Tea, Earl Grey, Hot: Designing Speech Interactions from the Imagined Ideal of Star Trek

BENETT AXTELL

University of Toronto, TAGlab, Toronto, Canada

COSMIN MUNTEANU

University of Toronto Mississauga, ICCIT, Mississauga, Canada

Speech is now common in daily interactions with our devices, thanks to voice user interfaces (VUIs) like Alexa. Despite their seeming ubiquity, designs often do not match users' expectations. Science fiction, which is known to influence design of new technologies, has included VUIs for decades. *Star Trek: The Next Generation* is a prime example of how people envisioned ideal VUIs. Understanding how current VUIs live up to *Star Trek*'s utopian technologies reveals mismatches between current designs and user expectations, as informed by popular fiction. Combining conversational analysis and VUI user analysis, we study voice interactions with the *Enterprise*'s computer and compare them to current interactions. Independent of futuristic computing power, we find key design-based differences: *Star Trek* interactions are brief and functional, not conversational, they are highly multimodal and context-driven, and there is often no spoken computer response. From this, we suggest paths to better align VUIs with user expectations.

CCS CONCEPTS • Human-centered computing~Natural language interfaces • Human-centered computing~User interface design

Additional Keywords and Phrases: Voice user interfaces, Speech interaction design, Conversational analysis, Star Trek

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Smart voice assistants, like Alexa, have quickly become ubiquitous in homes and in phones. These are seen as the "next thing" in industry, with much media hype being attached to the emerging tools and applications [52], and their speech-only modality has been heralded as ideal tools for users that are less digitally literate, such as older adults [12]. Though their use continues to rise, usability remains low [14]. Given their ubiquity and the long history of design for these voice user interfaces (VUIs), dating back to Put That There [7], one would expect broader and deeper use, as has been seen with other modalities.

VUIs continue to focus on supporting human-like conversations for a broad, nearly all-inclusive set of tasks. However, previous research into use of these VUIs and into human expectations for human-computer voice interactions find that replicating human-to-human conversation is not desired or helpful [16,31]. Further, there is limited evidence that voice alone will grow to be a default or universal modality [4]. Users have a limited sense of what can be done with these tools [13], and select which tasks to do with voice based on external criteria, like their location [16].

Across the board, not enough is known about how VUIs should be designed to best support users and what they should support. Some work has begun to explore design heuristics for this area that will greatly increase the knowledge base [35].

As this is still a relatively new modality for general use, the knowledge base of VUI design would be significantly improved by gathering how users imagine these systems could work and analysing that hypothetical use. This would lead to designs that better match users' expectations, as influenced by the common popular media of *Star Trek*.

These types of speech-based interactions have been a consistent presence in science fiction media since the mid-19th century, with some of the most famous examples being introduced in the mid-20th. Some take the form of anthropomorphic robots (e.g., Robbie from Asimov's *Robot* series [3]). Closer to today's VUIs, which function without a physical interface, is HAL9000 from *2001: A Space Odyssey* [27]. *Star Trek* has been a particularly strong source of imagined future technology from its original airing in the 1960s through the spin off series from the late 1980s to today, including a fully-integrated VUI system. The influences of these designs on modern technology are broad, from cell phones and automatic doors, to tablets and voice assistants. While today's VUIs are becoming as ubiquitous as seen in these futuristic settings, the uses and usability have a long way to go before we can ask for "Tea, Earl Grey, Hot", as the crew of the *Enterprise* [36].

Science fiction has long been an influence on new designs and interactions. HCI researchers often look to science fiction as a source of innovation [38,41]. *Star Trek* has been a notable source of cultural impact, as found by various works in digital media and design [1,10,17,25]. Further, the genre has not only contributed to the design of specific technologies that exist today, but has directly influenced tech designers, developers, creators, and entrepreneurs [32]. As *Star Trek* is based in a near-utopian future with our ideal technology and interactions, it offers a unique opportunity to compare what we can do now with what we had imagined. This can be seen in innovations like the flip phone and the touchscreen tablet.

In order to better understand how actual speech interactions with VUIs can improve, we review speech interactions on *Star Trek: The Next Generation* (*TNG*). As a long-running TV series, it presents consistent, long-term VUI use, creating a large and meaningful dataset. This includes both repeated use, especially for mundane tasks (e.g., making a cup of tea), and exceptional use, such as problem solving through a back-and-forth conversation. VUI use across this series reflects user-designed ideal interactions. We decide to treat these interactions as though collected in-the-wild in order to better compare to existing use and to reflect *Star Trek*'s strong influence on actual design and use. We analyze these interactions to create design recommendations towards closing the gaps between our fictional future and today. As *Star Trek* shapes both popular culture and our expectations of technology, analyzing the speech interactions in *TNG* will help us understand the gap between the promises of *TNG* and our current reality.

TNG aired from 1987 to 1994, during which time many technological advances became mainstream, most notably the Internet. These advances were reflected in key changes from the original series of *Star Trek*. We focus on this series over the original as this time period allowed the designers of the show to imagine speech interactions based on their new experiences with home computers and digital communication as was not possible when the original series aired in the 1960s. We focus on this series alone, rather than including the overlapping spin-off series (*Deep Space 9* and *Voyager*), as it presents a more cohesive and standard use of the available technologies, where the others are literary edge cases as the shows intentionally try to be different than the main show.

