ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/324670825

Patterns for How Users Overcome Obstacles in Voice User Interfaces

Conference Paper · April 2018

DOI: 10.1145/3173574.3173580

CITATIONS
65

READS
780

5 authors, including:

Chelsea M. Myers
Drexel University

9 PUBLICATIONS **101** CITATIONS

SEE PROFILE

Jichen Zhu
Drexel University

79 PUBLICATIONS **387** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Adaptive Techniques for Voice User Interfaces (DiscoverCal) View project

Project Parallel View project

All content following this page was uploaded by Chelsea M. Myers on 10 January 2019.

The user has requested enhancement of the downloaded file.

Patterns for How Users Overcome Obstacles in Voice User Interfaces

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

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

Jessica Nebolsky
Drexel University
Philadelphia, PA 19104, USA
jjn63@drexel.edu

Karina Caro
Drexel University
Philadelphia, PA 19104, USA
karinacaro@drexel.edu

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

ABSTRACT
Voice User Interfaces (VUIs) are growing in popularity. However, even the most current VUIs regularly cause frustration for their users. Very few studies exist on what people do to overcome VUI problems they encounter, or how VUIs can be designed to aid people when these problems occur. In this paper, we analyze empirical data on how users (n=12) interact with our VUI calendar system, *DiscoverCal*, over three sessions. In particular, we identify the main obstacle categories and types of tactics our participants employ to overcome them. We analyzed the patterns of how different tactics are used in each obstacle category. We found that while NLP Error obstacles occurred the most, other obstacles are more likely to frustrate or confuse the user. We also found patterns that suggest participants were more likely to employ a "guessing" approach rather than rely on visual aids or knowledge recall.

ACM Classification Keywords
H.5.2 User Interfaces: Voice I/O

Author Keywords
Voice control; Voice User Interfaces; User Experience

INTRODUCTION
Voice User Interfaces (VUIs) are rapidly being integrated into our daily lives. Major companies such as Amazon and Google are investing further into Automatic Speech Recognition (ASR) technologies [1] which power VUIs. As a result, we have VUI products like Amazon's Alexa-enabled devices, Google Home, and Apple's HomePod being embedded into people's homes and becoming a part of their daily routines. Similar to Graphical User Interfaces (GUIs), VUIs are becoming a common way to interact with computers.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

© 2018 ACM. ISBN 978-1-4503-5620-6/18/04. . . $15.00

DOI: https://doi.org/10.1145/3173574.3173580

However, today's VUIs face many unresolved challenges. In addition to the error rates of natural language processing (NLP), users are struggling to make correct mental models [8, 3, 14]. For example, Shneiderman argued that "Speech is slow for presenting information, is transient and therefore difficult to review or edit, and interferes significantly with other cognitive tasks" [14]. Others have observed that the invisible nature of VUIs, compounded by the imperfection of NLP technology, makes them difficult to learn and use [2, 3, 8].

To improve the learnability and usability of VUIs, existing research has focused on new interaction techniques [17, 9, 2] and better algorithms [12] to increase speed and reduce error rates. Meanwhile, we need to understand how users interact with current VUIs, what challenges they face, and how they try to resolve them. Currently, we have limited understanding of these questions, with a few exceptions [8, 10, 6, 5]. The paper extends this line of work by examining *empirical usage data* (compared to self-reported data [8, 10]) on a new VUI system with modern NLP technology (compared to early generation VUI systems [6, 5]).

In particular, we present results from a user study of *DiscoverCal*, our VUI calendar manager with adaptive discovery tools (ADT) [3]. We analyzed how 12 participants, with self-reported technical backgrounds, interacted with *DiscoverCal*; each participant used the system for three sessions over 7 days. Our main research questions are: 1) what are the categories of obstacles the users encountered, and 2) what tactics they developed to overcome these obstacles? Based on our analysis, we present 4 obstacle categories, 10 tactic categories, and the patterns that exist between them. By analyzing these patterns, we found that NLP Error obstacles are the most common, but other obstacles hindered our participants more. Also, we found our participants relied more on guessing to overcome obstacles than knowledge recall or visual aids.

The rest of the paper is structured as follows. First, we review related VUI user experience (UX) research. Then we describe our previous study using *DiscoverCal* and our methodology for this paper's study. Finally, we present our results and discuss their design implications for future VUIs.

