
Adaptable Utterances in Voice User Interfaces to Increase Learnability

Chelsea Myers

Drexel University
Philadelphia, PA 19104, USA
chel.myers@gmail.com

Anushay Furqan

Drexel University
Philadelphia, PA 19104, USA
anushay.furqan@gmail.com

Jichen Zhu

Drexel University
Philadelphia, PA 19104, USA
jichen.zhu@gmail.com

Abstract

Voice User Interfaces (VUIs) are growing in popularity as a method of controlling smart home features. However, as VUIs grow in popularity, major obstacles still negatively impact their performance and user experience. Since VUIs are invisible by nature, users find it difficult to learn the supported features and verbal commands. Discovering the correct verbal commands is made more difficult since users have a verbiage preference and developers of VUIs cannot pre-program all possible commands. We propose adaptable verbal commands, termed adaptable utterances, and Open User Models (OUMs) as a method to allow customization of a VUI's commands to match the individual user's preference. We review relevant research on adaptive and adaptable VUIs and identify the limitations adaptable utterances and OUMs could address. Finally, we present a sample study design to evaluate our proposed methods.

Author Keywords

Voice User Interface; Learnability; User Experience; Adaptive; Adaptable; Open User Models

ACM Classification Keywords

H.5.2 [User Interfaces]: Voice I/O

Copyright © 2018 for this paper held by its author(s). Copying permitted for private and academic purposes.



Figure 1: Screen shot of *DiscoverCal*

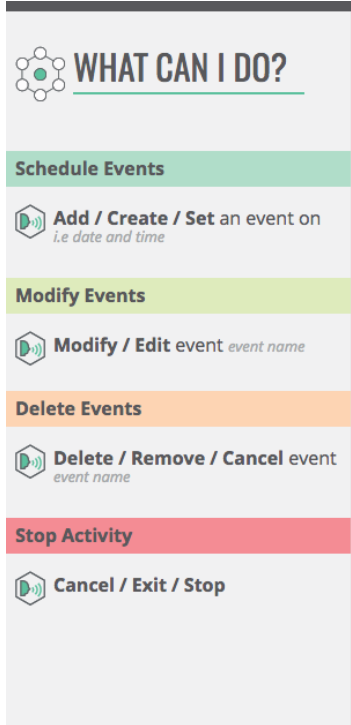


Figure 2: Close up of *DiscoverCal* Menu

Introduction

Voice User Interfaces (VUIs) started becoming more common in everyday objects with the release of iPhone's Siri and the Google Assistant on Google phones. In addition to the VUIs embedded in smart phones, standalone smart objects, such as Amazon's Echo and Google Home, made VUIs more popular because of their accessibility and error prevention [23]. As smart objects and smart homes become increasingly popular, we believe VUIs will be an important interaction method for them. However, the invisible nature of VUIs and their high cognitive load are major obstacles. Users struggle to learn VUI's intents (i.e. features) and the utterances (i.e. verbal commands) that trigger them [22, 20]. This issue can compound if standalone VUIs become a hub on control for smart homes.

In this paper, we propose to explore the use of adaptation and open user models (OUM) to make VUIs more transparent and learnable. OUMs are user models accessible to the user of a system. Ahn [2] defines OUMs as aiming to "transparently reveal the personalization process so that users can easily understand the internal mechanisms and anticipate the behavior of the system." In existing literature, techniques for increasing a VUI's learnability through starter tutorials and in-app tutorials have been explored [7]. *Adaptive* and *adaptable* techniques are often applied to VUIs to support usability and learnability. Although VUI research focuses more on adaptive techniques [14, 6, 16, 17, 10, 7], adaptive VUIs have the potential to confuse their users if the correct mental model of the adaptation is not formed [15, 10].

We propose an adaptable VUI with a GUI OUM that allows user to edit the utterances a VUI accepts. Developers of VUIs cannot pre-program all the possible utterances to meet the preferences of each individual users. We be-

lieve that adaptable utterances will allow users to customize the verbiage of their VUI to match their preferences. We hypothesize that the GUI OUM of the VUI's intents and utterances, coupled with the act of adapting utterances, will increase the transparency of the VUI and its learnability.