Recent studies have reviewed in-the-wild speech data from generally available devices [28,29,45,46]. These previous works allow us to align our analysis with the actual tasks performed by users today, meaning that the results, though stemming from fiction, can be meaningfully compared to current use. The tasks completed by the computer's VUI on *TNG* largely match what home devices are used for now, for example, simple searches and controlling the internet of things.

We analyze the speech data collected from all *TNG* episodes. We build an understanding of this data by following methods that are now the de facto standard for analysis of human interactions with VUIs such as Alexa [26,40]. These methods are derived from conversational analysis (CA) to match the computer-focused setting. This analysis is not as

detailed as a human-to-human one would be, as the limitations of the digital system allows for a simpler approach (e.g., gestures are not used). We then present design recommendations based in this analysis to bring us closer to the future.

The coding and analysis of all speech interactions with the *Enterprise* computer system supports our primary contribution: design recommendations that can lessen the gaps between Alexa and *Star Trek*, based in those user-imagined ideal interactions. These design recommendations, which are technically feasible today, bring us closer to matching users' expectations, which is still much needed in the design of VUIs. *Make it so.*

2 BACKGROUND

Recent years have produced several reviews of actual use of VUIs, mainly Alexa [28,29,45,46]. These provide insight into the primary domains of use and user demographics. These datasets are limited in that they are not able to access the larger context around the interaction. In particular, they are not able to see which interactions are part of a longer back-and-forth and which are one-offs.

Past works have explored how VUIs are perceived and used in a variety of settings, including with children [11,30,48], in families [5], and with older adults [50]. Much research has been focused on making these interfaces more human and more conversational [19], though others question whether that should be the goal [4,16].

2.1 Conversational Analysis

While the reviews of speech interactions have been valuable to understand real use of these systems, we are not aware of research that does a conversational analysis on this in-the-wild data, though several have explored use of spoken language with computers through other methodologies [22,23].

Conversational analysis is a sociological method that reveals how conversations unfold. Research in this area has defined the adjacency pair (AP) to be a two-part exchange between two speakers and outline common types of APs in human conversation [42]. We use the AP to explore the differences between the initial interaction between person and computer (e.g., person gives a command and the computer turns on the lights in response) to the interactions that follow after that first interaction, if any.

CA has been used within dialog segmentation research [34] and has been adapted to study voice interaction corpora such as collected from Alexa device logs [26,40] . Although methods used to study VUI interaction logs seem limited when compared to human-to-human CAs, these methods are nevertheless powerful in understanding design needs of human-to-computer dialogue and conversation [44]. Applying this to actual or envisioned speech interactions with VUIs could help reveal how our speech differs with computers than with humans, expanding on previous works [15]. Similar previous works with VUIs have explored how the users structure voice queries and have found that they do not match human-human spoken language [24]. This will provide a stronger base understanding for future designs. We use a simplified conversational analysis with this fictional data, not considering gestures and other aspects of human conversation that do not apply to these VUI interactions, in the hopes that this will be expanded to actual data in the future.

2.2 Digital Media and Science Fiction

Digital media research has long explored how our media influences the design of technologies. Examples include comparing science fiction interactions and ubiquitous computing [18], exploring how it shapes expectations from technology and, therefore, its design [21] and even how it affects e-commerce [49]. *Star Trek* is a particularly common topic within these works, including how it influences programmers and developers [17], scientific researchers and their writing [1,25], and the early phases of what would become Siri [10].

Clearly, science fiction, especially *Star Trek*, has been a consistent influence on the design of new technology. However, its effect on VUIs and speech interaction, specifically, is yet to be explored.

2.3 *Star Trek* Definitions

We provide here some definitions of common terms from the *Star Trek* universe that we will be using throughout this paper, adapted from Memory Alpha, the primary *Star Trek* wiki [53].

Enterprise: the starship at the centre of the *TNG* series.

Federation: the union of planets, including Earth, that share technology, among other benefits, including the computer system reviewed here. All Federation ships, including the *Enterprise*, use the same computer voice interface.

Turbolift: or turbo-elevator, provides both vertical and horizontal transportation for personnel through turboshafts between key sections of a ship.

Holodeck: a room which enables holographic projections and holograms, which have the illusion of substance. Used for entertainment, training, and scientific analysis.

Replicator: a device that transforms matter from one form into another. Mainly used to create food and drink.

These last three technologies are controlled primarily, if not solely, through voice commands.

3 METHODS

Our human-to-computer CA uses the dialog produced across the *TNG* series to build a database of speech interactions with the Enterprise computer. Human-to-human CA incorporates many aspects of conversation, such as gestures and how parties opt in and out of the interaction. Given that our data is exclusively with a computer that does not recognize gesture and has no choice whether to join the conversation or not, we consider the conversation, in this setting, to be what each party has said, plus resulting actions performed by the computer. In this way, the *TNG* computer is similar to Alexa and other VUIs, going back to early VUIs, such as air travel information systems [47] or dialogue systems to elicit weather forecasts [51].

The *TNG* interactions were coded, using previous works analyzing actual in-the-wild speech data [28,29,45,46] and previous conversational analysis works [42] to build the initial code book. The analysis focuses on what factors influence a person's use of speech in the futuristic, idealized setting and how that compares to modern, actual use. Below we detail the data gathering and analysis processes before discussing our findings.