RELATED WORK

Many existing works on VUIs focus on improving its learnability and user satisfaction. For instance, Purlington et al. [10] analyze VUI sentiment through Amazon customer reviews of Alexa's Echo to uncover characteristics that lead to higher user satisfaction. Research shows that users significantly prefer the personification of VUIs and that this anthropomorphism can improve the user's experience [11]. Other UX-focused research tests how to support VUI learnability through adaptive visuals and commands [2, 4]. Researchers also focus on the VUI's medium; for instance, designing for smaller mobile displays using adaptive visuals to support a VUI system's discoverability [17].

Regarding how users interact with VUIs, research found that there are "important differences" between how human users speak to computer systems and other humans [13]. For example, communication with a computer system is brief. Participants also experienced difficulty with the retention of information and commands when using a VUI [6]. Comparing VUIs to GUIs, users reported that "it is hard to talk and think than type and think." [5] This research highlights the importance of understanding how users interact with VUIs, showing that the conventions of human-human and human-GUI interaction do not always apply.

Relatively little work has been done on close empirical analysis of user interactions with VUIs. Luger and Sellen [8] interviewed "regular" users of VUIs to analyze what fosters and deters everyday use. They found differences in technical knowledge influenced participants' initial expectations. Purington et al.'s [10] sentiment analysis of online reviews showed that the people who personified Alexa had higher satisfaction with the VUI system, but most reviewers referred to it with object pronouns. In comparison, our work also strives to understand how VUIs are used, but our approach focuses on directly analyzing empirical data to do so.

Similar studies have been conducted with previous generations of VUIs in the 1990s [6, 5]. Schneiderman elaborates on these in his well-known paper on the limitations of VUIs [14]. These studies are examples of the obstacles VUIs pose. Karat et al. [5] categorized different types of dictation errors and techniques to correct them. We extend and re-evaluate their work by studying modern VUIs with more sophisticated NLP and interaction methods.

Several studies [15, 7] examined the role of hyperarticulation, a tactic users of VUIs often deploy to resolve a problem "by speaking more slowly, more loudly, and more clearly." [15] A command in VUI is broken into two parts, its *intent* (what the user wants to do) and its *utterance* (how the user says it). Stent et al. found participants more frequently hyperarticulate content words in an utterance after an error rather than the keywords [15]. For instance, the utterance "Add an event called Morning Meeting," "Add an event" are keywords because they declare this utterance maps to the Create Event intent, while "Morning Meeting" are content words, providing more information about the event being created. In our research, we look at a wide variety of tactics, including hyperarticulation, employed by users with technical backgrounds when they interact with unfamiliar VUIs.

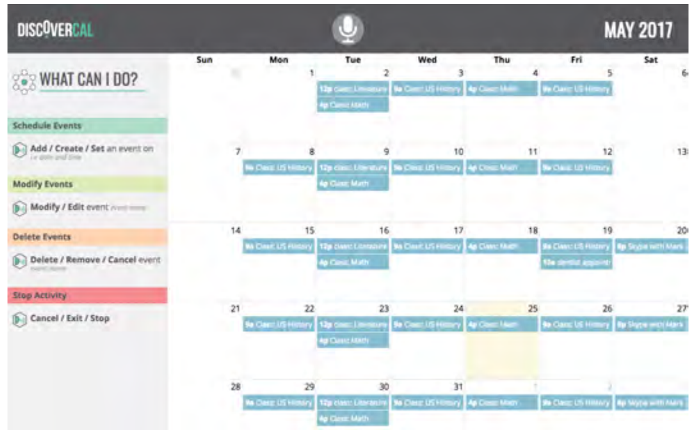

Figure 1. Screen shot of *DiscoverCal* with ADT

METHODOLOGY

We analyze data from *DiscoverCal*, a VUI calendar manager (Figure 1). The primary method of interaction is voice, using Dialogflow[1] (a machine learning NLP library) for processing voice commands. A GUI was designed for *DiscoverCal* to display the calendar and provide information on possible voice intents and utterances.

DiscoverCal uses adaptive discovery tools (ADT) to help users learn supported utterances and intents and discover new ones. ADTs adapt to the system's context and provide users with contextually relevant information to aid in interactions. *DiscoverCal's* menu adapts based on the participant's successful usage of intents and shows more complex intents progressively for the system. More details of the design of *DiscoverCal* can be found in our prior work [3].

