about specific meals (either all details or particular details, e.g. food price) or restaurants (description, opening times, location). The user could also request help from the system, if he or she was unsure about what to say. Furthermore, the system was able to prompt the user again in case of low speech recognition accuracy. After participants repeated the task for 5

<i>Dialogue Strategy</i>	<i>Acceptance</i>	<i>SRA</i>	<i>Likeability</i>	<i>Cognitive Demand</i>	<i>Satisfaction</i>	<i>Habitability</i>	<i>Motivation</i>
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Reactive	4.09 (1.26)	3.50 (.88)	3.74 (1.04)	3.79 (1.22)	4.25 (1.04)	3.67 (.98)	5.28 (.61)
Proactive (explicit)	4.79 (1.66)	3.81 (.73)	4.55 (.94)	4.58 (.88)	5.54 (.70)	4.38 (.93)	6.22 (.54)
Proactive (implicit)	4.30 (.84)	4.17 (.77)	4.67 (.63)	4.64 (.64)	5.36 (.93)	5.21 (.80)	6.19 (.74)

Table 2: Descriptive statistics of the measured dependent variables with reference to the the dialogue strategies. Results for cognitive demand are inverted (the higher, the better). SRA implies system response accuracy.

days, they received access to the online questionnaire. Upon successful completion, they were then given the reward for participation.

19 German participants (52.6 % female) with an average age of 23.37 ($SD = 4.06$) were recruited and received 10 Euro in return for their participation. 15 subjects were students, while the other 4 were employees of our university. A condition of participation was actually visiting the canteen on a regular basis.

5. Results

For data analysis, a one-way Analysis of Variance (ANOVA) was used for guaranteeing no significant confounding variables and for testing the significance of described reactive and proactive (explicit, implicit) dialogue strategies. As usual for small sample sizes, a few deviations of normal distribution (Shapiro-Wilk test) were detected. According to Glass et al. (1972) and Field (2009), the F-statistic of an ANOVA can be robust to violations of assumptions, as long as the group sizes are equal. All depending variables showed equal homogeneity of variances according to Levene's test.

In order to rule out confounding group differences for the study conditions, the subjects' experience with existing speech assistants (Alexa, Google Home, Siri,...) were controlled for. There were no significant group differences ($F(2, 16) = 0.001, p >> .05$).

In addition, participants age and gender was similarly distributed in the different experimental groups and no outliers were found in the data set. An overview of the results is presented in Table 2.

The data analysis revealed significant differences between conditions for the dependent variables *Satisfaction* ($F(2, 16) = 3.65, p < .05, \eta^2 = 0.31$), *Habitability* ($F(2, 16) = 4.81, p < .05, \eta^2 = 0.38$), and *Motivation* ($F(2, 16) = 4.25, p < .05, \eta^2 = 0.35$). According to Cohen (1988), the limits for the size of the effect are .01 (small effect), .06 (medium effect) and .14 (large effect). According to these rules of thumb, the effect of our ANOVA is considered as large.

For clarifying which conditions differed significantly, post-hoc t-tests were conducted. For *Satisfaction*, the explicit proactive condition was rated significantly higher than the reactive condition ($t(10) = 2.53, p < .05$). There was also a notable difference between implicit and reactive strategy ($t(11) = 2.03, p = .07$). *Habitability* was rated

significantly higher for the implicit vs. reactive condition ($t(11) = 3.14, p < .01$). Both proactive strategies were rated significantly higher than the reactive condition for *Motivation* (explicit vs. reactive, $t(10) = 3.03, p < .05$; implicit vs. reactive, $t(11) = 2.39, p < .05$). There were no significant differences between the proactive test conditions.

6. Discussion

The study revealed differences between proactive and reactive conditions. Both proactive dialogue strategies received, in case of the explicit strategy significantly, higher ratings for user's satisfaction with system interaction. Hence, it could be shown that by providing a obliging system, the interaction was tailored more appropriate to user needs in spoken information retrieval dialogues. Furthermore, both proactive strategies motivated the user significantly more to interact with the system than the reactive strategy. This is a strong indicator that the system, by its activity, was considered more encouraging. Since motivation was measured after 5 days of usage, proactivity could be seen as a factor for engaging the user more for a permanent use of a recommendation system. However, more studies regarding this topic need to be conducted in order to get more valuable results.