3.1 Source Data

The data consist of all speech interactions with or by the *Enterprise* computer in the 176 episodes of *TNG*. The data are in the form of complete and verbatim transcripts of each episode, including markups such as which character is speaking and other stage directions (e.g., description of a specific setting or object the characters focus on or interact with). Each episode is stored as a JSON file containing all dialog and any relevant setting and action information, for example, when a character is in a turbolift or when the computer responds with a visual output on a monitor.

In all, from 69,355 lines of dialog across the series, our dataset consists of 1,372 individual utterances made either by a person to a computer or vice versa, from 587 interactions, ranging from 1 to 32 turns (16 APs). These utterances do not include 411 action-only responses by the computer, such as creating a cup of tea without saying anything. 67% of interactions are one AP, and 95% of interactions are under 10 turns. 62% of individual utterances are by a person.

3.2 Coding Method

These interactions were identified through a combination of algorithmic and manual coding of the transcribed episode data. Algorithmic coding was used to flag likely candidates in the dialog based on speaker (i.e., the computer), setting (e.g., the holodeck), and dialog content (e.g., use of a wake word, like "Computer" or "Alexa"). This data was then manually coded by the researcher for type of interaction (e.g., command, question, response), and domain of use (e.g., entertainment, IoT, etc.). The domain codes are pulled directly from previous works which analyze actual speech interactions [2,29], as can be seen in Table 1, with the addition of "Analysis", which includes tasks like "Is there a pattern?" or "Impact analysis, Computer" . These complicated tasks are largely not possible with modern computing power and represent the only domain that leverages the futuristic capabilities of this dataset.

Table 1: Domains and sources from previous works.

Domain	Ammari et al [2]	Lopatovska et al [29]
InfoSeek	Search	Quick Information Searches
Entertainment	Music	Entertainment
IoT	Internet of Things	Control External Devices
	Smart home and IoT hubs	
Analysis		
Other (e.g., Communications)	Timers	
	Macros and Programming	
	Family Interactions	
	Privacy	

The interactions are drawn from conversational analysis [42] and dialog segmentation [34], plus iteratively generated codes for computer-specific types, like a system alert (e.g., "Red Alert"). The full dataset, including the algorithmic flagging, and the codebook defining all domains and interaction types have been made publicly available[1].

The interaction types are intentionally at a higher level than an in-depth CA would use (e.g., question rather than subtypes of questions) to match the data available from those actual reviews of use in order to enable the comparison. Future work will be able to dive into this deeper analysis based in the higher-level understanding gained here. Using the final codebook, a second, unaffiliated researcher coded a randomly selected 15% of dialogs, with an inter-rater reliability above 85% for all codes.

Many interactions include more than one type, for example "Computer, where is Captain Picard?" which has both a wake word and a question. In order to analyze frequencies of various types, we consider all types together. We have also defined a ranking of types based on their relevance to the interaction, which we use when analysing summative ratios of use across the entire dataset (i.e., so percentages of types sum to 100%). This was created from their frequency of use and from expected use from available tools, in order to determine the primary type for each interaction. These rankings for both person and computer interactions, along with definitions and examples for each type, can be seen in Tables 2 and 3, respectively. When multiple types are present, the highest ranking one for the speaking character is the primary type, so in the case of the example above, the primary type is the question as that is the most substantive part of the query; it is more relevant to the interaction that the person is asking a question than that they are using a wake word.

[1] http://www.speechinteraction.org/TNG/

Table 2: Ranking of interaction types for person.

Rank	Person	Definition	Example(s)
1	Command	Utterances that directly tell the computer what to do.	Run a diagnostic on the port nacelle.
2	Question	Utterances that ask the computer for something.	Where is Captain Picard?
3	Statement	Utterances that don't tell the computer or ask it but meaning is inferred.	Deck four. I wish to learn about jokes.
4	Password	Utterances that contain a password.	This is Captain Picard.
5	Wake Word	Key phrases used to activate the computer VUI.	Computer Holodeck
6	Comment	Utterances that have no intended action for the computer	Excellent. Ferrazene has a complex molecular structure.
7	Conversation	Utterances that are more like human conversation, such as phatic expressions, formalities, and colloquial speech	Well, check it again! Then run it for us, dear.

Table 3: Ranking of interaction types for computer.

Rank	Computer	Definition	Example(s)
1	Clarification	Utterances asking for more information.	What temperature?
2	Response	Utterances that respond to a person's query or action	
3	System Alert	Utterances that either respond to a person's requested action (often a warning) or are prompted by the system, rather than a person	That is not recommended. Outer hull breach.
4	Information	Utterances that inform users of what is happening without being prompted to do so (which would be a response)	Opening main shuttle bay doors
5	Countdown	Utterances that are part of a countdown.	Self destruct in 30 seconds. Self destruct in 20 seconds.
6	In Progress	Utterances that indicates that a process is ongoing or updates on that process	Accessing.
7	Conversation	Utterances that are more like human conversation, such as phatic expressions, formalities, and colloquial speech	You're more than welcome, Commander Riker.

59% of interactions include more than one interaction type and 32% contain a single type, ignoring wake words, including less than 20% for each of the most common interactions (commands, questions, statements, and responses).

All speaking characters, other than the computer, are treated as people in our database. In particular, we choose to consider the speech from Data, a near-human android, as person interactions and never a computer interaction. He is treated as another person by the crew, and, although he has direct, internal access to the computer system at all times, he often interacts with it through the same modalities as the other people, namely touch and speech.