To test ADT's impact on extended learnability, we conducted a between-subject study (n=26) balanced for gender and previous VUI usage. Our experimental group used *DiscoverCal* with its ADT menu, and the control group used a version of *DiscoverCal* with a static menu. For both groups, each participant carried out pre-written tasks with DiscoverCal in three 30 minute sessions over the length of one week. All participants had the same tasks; with 3-4 tasks per session. In each session, they were asked to complete the same set of tasks including: creating events, modifying events, deleting events, and inviting attendees to events. We used a lab with one facilitator handing the participants their tasks and recorded each session with both a stationary camera and screen recordings. Once a task was done, the participant was handed the prompt for the next task. Each session was followed by a semi-structured interview that was also recorded. *DiscoverCal* also captured every utterance as it was processed by Dialogflow, with the intent it mapped to, and *DiscoverCal's* response. With this data, we were able to have two transcripts for each session: what the participant said, and what *DiscoverCal* processed it as. We used the Think Aloud protocol to gather data on users' reactions. For it not to interfere with the voice commands, we

[1]https://dialogflow.com/

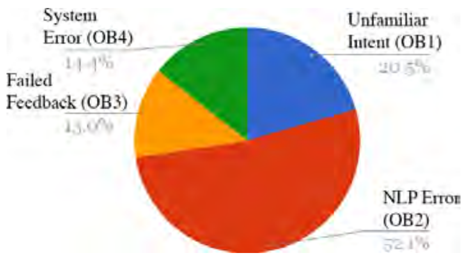

Figure 2. % of Total Obstacles Recorded (146)

designed *DiscoverCal* to only process voice input when the user presses a button, indicating that they are speaking to the system.

In this paper, we focus specifically on analyzing what categories of obstacles users encountered and the tactics users developed to work around them. We randomly selected 12 participants, controlled for gender, experimental and control groups (6 from experimental and 6 from control group), and previous VUI usage, for in-depth analysis. In a post-test questionnaire, each participant self-identified as having a technical background. Participants were ages 18-44 years old (6 male, 6 female). These participants were recruited from the departments of Digital Media and Computer Science majors from a university located in a major U.S. city.

We followed a systematic procedure to code 36 transcripts, three transcripts per participant, as follows. In the beginning, one researcher generated an initial code scheme performing the first iteration of open and axial coding [16] on each transcript while watching its corresponding video recording. The code scheme was presented and discussed with three other researchers. Once the codes included in the code scheme (i.e., obstacles and tactics) was agreed upon by this group, coding was performed by two researchers using the iterated code scheme. Multiple coding for the tactics was used because participants could use multiple tactics per utterance. With this data, we generated transition tables by calculating the probability of transitioning from using one tactic to using another to uncover tactic patterns.

RESULTS

In this section, we present our results regarding 1) what are the common obstacle categories users of *DiscoverCal* encountered, and 2) what are the categories of tactics applied accordingly.

Obstacle Categories

We identified four main categories of obstacles in our data. A total of 146 obstacle occurrences were recorded. The percentage breakdown per obstacle category of that total is summarized in Figure 2. As expected, the most encountered obstacle was NLP Error obstacles; making up 52.1% of the total obstacles counted. Unfamiliar Intent obstacles were the second most common while Failed Feedback and System Error showed similar results.

OB1: Unfamiliar Intent

When a participant is unfamiliar with a VUI system, this can manifest itself in two ways: 1) The participant says something (an utterance) that is structured in a way the NLP cannot parse, or 2) the participant tries to execute an intent that the VUI system does not support. This obstacle is a reflection of the participant's mental model of the VUI system. Encountering these errors shows the participant is not fully aware of what the VUI system can do and/or how to structure utterances to achieve desired intents. For example, Participant 1 first said, *"Schedule study session,"* to create an event which is not supported by *DiscoverCal*. "Add event study session" would be supported though.

OB2: NLP Error

NLP Error obstacles occur when the NLP "misheard" the participant and/or maps their utterance to the wrong intent. We only categorized NLP errors that were perceived by the participant as an NLP Error obstacle. NLP errors were very common, but sometimes they did not interfere with the participant achieving their goal. As the most common obstacle, the participants commented on this frequently in their follow-up interviews. Participant 12 commented, *"...it's a great new, blooming technology [voice control], but it still has it's little kinks like, 'N,' and 'M,' sounds are a little weird for it."* Even with these "kinks," we coded that half our participants (6 out of 12) expressed faith in NLP to advance, making these obstacles less of an issue in future VUIs.