In this paper, we present our adaptable utterance and OUM techniques in relation to adaptive and adaptable VUI research. We identify the limitations that adaptable utterances could potentially address. To evaluate the impact of adaptable utterances, we propose a study design for three different adaptable techniques: adaptable tutorials, OUM adaptation, and in-app adaptation. We have designed a calendar-management VUI called *DiscoverCal* (Figures 1 & 2) that has been evaluated in previous studies[10, 20]. *DiscoverCal* runs on a wall-mounted display, with voice control being its primary input. Our study design incorporates these adaptable techniques into *DiscoverCal* and a companion website. The rest of the paper is structured as follows. First, we review related adaptable VUI research and then we present our methodology for this paper's study.

Adaptive and Adaptable VUIs

Research explores different approaches to personalize VUIs for increased learnability and reducing the user's cognitive load. Among these approaches are creating an adaptive or an adaptable VUI system. Adaptive VUI systems are personalized automatically to the user by adapting the system's feedback [15, 19], initiative [16], context [21, 12], and/or visual aids [10, 7]. Adaptable VUI systems are personalized, but only through the user's initiative. Users can change the settings of a VUI to their preference.

Adaptive VUIs

Popular adaptive VUI techniques alter feedback and initiative. Karsenty and Botherel [15] developed a VUI, *TRAV-*

ELS, with a guided and unguided mode that relies on changing the initiative of the system and the amount of feedback provided. Adaptation occurs after users were deemed no longer novices with the system. Karsenty and Botherel found that the switch to the unguided mode increased system errors temporarily. They observed that "several users were confused by the system's change of behavior." [15] Research has found adaptive techniques can increase usability of VUI research [16, 17, 6], but if the correct mental model of the adaptation is not formed by the user, it can be confusing [15, 10]. In a previous study with *DiscoverCal* we explored using a visual adaptive menu to show more complex commands for the system as the user's expertise grew (Figure 2). We found that several users using the adaptive version of *DiscoverCal* were confused by the adaptation and were not sure why the menu changed. This observation lead us to explore other tools to support learnability and transparency for VUIs. Adaptive techniques are a useful tool for VUI design. However, we believe adaptable techniques are under-explored and can address the transparency issues that adaptive techniques pose as seen with *TRAVELS* and *DiscoverCal*.

Adpatable VUIs

Dusan and Flanagan [8, 9] evaluated adaptable intents and utterances with a multimodal VUI drawing application, using a digital touch display with pen and voice input. Participants could create a new intents and utterances for colors, shapes, and other features of the drawing app. Although this study does include adaptable utterances, the application required intensive training to use and memorization of a programmatic language to create intents and utterances. The evaluation also focused on the system's technical performance and not its learnability. To our knowledge, there is no empirical evaluation on how adaptable VUI techniques can be used to allow the user to alter a VUI system's utter-

ances to increase learnability. Adaptable research instead focuses on initiative [18], content words, and self-learning VUIs.

Adaptable Initiative

VUI initiative is determined by who is leading the conversation. If the VUI is asking the user questions, the VUI has the initiative (and vice-versa). Adaptable VUI research so far has empirically evaluated an adaptable and non-adaptable initiative of a VUI and found the adaptable version "outperformed" the non-adaptable [18]. Users of the adaptive version were allowed to swap the initiative of the VUI themselves to navigate through errors. Although this VUI is adaptable, the user only has control over the initiative. We propose to evaluate the affect of adaptable utterances on a VUI's learnability.