Manual coding also identified and removed speech interactions with non-Federation computer systems. These data points are equivalent to the real-life analysis of an Alexa conversational corpus that captures, e.g., a few Siri interactions, so they are excluded in the name of consistency of use. This also allows us to include only English language interactions, rather than Romulan ones. We also exclude one interaction on the *Enterprise* with a children's computer, which while part of the *Enterprise* system uses a different VUI system, including a different computer voice. This single interaction is explored further in the discussion section.

3.3 Domains

As discussed in the previous section, three of the four domains identified by our analysis are common to existing speech interactions, and match previous findings [2,28,29,45]: Entertainment (e.g., holodeck and music tasks), IoT (Internet of

Things: e.g., replicator and turbolift), and InfoSeek (general data search tasks or locating a crewmember). 71% of all interactions fall within these pre-defined domains. We compare the frequency of domain use on the *Enterprise* to those seen in reviews of modern VUI use. The fourth domain is Analysis, which reflects the futuristic computing power available to the *Enterprise*, as they are able to verbally prompt the computer to complete complex analysis of data, such as hypothesizing the source of a theoretically impossible signal. Other domains, rare in our dataset include Communications (e.g., triggering a call to another crew member) and Help (e.g., asking the computer how to use the system).

4 FINDINGS

We present first an overview of use of types, domains, and wake words. We then explore how the types people use changes across domains (IoT, Entertainment, InfoSeek, and Analysis), and how the computer response changes based on both domain and interaction type. We also compare how these changes between the first adjacency pair (AP1) (i.e., one utterance by either person or computer and one response by the other) and the remaining interactions (AP2-N), consisting of the subsequent interactions beyond the first AP. For example, in the following exchange [8]:

Data: Computer, request all biographical information on fictional character Dixon Hill.

Computer: Working. Character first appeared in pulp magazine, 'Amazing Detective Stories,' copyright 1934, A.D. Second appearance in novel 'The Long Dark Tunnel,' circa 1936.

Data: Request complete text of all stories involving said character.

Computer: [Action: shows text on console screen and scrolls]

Data: Increase speed.

Computer: [Action: increases scroll speed]

AP1 is the Data's initial command and the computer's response. The next four lines or actions form two turns (APs): AP2 and AP3. These are AP2-N, everything that follows the first AP.

This allows us to elaborate on differences between the initial interaction versus extended interactions to see how the quick interactions compare to cases that require further back-and-forth (e.g., what types are used by people after the first AP) as well as how interactions are initiated and how they are continued (e.g., when is a wake word used). This data cannot be compared to existing studies of actual use, as the restrictions on that data do not reveal the turn-taking information. Additionally, multi-turn interactions are limited by the design of current tools, which generally require another wake word before subsequent interactions. 67% of interactions in our data set consist of a single AP.

4.1 Overview of Speech Use

Of all interactions by a person, 59% are commands, and 50% of all interactions use a wake word (e.g., "Computer"). Nearly half (48%) of computer interactions are action-only responses, meaning it performs an action (e.g., creating a cup of tea) without any verbal response. In this case, the computer relies on other cues, generally visual, to communicate task completion. Spoken computer interactions are largely responses (47%), followed by system alerts (12%), which are analogous to smart reminders (e.g., a notification that it will be raining and the user should bring an umbrella). This shows us that interactions are largely practical and targeted, rather than extend, conversational exchanges (**F1: Practical interactions**). Table 4 shows all frequencies of interaction types by person and computer.

Table 4: Frequency of interaction types by people and by the computer.

Type	% by Person	% by Computer
Command	59%	1%
Wake Word	50%	

Type	% by Person	% by Computer
Statement	24%	
Question	16%	1%
Conversation	10%	
Comment	3%	
Password	4%	
Action		48%
Response		47%
System Alert		12%
Information		6%
Countdown		6%
Clarification		5%
Error		4%
In Progress		3%

85% of interactions fall into the four main categories defined in the previous section: InfoSeek, Entertainment, IoT, and Analysis (see Figure 1). Analysis interactions are fairly even split between people (55%) and the computer (45%). However, people provide 75% of the interactions of IoT and Entertainment tasks, and 61% of InfoSeek. The three pre-existing domains, fall within similar ranges of frequencies as those seen across previous studies [28,29,45,46], meaning the supported tasks are largely in line with what is currently available.

The other domains were related to communications and emergency interactions, such as asking the computer to call an individual or the computer announcing a self destruct countdown. As neither of these were represented in the available real word data and were rare in our dataset, they are not analyzed as their own domains.

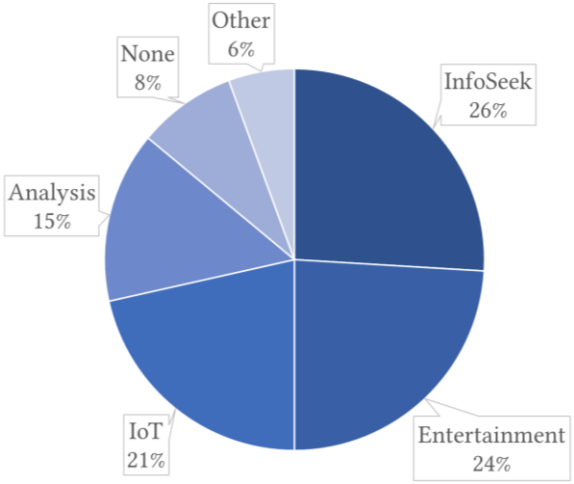

Figure 1. Distribution of Domains

Initially (for AP1) wake words are used in 69% of interactions, but this drops to almost a third of that (26%) for AP2-N. Users assume that after the initial interaction, they will be able to continue without re-triggering the computer. Figure 2 shows distribution of wake word use for AP1 and AP2-N, overall and across common domains of use. Use of wake word drops by at least 50% in all domains after AP1.