OB3: Failed Feedback

DiscoverCal is a VUI with both visual and audio feedback. At times, participants were observed to have ignored or misinterpreted this feedback which caused future errors. These instances were categorized as Failed Feedback obstacles since the feedback was not clear enough or did not provide the participant with the information they needed to achieve their goal. These obstacles reflect the VUI's design, and show when feedback design needs to be further improved.

For example, Participant 4 created an event with a long title and later could not edit it. This is a failure of *visual* feedback. The participant struggled to find her event on the calendar, and remember the event title to say to the GUI. She commented on this, asking *"How do I like, figure out, the name, of the group? If I want to change it?"*

For *audio* feedback, an obstacle can occur when VUI's feedback is ambiguous. A limitation of *DiscoverCal* is that the date and time must be given at *at the same time.* This was not always clear to the participants that they must say both the date and time in one utterance. *DiscoverCal's* feedback in this instance is not clear to all users and can confuse them; causing a Failed Feedback obstacle.

OB4: System Error

A System Error obstacle comes from a flaw in the VUI system's architecture. For example, a known bug was participants sometimes needed to execute the Cancel intent twice to cancel an event properly. We believe this issue came from Dialogflow, since participants' first attempt mapped to an unknown Cancel intent we did not make or design.

Tactic Categories

Based on the usage data, we identified ten main categories of tactics. Their breakdown is summarized in Figure 3.

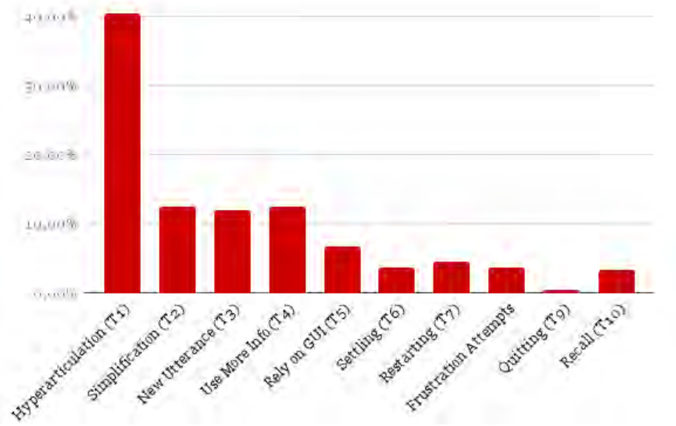

Figure 3. % of Total Tactics Recorded (457)

T1: Hyperarticulation. The participant speaks louder and/or slower; trying to clearly articulate her utterance.

T2: Simplification. A participant reduces the complexity of her utterance, normally removing content words and leaving keywords.

T3: New Utterance. A participant tries a different utterance to achieve the same intent, by using one that is supported and specified in the menu, or inventing their own. For instance, Participant 11 experienced an NLP Error obstacle when his utterance, *"Modify math group meeting"* was processed as "Modify *mask* group meeting." Participant 10 next said *"Edit"* instead of "Modify" because he was not sure what went wrong the first time, and tried out a different utterance.

T4: Use More Info. A participant adds more information to an utterance, believing it will help the system understand what they want to achieve. When Participant 8 experienced an NLP Error obstacle while trying to edit an event, he then said *"Modify group meeting* ***on Wednesday.****"* By saying, "on Wednesday," he hoped the VUI system would have more information to retrieve the event he was trying to edit correctly. This tactic suggests the participant overestimates the VUI's intelligence.

T5: Rely on GUI. Participant looks at *DiscoverCal*'s GUI for direction. For example, Participant 1 thinks aloud while trying to figure out how to create an event and looks at the GUI's menu. *"So on the side is says, 'Add an event on.' So I am going to say that..."* Since we had video footage of the participants using *DiscoverCal*, we could observe when participants look at its menu for help. Participants would also note this aloud during testing.

T6: Settling. Participant settles and accepts a flaw in creating or modifying an event. For Participant 12, the NLP misheard "Ethics" as "FX" when he was trying to create a study group meeting for an Ethics course. After several attempts to correct the title, Participant 12 settled, and accepted the incorrect title.

T7: Restarting. If a participant cannot overcome an obstacle using the previous tactics, the participant may cancel or delete the event and restart. For example, Participant 6 was not sure why she could not change the end time for an event when encountering a Failed Feedback obstacle and said, *"Not sure...It keeps asking for the start time. But it never asks for when it ends...So maybe I should start over. Ok, I am going to start over and try again."*

T8: Frustration Attempts. A participant rapidly tries many different tactics in a burst of frustration. There is not much thought put in each utterance. Even though this tactic is a combination of others, we decided to give it its own category because unlike the other tactics, this is not as deliberate. Instead, the participant is trying anything and everything to overcome the obstacle.