Adaptable Content Words

Content words are the entities in an utterance. They are not the keywords that map to an intent, but instead provide information. In the utterance, "Add an event titled Lunch located at the cafe," "Lunch" and "the cafe" are content words. The rest are the keywords that would be detected by a VUI and mapped to their corresponding intent. Adaptable research explores increasing the vocabulary of VUIs by adding unidentified content words. Seneff et al. [21] developed a VUI that automatically detected out-of-vocabulary words (OOV). The user could then choose to add that OOV word to the system by spelling it out. Brill [3] developed an algorithm for NLP that also detected OOV words but would guess their category. The user could confirm this categorization (e.g. labeling "the cafe" as a restaurant). Both of these studies do not evaluate their system with users and instead test with text corpora. We aim to take this adaptable technique further by allowing users to also change the keywords in an utterance that maps to an intent and evaluate

its impact on the learnability.

Self-Learning VUIs

Relevant research allowing users to generate their own utterances includes self-learning VUIs like *ALADIN* [11, 13]. *ALADIN* is used to control smart homes for disabled users and/or users with dysarthria speech. *ALADIN* has no vocabulary, grammar, or feature libraries initially, and instead is completely trained by the user on first use and continually through further usage. However, this type of research has only been evaluated in a demonstration setting. No empirical analysis has been completed on the implications of a self-learning VUI on its users' experience. Our research differs by examining a VUI that does have pre-defined vocabulary, grammar, and feature libraries; mirroring modern VUIs. We aim to analyze how learnability is impacted by allowing users to change the VUI's utterances, while keeping the intents the same.

Open User Models

OUMs have been found to be generally beneficial for adaptive systems [2]. Open Student Models, a category of OUMs, are used in adaptive personalized E-Learning to encourage students to reflect on their progress, strengths, and weaknesses [5, 4]. OUMs provide a layer of transparency for the user to view the information stored about them. Because of the benefits OUMs have been observed to provide, we propose to evaluate the impact of an OUM on a VUI's learnability. Since users of our OUM will be able to edit utterances, our OUM can be categorized as an editable OUM. An editable OUM's adaptive search/filtering system, similar to our adaptable utterances, was initially found to hurt the system's performance [1]. The study was later redone and it was found that the design of the OUM was too complicated and user edits to the OUM negatively affected performance. The editable OUM was redesigned to be more visual, and

researchers found the OUM positively impacted performance [2]. From this, we learn it is important to design a clear OUM where the user is aware of the information presented and how to manipulate it.

Study Design

To analyze adaptable utterances and OUMs, we propose a between-subject study evaluating three adaptable utterance techniques with *DiscoverCal* compared to non-adaptive version focusing on extended learnability. The three adaptable techniques are as follows: adaptable tutorials, OUM adaptation, and in-app adaptation.

Adaptable Tutorials - Tutorials are a common learning tool for VUIs. Our technique creates a tutorial for *DiscoverCal*'s VUI that also includes utterance "training". In this training, users will declare what utterances they wish to associate to each intent. This is similar to *ALADIN*'s self-learning training, but users can only change utterances, not the intents *DiscoverCal* supports.

OUM Adaptation - Allowing users to view the OUM in *DiscoverCal*'s companion website. Here users can update, add, and delete utterances for each intent.

In-app Adaptation - Allowing users to update utterances while using *DiscoverCal* (not in the OUM) to their preference.

We have selected these three techniques to evaluate the impact of different adaptable utterance methods that occur in different parts of a VUI: the tutorial, OUM, and in the application. Recruited participants would be balanced for technical skill and previous VUI experience. Each participant would interact with *DiscoverCal* and perform a select set of tasks. Tasks would have the participants add, edit, and delete events from *DiscoverCal*. To analyze the impact

on extended learnability, the study would be broken into three sessions over the course of one week. Performance metrics (e.g. time spent per task, total time per session, errors encounter) will be collected and a subjective usability completed per session to measure extended learnability.

Conclusion

In this paper we propose the evaluation of adaptable utterances and OUMs as a tool for increasing the learnability and transparency of VUIs. We next present an overview of adaptive and adaptable techniques and their limitations. Although adaptive VUI techniques generally benefit a VUI, the dynamic adaptation can confuse users. We hypothesize that adaptable utterances and OUMs can customize a VUI's vocabulary to the preference of the user while avoiding this confusion. Finally, we propose a sample study design to evaluate these methods.