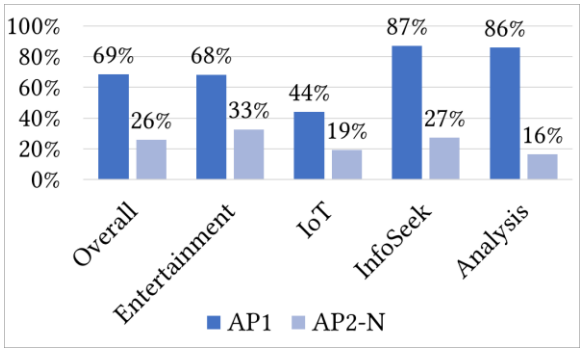

Figure 2. Use of wake word by domain for both AP1 and AP2-N

Wake words are least common for IoT tasks, as the physical context of the task implies the use without the need for a verbal trigger. Entertainment has the highest re-use of wake words (33%). Subsequent interactions with music or the holodeck have to contend with the sound of the ongoing interaction, that is, the music playing or the characters talking on the holodeck, so the computer is less likely to understand a repeat interaction as such without another wake word.

InfoSeek and Analysis have the highest rates of wake word use in AP1, and both have a significant drop in use for AP2-N. As with IoT, this shows the role of physical context in predicting the ongoing speech interaction. These are also spaces that are more likely to take multiple turns to complete than Entertainment or IoT tasks, which often are simply one query followed by a computer action.

Wake words are used only when necessary, and the users understand the contexts in which subsequent wake words would be necessary. The system uses the domain context of the initial interaction to determine when further interactions are likely (**F2**: **Context over wake words**).

4.1.1 Key Takeaways

Short, practical interactions are preferred over longer conversations, seen both in the prevalence of simple commands and in users' understanding of when wake words will be necessary based on the larger context of the interaction.

4.2 How People Interact with the Computer

Figure 3 show the distribution of primary types spoken by people for AP1 and AP2-N, overall and across common domains. This demonstrates the extent to which a user's interaction type is influenced by the domain of use.

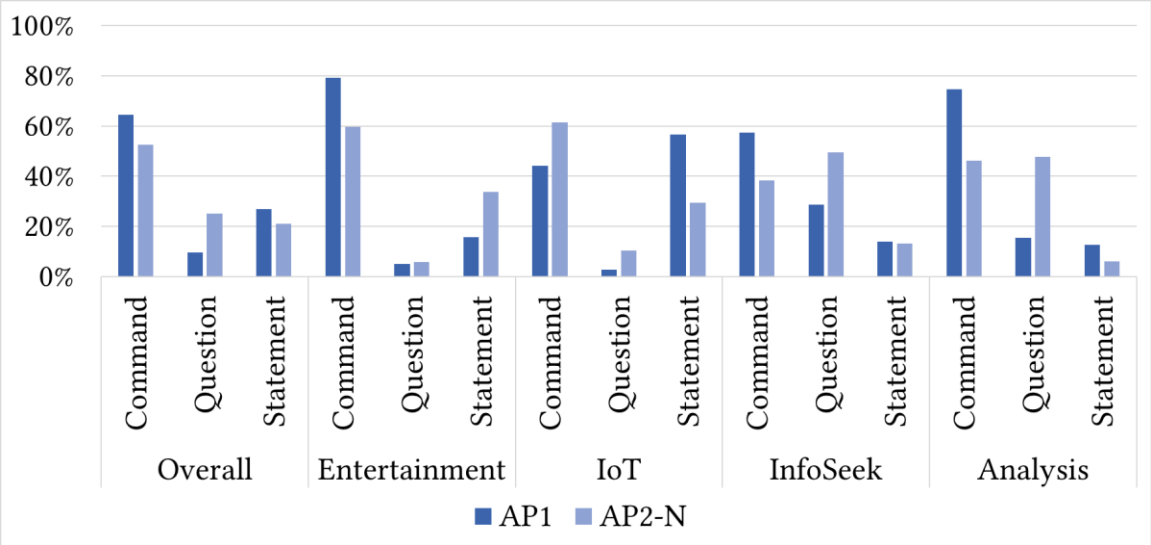

Figure 3. Use of commands, questions, and statements by people across domains in AP1 vs. in AP2-N.

There is a drop in frequency of commands after AP1, while questions increase to 25%. This demonstrates that, while the large majority of overall interactions from people are commands, after the initial query there is likely a need to extrapolate, which lends itself to the use of questions (**F3: Commands first**). This use of questions later in interactions also indicates to the computer that a verbal response is required, as we discuss below.

Statements are only common in IoT settings, where they make up 57% of all AP1 interactions, compared to 27% overall for AP1. The physical context of the IoT interactions allow users to state the least amount of information possible. For example, once on the turbolift, it is sufficient to say, "Deck four", without a wake word or a complete sentence like, "Take me to deck four".