T9: Quitting. The participant quits the task and does not overcome the obstacle at all. While this is not a tactic to overcome obstacles, it is an important category to record. It shows when a participant no longer wishes to continue with a task and gives up.

T10: Recall. Through the Think Aloud process, we observed participants recall the correct utterance without aid of the GUI while facing an obstacle. This shows the participant is learning the VUI and pauses to reflect on that knowledge to overcome an obstacle.

For tactics, Hyperarticulation (T1) was used the most, as past research has pointed out [7, 15, 8]; totaling 40.48% of the total tactics used. We found that tactics 1-4, Hyperarticulation (T1), Use More Info (T4), Simplification (T2), and New Utterance (T3), were the most used, totaling 77.46% of the totals tactics employed.

Obstacle and Tactic Patterns

The breakdown of what percentage of tactics used for each obstacle can be seen in Figure 5. Hyperarticulation (T1) was the most used tactic for each obstacle but was employed the most for NLP Error obstacles. Tactics we expected would occur often were surprisingly low, with Rely on GUI (6.78%) and Recall (3.28%).

Fallback Tactics

We observed that tactics 6-9 (Settling, Restarting, Frustration Attempts, and Quitting) were fallback tactics that were commonly coupled with a participant's confusion and frustration. The count of these fallback tactics were relatively low, making up 12.47% of total tactics recorded. In Figure 5, we see Settling (T6) used the most for Unfamiliar Intent (OB1) and System Error (OB4) obstacles. The Restarting tactic (T7) was used more for Failed Feedback (OB3) obstacles compared to the other tactics employed, with Unfamiliar Intent (OB1) and System Error (OB4) obstacles getting lower but similar results. The Frustration Attempts (T8) tactic made up a similar percentage of total tactics for Unfamiliar Intent (OB1) and NLP Error (OB2) obstacles. Quitting (T9) was only recorded once for a Failed Feedback (OB3) and once for a System Error (OB4) obstacle. Since we see these fallback tactics used more for Unfamiliar Intent (OB1), Failed Feedback (OB3), and System Error (OB4) obstacles, this could suggest participants experience a higher level of frustration and confusion when encountering these obstacles.

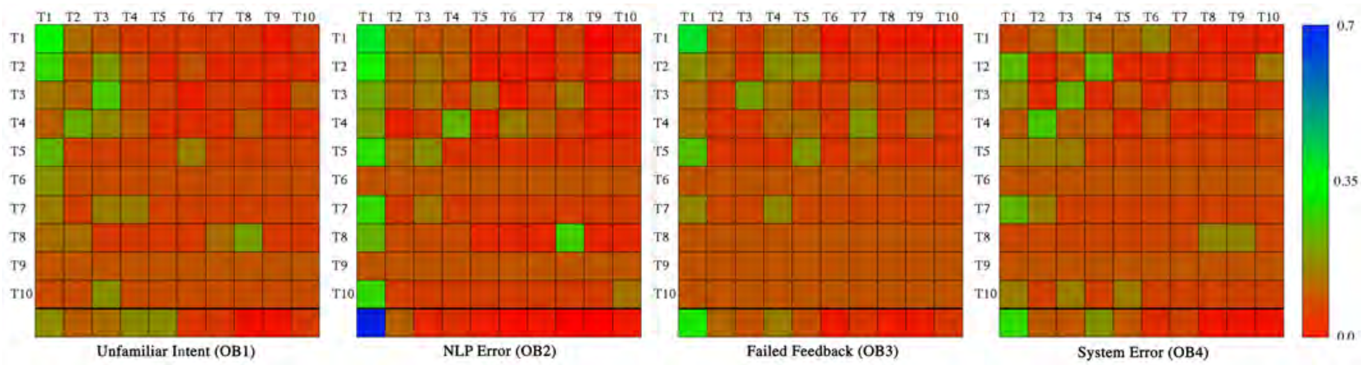

Figure 4. Heat maps representing the probability of transitioning from one tactic (row) to another (column) for our obstacle categories

Simplification (T2) v. Use More Info (T4)

Existing literature suggests the Hyperarticulation (T1) and Simplification (T2) tactics should be the most used [15, 8]. Our data shows that the Use More Info (T4) tactic was used almost as much Simplification (T2), its opposite tactic (Figure 3). Participants would provide more information when facing an obstacle in an effort to "help" the VUI process their intent correctly. The Use More Info (T4) tactic was often seen when the participants overestimated the VUI's intelligence. We speculate that the rise in Use More Info (T4) tactic could mean the VUI's intelligence is not being properly conveyed.