Implicit proactive behaviour was also rated higher on habitability. According to Hone and Graham (2000) a "habitable system may be defined as one in which there is a good match between the user's conceptual model of the system and the actual system.". Consequently, users could build a better model of the system's functionalities, when it implicitly tracked their behaviour for recommendations. This implicates that the tracking of user behaviour worked quite well. However, it seems strange that the explicit strategy in comparison was rated quite lower. Especially, as users were in charge of organising their meal suggestions in this condition. Why this effect did occur needs to be investigated more.

Overall, the proactive strategies were perceived as more user friendly which is additionally backed up by positive tendencies in the measurements of acceptance, cognitive demand, and likeability. This is in line with the results provided by Peng et al. (2019). There, a medium-proactive robot was perceived as more appropriate and helpful in decision-making tasks. However, they used a Wizard-of-Oz setup in contrast to our approach. Hence, our study recommends that

the results can be transferred to real functioning interactive systems. Although there were no significant differences between the two types of proactivity, some observations were possible. The explicit strategy had higher ratings for user acceptance. This could indicate that users want to be in charge when providing information about their preferences and may not fully trust a system that autonomously collects their data. This behaviour can also be seen in the work of Rau et al. (2013), in which a reactive robot had higher trust ratings than a proactive version. However, this may also originate from using Alexa Echo Dots for evaluation. Hence, subjects could have been biased towards the brand. For a better understanding of the reasons, more studies regarding trust in proactive dialogue systems need to be conducted.

Contrary, the implicit strategy was rated higher in system response accuracy and habitability. A cautious assumption can be made, that as more proactive a system gets, the more competent and reliable it is perceived. This only holds true when the system acts in favour of the user and does not give incomprehensible recommendations (Nothdurft et al., 2015). Consequently, the implicit strategy has a higher cost when making errors but also a higher benefit when acting in favour of the user.

Comparing the strengths and weaknesses of the study design, the advantages of the setup formed a realistic test scenario and an easy rapid prototyping of the dialogue strategies. A disadvantage was the quite low number of participants. Using a higher number could have provided more comparable results between the explicit and implicit strategies. Furthermore, the speech recognition errors of Amazon Alexa were troublesome, as several participants of the study reported issues. For example, subject 8 (female, 26) reported that Alexa understood several meals, but some not. In addition, subject 2 (female, 27) described that Alexa often did not understand the meals, when she had asked for details about a dish. We are aware that a Wizard-of-Oz setup would have prevented such errors. However, for the creation of a realistic setup with an experimental duration of 5 days, this was deemed impracticable and not expedient in the context of our evaluation.

7. Conclusion

In this work, a pilot study on different types of proactive system behaviour was presented. Both strategies were implemented in a realistic test scenario using Amazon's Alexa as information retrieval system for planning university-restaurant visits. The study results provided evidence, in line with recent research, that proactive speech assistants can positively effect the perception of HCI. However, it was only concentrated on proactive recommendations in the context of this work. Proactive system behaviour should be far more than only suggestions. Therefore, future work should concentrate on generalisable proactive dialogue strategies for all kinds of application domains.

8. Acknowledgements

This work is done within the technology transfer project "Do it yourself, but not alone: Companion Technology for DIY support" of the Transregional Collaborative Research

Centre SFB/TRR 62 "Companion Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG). The industrial project partner is the Corporate Research Sector of the Robert Bosch GmbH. Additionally, this work receives funding within the project "RobotKoop: Cooperative Interaction Strategies and Goal Negotiations with Learning Autonomous Robots" by the BMBF.