Questions are very rare in Entertainment and IoT, as questions about those areas are often actually InfoSeek ("e.g., Who wrote this *Star Trek* episode?"). Questions are generally uncommon in AP1, with most use seen in InfoSeek (29%), but make up almost half of all interactions for InfoSeek and Analysis in AP2-N. This reflects the inquiring nature of these tasks. For Analysis specifically, while AP1 is likely to initialize the data to be analyzed (e.g., "Show me the results of the diagnostic"), subsequent interactions would want to understand something about the initial result. For example [20]:

Data: Computer, perform a level one diagnostic on the exocomp's command module.

Computer: The command pathways are functioning normally.

Data: How can that be if the interface circuitry is burned out?

Computer: The interface circuitry has been repaired.

Data's second query here, a direct follow up to the first relying on that existing context, expects the computer to understand the potential relationships between different subsystems and analyze them in order to determine the original problem. How a user phrases the initial query can provide important context to the system informing the intended domain (**F4: Grammar provides context**).

Commands are initially the most common type in all domains except IoT, where quick statements are preferred, though commands still make up nearly half of interactions. Use of questions increase across domains after AP1, though commands are still the primary type for all domains, except InfoSeek. The only domain in which commands increase for AP2-N is

IoT. This reflects again, that using brief interactions, like commands and statements, is common for the initial interaction, but subsequent turns need more detail.

4.2.1 Key Takeaways

Users structure their interactions based on the intended domain of use, such as using statements for IoT where the majority of the command is easily inferred. Users also adjust how they structure after the initial interaction to extrapolate on the initial turn.

4.3 How the Computer Interacts with People

Figure 4 shows frequency of answer types for commands, questions, and statements for both AP1 and AP2-N, highlighting how common adjacency pairings are used across the interaction. In AP1, 81% of statements receive an action, and only 12% receive a response, but after AP1, actions drop to 37% while responses double to 28%. Statements do not provide a lot of information and rely on the computer being able to infer the implying command. As such, they lend themselves to quick, clear interactions. Subsequent interactions that use statements are more likely to need further clarification, meaning the computer will need to respond verbally.

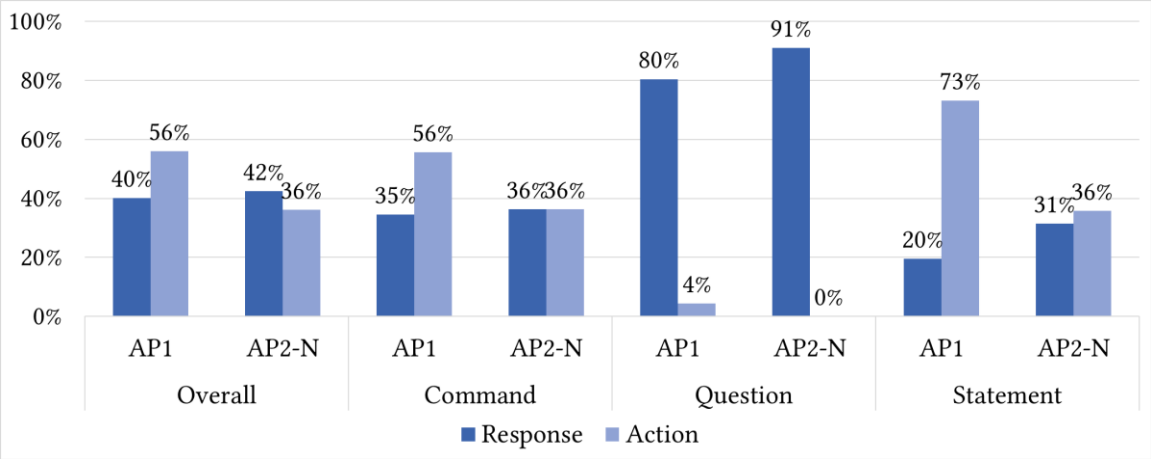

Figure 4. Frequency of responses and actions in response to interaction types for AP1 and AP2-N

Questions almost always require a verbal response for both AP1 and AP2-N. Due the nature of conversation, it is difficult for the computer to respond to a question without speaking. This reflects the need for further clarification, on the user's part, after the first pair, as well as the general increase of verbal response by the computer after AP1.

Commands are more evenly split between receiving an action (56%) or a response (35%), as they cover a broader range of use. As with statements, frequency of actions in response to commands drop after AP1 (seen in both Figures 3 and 4), as these similarly require further follow up from the computer. How a person structures their interaction provides key context to the system, and informs the computer's choice of action vs. spoken response (**F5: Grammar infers response**).

Figure 5 show frequency of computer actions and responses for both AP1 and AP2-N, across domains of use. From AP1 to AP2-N, frequency of responses is similar, but computer actions are present in over half of initial interactions and drops to 36% in AP2-N. As with the shift from command to question in people's interactions, this likely reflects the computer's need for clarification when the interaction goes beyond the one AP, which is further influenced by the domain of the interaction.