	OB1	OB2	OB3	OB4
Hyperarticulation T1 (%)	24.74	52.70	40.00	24.66
Simplification T2 (%)	14.43	11.71	9.23	15.07
New Utterance T3 (%)	18.56	9.46	7.69	15.07
Use More Info T4 (%)	12.37	9.46	18.46	16.44
Rely on GUI T5 (%)	9.28	3.60	12.31	8.22
Settling T6 (%)	5.15	3.15	0.00	6.85
Restarting T7 (%)	5.15	2.70	9.23	5.48
Frustration Attempts T8 (%)	5.15	4.50	0.00	2.74
Quitting T9 (%)	0.00	0.00	1.54	1.37
Recall T10 (%)	5.15	2.70	1.54	4.11

Figure 5. Percentage of each tactic used for each obstacle

Tactic Transition Patterns

To analyze common patterns in the sequential deployment of the different tactics for each category of obstacle, we calculated the probability of each tactic being deployed after another tactic. To avoid the combinatorial explosion of 10 tactics, we simplified the multiple coding in the following way: When an utterance using T1 T2 is followed by another utterances with T3, we considered as if T1 is followed by T3 and also T2 followed by T3. The converse is true. This way, we can more easily spot trends in the transition. The limitation is that our graph does not show the common types of tactic combinations. We illustrate our findings with heat maps, seen in Figure 4, with a scale of 0-70% indicating the probability of a transition.

Since we identified 10 different tactics, this resulted in a 10 by 10 table of probabilities. The probability in row i and column j represents the probability of a participant deploying tactic j after having just deployed tactic i. In other words, Ti (row number) is used before Tj (column number). We calculated an additional 11th row for each table showing the probability for each tactic being the first one deployed for an obstacle.

Given that the number of samples available to estimate these probabilities is small, we employed Laplace smoothing. Figure 4 shows a visualization of the resulting probability tables for each of the 4 obstacles, where we have shaded the cells with higher probability green/blue and those with lower probability red. As can be seen, several interesting patterns emerge.

Hyperarticulation (T1) is a tool used across obstacles. For NLP Error (OB2) obstacles, the Hyperarticulation (T1) tactic has a high probability of being used first and again after other tactics, except for Use More Info (T4), Frustration Attempts (T8), and Quitting (T9). This reflects our observations during the sessions of Hyperarticulation (T1) being combined with other tactics and occurring more often if the participant knew an NLP error happened.

Frustration leads to more frustration. For NLP Error (OB2) obstacles, we see the pattern of Frustration Attempts (T8) followed by itself. We see this again, but with a lower probability for Unfamiliar Intents (OB1) and System Error (OB4) obstacles. This could indicate that the Frustration Attempts (T8) tactic "episodes" occur for these obstacles.

Different initial approaches for obstacles Unfamiliar Intent (OB1) and System Error (OB4). We can also see that for obstacles NLP Error (OB2) and Failed Feedback (OB3), participants tend to always start with Hyperarticulation (T1). This is particularly notable for NLP Error (OB2) obstacles, where the probability that a participant starts with tactic T1 is 0.658. Whereas for obstacles Unfamiliar Intent (OB1) and System Error (OB4) other tactics such as Use More Info (T4) and Rely on GUI (T5) are common for Unfamiliar Intent (OB1) obstacles and Use More Info (T4) is also common for System Error (OB4) obstacles. This lower probability could indicate that participants have more diverse approaches to these obstacles and use Hyperarticulation (T1) less.

Participants exploring utterances for Unfamiliar Intents (OB1) and System Error (OB4) obstacles. When looking at System Error obstacles (OB4), we see that the common patterns of Use More Info (T4) followed by Simplification (T2), and

New Utterance (T3) followed by New Utterance (T3) (that we observed in Unfamiliar Intents (OB1)) again appear. But we also see some new common sequences such as Simplification (T2) followed by Use More Info (T4) and the inverse. Tactics 2-4 are exploratory tactics where the participant is creating new utterances. This could indicate participants took an exploratory/guessing style approach to overcoming obstacles for these tactics.