REFERENCES

1. Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. 2007. Open User Profiles for Adaptive News Systems: Help or Harm?. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 11–20.
2. Jae-wook Ahn, Peter Brusilovsky, and Shuguang Han. 2015. Personalized search: reconsidering the value of open user models. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 202–212.
3. Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics* 21, 4 (1995), 543–565.
4. Susan Bull. 2004. Supporting learning with open learner models. *Planning* 29, 14 (2004), 1.
5. Susan Bull, Paul Brna, and Helen Pain. 1995. Extending the scope of the student model. *User Modeling and User-Adapted Interaction* 5, 1 (1995), 45–65.
6. J. Chu-Carroll. 2000. MIMIC: An adaptive mixed initiative spoken dialogue system for information queries. *Proceedings of the sixth conference on Applied natural language processing* 6 (2000), 97–104.
7. Eric Corbett and Astrid Weber. 2016. What can I say? *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '16* (2016), 72–82.
8. S. Dusan and J. Flanagan. 2002. Adaptive dialog based upon multimodal language acquisition. *Proceedings - 4th IEEE International Conference on Multimodal Interfaces, ICMI 2002* (2002), 135–140.
9. Sorin Dusan and James L Flanagan. 2001. Human language acquisition by computers. In *Proceedings of the International Conference on Robotics, Distance Learning and Intelligent Communication Systems*. 387–392.
10. Anushay Furqan, Chelsea Myers, and Jichen Zhu. 2017. Learnability through Adaptive Discovery Tools in Voice User Interfaces. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (2017), 1617–1623.

11. Jort F. Gemmeke, Bart Ons, Netsanet Tessema, Hugo Van Hamme, Janneke Van De Loo, Guy De Pauw, Walter Daelemans, Jonathan Huyghe, Jan Derboven, Lode Vuegen, Bert Van Den Broeck, Peter Karsmakers, and Bart Vanrumste. 2013. Self-taught assistive vocal interfaces: An overview of the ALADIN project. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2013), 2039–2043.
12. James Glass and Stephanie Seneff. 2003. Flexible and personalizable mixed-initiative dialogue systems. *Proceedings of the HLT-NAACL 2003 workshop on Research directions in dialogue processing - 7* (2003), 19–21.
13. Jonathan Huyghe, Jan Derboven, and Dirk De Grooff. 2014. ALADIN: Demo of a Multimodal Adaptive Voice Interface. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (2014), 1035–1038.
14. Pontus Johansson. 2002. NLP Techniques for Adaptive Dialogue Systems. *Techniques* (2002).
15. Laurent Karsenty and Valerie Botherel. 2005. Transparency strategies to help users handle system errors. *Speech Communication* 45, 3 SPEC. ISS. (2005), 305–324.
16. Diane J Litman and Shimei Pan. 2000. Predicting and adapting to poor speech recognition in a spoken dialogue system. *Proceedings of the National Conference on Artificial Intelligence* (2000), 722–728.
17. Diane J Litman and Shimei Pan. 2002a. Designing and Evaluating an Adaptive Spoken Dialogue System. (2002), 111–137.
18. Diane J Litman and Shimei Pan. 2002b. Designing and Evaluating an Adaptive Spoken Dialogue System. (2002), 111–137.
19. François Mairesse and Marilyn Walker. 2005. Learning to Personalize Spoken Generation for Dialogue Systems. *Proceedings of Interspeech'2005 - Eurospeech: 9th European Conference on Speech Communication and Technology* October (2005), 1881–1884.
20. Chelsea Myers, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. (2018).
21. Stephanie Seneff, Grace Chung, and Chao Wang. 2003. Empowering End Users to Personalize Dialogue Systems through Spoken Interaction 1. *Processing* January (2003), 749–752.
22. Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (2000), 63–65.
23. Kathryn Whitenton. 2016. Voice Interaction UX: Brave New World...Same Old Story. (2016). <https://www.nngroup.com/articles/voice-interaction-ux/>