9. Bibliographical References

- Baraglia, J., Cakmak, M., Nagai, Y., Rao, R., and Asada, M. (2016). Initiative in robot assistance during collaborative task execution. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 67–74. IEEE Press.
- Beer, J. M., Fisk, A. D., and Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2):74–99.
- Brasser, F., Frassetto, T., Riedhammer, K., Sadeghi, A.-R., Schneider, T., and Weinert, C. (2018). Voiceguard: Secure and private speech processing. In *Interspeech*, pages 1303–1307.
- Christakopoulou, K. and Banerjee, A. (2018). Learning to interact with users: A collaborative-bandit approach. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 612–620. SIAM.
- Christakopoulou, K., Radlinski, F., and Hofmann, K. (2016). Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 815–824. ACM.
- Christakopoulou, K., Beutel, A., Li, R., Jain, S., and Chi, E. H. (2018). Q&r: A two-stage approach toward interactive recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 139–148. ACM.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Hillsdale (NJ): Lawrence Erlbaum Associates*, pages 18–74.
- Dalton, J., Ajayi, V., and Main, R. (2018). Vote goat: Conversational movie recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1285–1288. ACM.
- Elsweiler, D. and Harvey, M. (2015). Towards automatic meal plan recommendations for balanced nutrition. In *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys '15*, pages 313–316, New York, NY, USA. ACM.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.
- Freyne, J. and Berkovsky, S. (2010). Intelligent food planning: Personalized recipe recommendation. In *Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI '10*, pages 321–324, New York, NY, USA. ACM.
- Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, 42(3):237–288.

- Grant, A. M. and Ashford, S. J. (2008). The dynamics of proactivity at work. *Research in organizational behavior*, 28:3–34.
- He, Y. and Young, S. (2003). A data-driven spoken language understanding system. In 2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721), pages 583–588. IEEE.
- Hone, K. S. and Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Engineering*, 6(3-4):287–303.
- Huang, C.-M., Andrist, S., Saupé, A., and Mutlu, B. (2015a). Using gaze patterns to predict task intent in collaboration. *Frontiers in psychology*, 6:1049.
- Huang, C.-M., Cakmak, M., and Mutlu, B. (2015b). Adaptive coordination strategies for human-robot handovers. In *Robotics: science and systems*. Rome, Italy.
- Ikemoto, Y., Asawavetvutt, V., Kuwabara, K., and Huang, H.-H. (2019). Tuning a conversation strategy for interactive recommendations in a chatbot setting. *Journal of Information and Telecommunication*, 3(2):180–195.
- Linden, G., Smith, B., and York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, (1):76–80.
- Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74:12–32.
- McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58.
- Nothdurft, F., Ultes, S., and Minker, W. (2015). Finding appropriate interaction strategies for proactive dialogue systems-an open quest. In *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication*, August 6-8, 2014, Tartu, Estonia, number 110, pages 73–80. Linköping University Electronic Press.
- Parker, S. K., Williams, H. M., and Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *Journal of applied psychology*, 91(3):636.
- Pazzani, M. J. and Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web*. Springer, pp. 325–341.
- Peng, Z., Kwon, Y., Lu, J., Wu, Z., and Ma, X. (2019). Design and evaluation of service robot’s proactivity in decision-making support process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 98. ACM.
- Rafailidis, D. (2018). The technological gap between virtual assistants and recommendation systems. *arXiv preprint arXiv:1901.00431*.
- Rau, P.-L. P., Li, Y., and Liu, J. (2013). Effects of a social robot’s autonomy and group orientation on human decision-making. *Advances in Human-Computer Interaction*, 2013:11.
- Raux, A., Bohus, D., Langner, B., Black, A. W., and Eskenazi, M. (2006). Doing research on a deployed spoken dialogue system: One year of let’s go! experience. In *Ninth International Conference on Spoken Language Processing*.
- Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, pp. 1–35.
- Sheridan, T. B. and Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Technical report, Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.
- Sun, Y. and Zhang, Y. (2018). Conversational recommender system. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 235–244. ACM.
- Van Der Laan, J. D., Heino, A., and De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5(1):1–10.
- Wagner, N., Kraus, M., Rach, N., and Minker, W. (2019). How to address humans: System barge-in in multi-user hri. In *Proceedings of the 10th International Workshop on Spoken Dialog Systems Technology (IWSDS 2019)*.
- Yang, H. and Lee, H. (2018). Understanding user behavior of virtual personal assistant devices. *Information Systems and e-Business Management*, pages 1–23.
- Young, S. (2006). Using pomdps for dialog management. In *2006 IEEE Spoken Language Technology Workshop*, pages 8–13. IEEE.
- Zhang, Y., Narayanan, V., Chakraborti, T., and Kambhampati, S. (2015). A human factors analysis of proactive support in human-robot teaming. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3586–3593. IEEE.