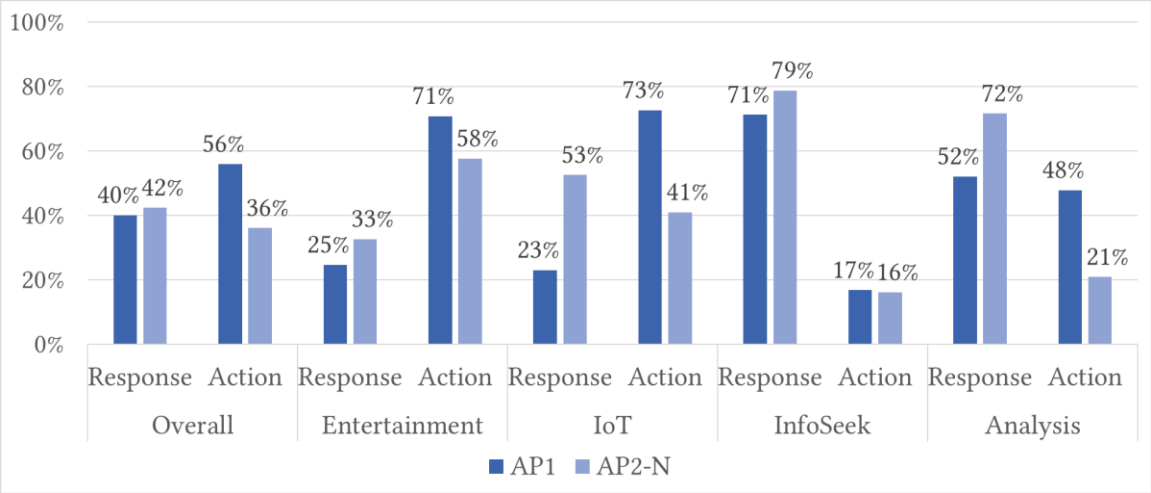

Figure 5. Frequency of computer's verbal responses and non-verbal actions overall and across domains, for both AP1 and AP2-N

Entertainment and IoT interactions receive actions in over 70% of AP1, while responses are less than 25%, as the primary uses of these domains are action-specific and their completion is clear from the action without need for verbal answer. Frequency of actions drop for AP2-N, while responses rise, likely again due to need for more information if the interaction goes past one turn each.

InfoSeek, however, often has no associated action to be preformed, requiring a spoken answer with little change between AP1 and AP2-N. For example [9]:

Mrs. Troi: Tell me, computer, is Commander Riker still on the bridge?

Computer: Negative. Riker is currently in Holodeck Three.

Mrs. Troi: Holodeck? Where is that?

Computer: Follow the com panel lights. They will lead you there.

Analysis in AP1 is evenly split between actions and responses, but, as with Entertainment and IoT, actions fall for AP2-N, while responses rise. As seen with people's commands and questions in this domain, AP1 is often an initialization, which can be shown through, for example, a visual cue, over a spoken answer. Subsequent interactions will be more question-heavy, requiring the computer to speak. Along with the user's interaction type, domain of interaction and turn (whether it is the initiating interaction or not) further informs the computer's appropriate response type (**F6: Domain infers response**).

4.3.1 Key Takeaways

Based on how the user structures their interaction and the associated domain of use, the computer can determine an appropriate type of response.

4.4 Taking Turns: Types in Adjacency Pairings

Figure 6 shows distribution of the four most common AP interaction types for AP1 and AP2-N. Of all interaction types seen in AP1 (e.g., a person gives a command followed by a computer response – command to response), the majority are commands followed by an action (36%) or by a response (22%), showing that quick commands are the primary use of speech interaction. 18% of AP1 pairings are statement to action, which shows the preference for short interactions when

the action is apparent and clearly the final step in the process (e.g., stating a beverage to the replicator does not require a response beyond creating that beverage). Questions to responses are very rare in AP1, but are the most common pairing in AP2-N, reflecting the decrease in action responses and the more complicated nature of longer interactions. This all supports that short, targeted tasks are primary intention of speech interaction (**F7: Short interactions**).

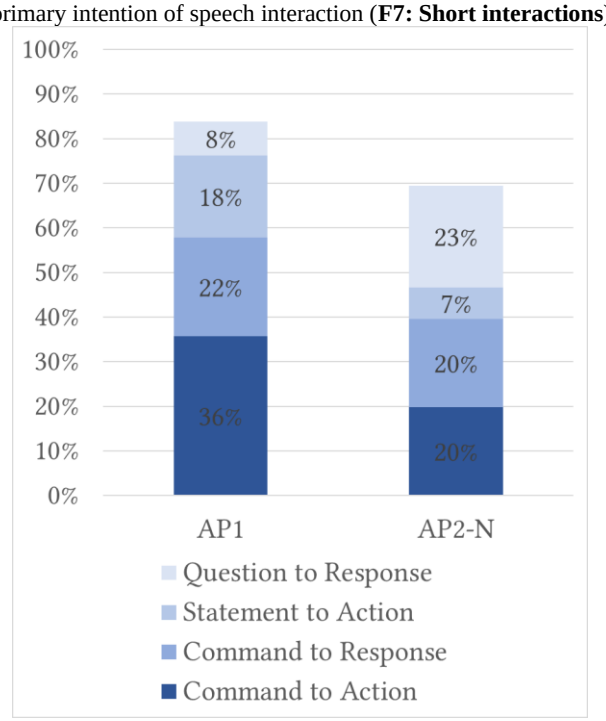

Figure 6. Frequency of most common AP interaction types.

4.4.1 Key Takeaways

Action responses are leveraged by the *TNG* computer to keep interactions short and practical. Longer interactions are less likely to complete with an action, reflecting the need for more information as an interaction continues.