Participant's Background
Since all of our participants self-reported they had technical backgrounds, we could not analyze the impact of technical proficiency. We performed significance tests comparing gender, previous VUI usage (e.g. Siri and/or Alexa), and control versus experimental groups (*DiscoverCal* with and without ADTs). We found these characteristics did not have significant impact on the obstacles they encountered or the tactics employed. We did find that among our participants, two with a non-English accent did encounter more NLP Error obstacles on average, but remained consistent with the rest of the participants for the other obstacles.

DISCUSSION
Our goal was to understand better what obstacles people face when using unfamiliar VUIs and how they overcome them. We believe our research indicates that although NLP Error obstacles are the most common, they are not the biggest threat to VUI UX. Also, based on our findings we suggest our participants approached obstacles by guessing more than relying on GUI support.

The Implications of VUI Obstacles.
There is a sentiment among some developers that the advancement of NLP will automatically resolve many existing problems of VUIs. This type of advancement is crucial, as the data shows that NLP Errors make up 52.1% of the obstacles participants encountered in our sample. We believe that although NLP Error obstacles are the most frequent, they are not the biggest threat to the UX of VUIs. We found that when encountering an NLP Error obstacle, participants relied the most on the Hyperarticulation tactic. For the majority of the cases, if it is clear to the participant an NLP Error occurred, she is confident that Hyperarticulation is the solution. For this obstacle, we observed participants having the most correct mental model of the error. Participant 12 explains his correct mental model for overcoming an NLP Error obstacle. *"So I stopped for a second and thought about the words that it was having trouble with, and what specifically it was having trouble with, with that word. So with the whole 'ethics' thing, it was the problem of the 'TH' sound versus the 'S' sound. And so I took a second to make sure the next time I did it I really enunciated the 'TH' sound so that it tried to do something a little better."*

The other obstacles, Unfamiliar Intent, Failed Feedback, and System Errors, had a more distributed range of tactics. This could indicate that while NLP Error obstacles occur more often, participants did not have a correct mental model of the other obstacles and were less clear on how to solve them. As discussed, the fallback tactics 6-9, Settling, Restarting, Frustration Attempts, and Quitting, indicate when participants were experiencing confusion and/or frustration. As seen in Figure 5, Settling, Restarting, and Quitting tactics made up more of the total tactics used for the Unfamiliar Intent, Failed Feedback, and/or System Error obstacles. Only for the Frustrated Attempts tactic do we see NLP Error and Unfamiliar Intent obstacles having the highest and similar percentage. This could indicate that users were forced to rely more heavily on the fallback tactics for the Unfamiliar Intent, Failed Feedback, and System Error obstacles. And since these tactics show frustration and confusion are present, these obstacles could pose a larger threat to the UX of VUIs by raising negative emotions.

However, the clearer mental model that participants had of NLP Error obstacles relies on the feedback given by the VUI system. *DiscoverCal* was designed to "read" back what it processed the user said for many different types of errors. Without this type of feedback, participants may not have known an NLP error occurred, and may not have been able to create a correct mental model of the error. We note this to highlight the importance of feedback in a VUI system and suggest future VUI designers consider the ways their system will let users know or find out if an NLP error occurred.

Guessing tactics observed more than GUI usage and recall.
Unfamiliar Intent obstacles were seen to have the most diverse tactics used initially in our probability tables (Figure 4). When looking at the breakdown of the tactics used for this obstacle in Figure 5, we see the the VUI menu was often neglected. The Recall tactic, which indicates when participants relied on their knowledge of the system and remembered a correct utterance or intent, was low for each obstacle too. Participants approached Unfamiliar Intent obstacles without consulting the menu shown much less than expected and said what they believed to be the correct utterance instead. The Rely on GUI tactic made up <10% of tactics used for Unfamiliar Intent, NLP Error, and System Error obstacles; even though our VUI's menu was always displayed. Relying on GUI support was not the first choice of our participants. Instead participants relied more on Hyperarticulation, Simplification, New Utterance, and Use More Info tactics.

The Simplification, New Utterance, and Use More Info tactics are different versions of guessing. These tactics suggest the participant is trying to figure out what amount of information the VUI needs to execute an intent without consulting the VUI's menu. And since these tactics made up a higher percentage of the total tactics used for Unfamiliar Intent, NLP Error, and System Error obstacles, we can speculate participants were more comfortable with guessing and exploring these obstacles than relying on the GUI. The exception here is the Failed Feedback obstacle with the Rely on GUI tactic making up 12.31% of the tactics used. This could indicate that with Failed Feedback obstacles, participants feel less comfortable guessing and revert to visual aid. Since all of our participants had technical backgrounds, further research is needed to analyze if this characteristic impacted this "guessing" approach.

As the Failed Feedback obstacle indicates, this does not mean that visual aides are obsolete. It could mean that VUI design needs to place a higher emphasis on designing for exploration. VUI systems can be built to support this. Research on effective VUI tutorials exists [2], but we believe further research on designing utterances to support this initial exploration by participants can significantly aid in reducing Unfamiliar Intent obstacles. Visual guides can be provided, if the VUI system has a visual component, during detected obstacles instead; providing the support users need.

CONCLUSIONS & FUTURE WORK

This paper identified 4 major categories of obstacles people face when using an unfamiliar VUI and 10 types of tactics they use to overcome them. We analyzed the patterns in which participants employed different tactics for each obstacle category and identified the transition patterns between tactics. In our results, the opposite tactics of Simplification and Use More Info are the most frequently used tactics, after Hyperarticulation. We found that although NLP Error obstacles are the most common, the other obstacles caused more frustration and confusion to our participants. This indicates that improving VUI's UX requires further research in both NLP and in interaction design (e.g., feedback). Regarding tactic patterns, our participants, even with their technical backgrounds, relied more on guessing and exploration than knowledge recall or visual aids. This highlights the need for further research on supporting user-directed exploration and learning, in addition to standard tutorial and menus in VUIs.

Our study is limited to a small size of 36 sessions with 12 participants. Future work can be extended to analyze a larger sample and different demographics; specifically examining the obstacles and tactic patterns of non-technology comfortable users. *DiscoverCal* is also a VUI with a visual display, certain tactics such as Rely on GUI will not apply to all VUIs. Our study analyzes data retrieved from user interactions with a single context VUI (calendar management). Our categories can be used to evaluate more advanced, multi-context VUIs, like Alexa and Google Assistant. These categories can also be used by VUI designers to evaluate their own VUI systems.

REFERENCES

1. Matthew P Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd: Relationship Counselling for HCI and Speech Technology. *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (2014), 749–760.
2. 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.
3. 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.
4. 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.
5. Clare M Karat, Christine Halverson, Daniel Horn, and John Karat. 1999. Patterns of entry and correction in large vocabulary continuous speech recognition systems. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* (1999), 568–575.
6. Lewis R. Karl, Michael Pettey, and Ben Shneiderman. 1993. Speech versus mouse commands for word processing: an empirical evaluation. *International Journal of Man-Machine Studies* 39, 4 (oct 1993), 667–687.
7. Ranjitha Gurunath Kulkarni, Ahmed El Kholy, Ziad Al Bawab, Noha Alon, Imed Zitouni, Umut Ozertem, and Shuangyu Chang. 2017. Hyperarticulation Detection in Repetitive Voice Queries Using Pairwise Comparison for Improved Speech Recognition. (2017), 4985–4989.
8. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (2016), 5286–5297.
9. Gabriel Lyons, Vinh Tran, Carsten Binnig, Ugur Cetintemel, and Tim Kraska. 2016. Making the case for Query-by-Voice with EchoQuery. *Sigmod* (2016), 2129–2132.
10. Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is my new BFF". *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (2017), 2853–2859.
11. Pernilla Qvarfordt, Arne Jönsson, and Nils Dahlbäck. 2003. The Role of Spoken Feedback in Experiencing Multimodal Interfaces as Human-like. *Proceedings of the 5th international conference on Multimodal interfaces ICMI 03* (2003), 250–257.
12. Xin Rong, Adam Fourney, Robin N Brewer, Meredith Ringel Morris, and Paul N Bennett. 2017. Managing Uncertainty in Time Expressions for Virtual Assistants. *Acm* (2017), 568–579.
13. Nicole Shechtman and Leonard M Horowitz. 2003. Media Inequality in Conversation. *Proceedings of the conference on Human factors in computing systems - CHI '03* 5 (2003), 281–288.
14. Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (2000), 63–65.
15. Amanda J. Stent, Marie K. Huffman, and Susan E. Brennan. 2008. Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication* 50, 3 (2008), 163–178.
16. A. Strauss and J. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for developing grounded theory.* (second ed.). SAGE Publications.
17. Yu Zhong, T V Raman, Casey Burkhardt, Fadi Biadsy, and Jeffrey P Bigham. 2014. JustSpeak: Enabling Universal Voice Control on Android. *Proceedings of the 11th Web for All Conference on - W4A '14* (2014), 1–4.

View publication stats