5 DISCUSSION

Our data analysis of computer interactions with the *TNG* computer reveals similarities in domains of use to available VUIs, such as Alexa, and highlights patterns in the structure of voice interactions. Our corpus includes all interactions from the series, meaning that interactions with novice users, negative interactions (e.g., errors and unexpected responses by the computer), and other situations common to modern VUIs are all included. Despite the fictional premise, this creates a realistic snapshot of use, as can be seen in the consistent trends observed across our analysis. Our key findings along with design implications are summarized in Table 5.

Table 5: Summary of findings and implications

TNG Findings	Current Design Standard	Implication

Most interactions are one turn (F7: Short interactions) and are mostly commands (F1: Practical interactions). After AP1, questions overtake commands (F3: Commands first) and wake words are uncommon (F2: Context over wake words).	Systems largely need to use a wake word every time and do not handle continued conversational context well.	D1: Assume interactions will be a single turn, and know how initial interactions can indicate that when longer exchanges are needed.
People use different query types based on intended domain (F4: Grammar provides context) as well as expected computer response (F5: Grammar infers response). Computer uses query type (F5: Grammar infers response) and domain (F6: Domain infers response) to determine action or verbal response.	Systems parse keywords from user queries without considering the structure of the query itself.	D2: The context provided by the structure of a query should be leveraged to determine intended domains and actions, including whether a verbal response is needed.
Interactions are rarely just speech and many leverage physical location (e.g., domain), simultaneous visual displays, and IoT interactions (F4: Grammar provides context, F6: Domain infers response).	Nearly all system interactions are speech only, even when a screen is present.	D3: Multimodality is key to useful speech interactions.
Commands are most common (F1: Practical interactions), and short, targeted interactions are preferred over extended dialogs (F7: Short interactions).	Systems design for near human-to-human interactions, regardless of the user's intent for a given interaction.	D4: Conversation is not the practical use of VUIS. Practically short user commands and computer actions cover most intended tasks.

5.1 The Fiction in Science Fiction

We interpret this data as if the interactions were in-the-wild, in order to present meaningful design recommendations for today's technology. However, there are aspects of these interactions that are clearly influenced by the futuristic computing power available on board the *Enterprise*. Still, most of the interactions are comparable to what users do today, but three sub-areas stand out as more futuristic.

Primarily, the starship is fully inter-connected and crewmembers are able to control any aspect of the ship from any location on board (a true Internet of Things) and the ship can access each crewmember's location at any time. This level of pre-packaged cooperating hard- and software is likely still far off, though a smart home can be manually enabled to perform similar tasks. Despite this, the type of tasks that leverage these futuristic abilities correspond to what users do today. IoT on the *Enterprise* is often used for environment control (e.g., lighting) and food preparation (e.g., turning on the oven, or ordering at a replicator). To find a crewmember, a user asks a question to cue the computer to respond verbally, similar to performing a quick voice search with a smartphone. These similarities in use demonstrate how IoT interactions are already moving along a path towards to capabilities seen in *TNG*.

The functionality of the holodeck is also well beyond our current capabilities, and yet many of the speech interactions match what would be done to interact with a music or video player. Command including finding the desired program and then playing, pausing, and stopping that program. The only interactions that set this apart are those that design the actual holodeck programs, which is akin to creating a movie through voice commands.

As discussed above, the Analysis type tasks often rely on a level of computing power and AI problem-solving that is not regularly available today. This is the only area in which the queries by people have little overlap with what is done today. "Run a diagnostic on the port nacelle" is unlike anything we could ask Alexa.

Despite these futuristic capabilities, the speech interactions on the *Enterprise* largely align with today's use and the majority are feasible by today's technology. They can reveal what we as a culture expect from voice interactions since, as discussed in the intro, science fiction media are known to shape how we perceive ideal use of modern technology and drive (near) future designs [32]. This is particularly true for *Star Trek,* which is designed to show our idealized technology use.

5.2 Design Recommendations

From these findings, we identify key design recommendations that do not go beyond the capabilities of modern tools and will begin to close the gap between our imagined ideal use and what is currently possible. As discussed above, science fiction, and *Star Trek* in particular, has a long history as a positive influence and inspiration for new designs and for HCI research [1,10,17,25,32,38,41]. Still, there is a limitation in drawing design recommendations from fiction, and these recommendations should be seen through that fictional lens and not assumed to be functional for all practical applications in modern tools. We have drawn these recommendations from our analysis of this fictional VUI, based explicitly in the use of existing modern VUIs. Ultimately, these recommendations, supported by the thorough analysis presented here, should be seen as exactly that, recommendations, and not definitive evidence of their practicality or usability in actual use of VUIs until they are explored by future research in those practical settings.

5.2.1 Design for short interactions, know when it will be long

The majority of interactions are brief and targeted, mostly only lasting one AP (F7: Short interactions) and using commands more than questions (F1: Practical interactions). Systems can prepare for the large majority of interactions to be a single command-answer or command-action. Additionally, rather than rely on wake words for every interaction, users of the *Enterprise* computer often complete simple context-specific tasks with just a statement (F2: Context over wake words), for example on the turbolift ("Deck four") or at the replicator ("Tea, Earl Grey, Hot"). In these cases, physical proximity, possibly determined through computer vision or location tracking, can be used to replace the need of a direct wake word. Alternatives to wake words have been explored in recent works using gaze [33] and specifically with regards to children's interactions [11]. Designs should leverage multimodality to allow for the simple, quick tasks to be done without overhead or formalities. This does not need to have the level of location awareness seen in the *Enterprise*, which brings up important discussions on privacy, but could do much by using already available information such as the current WiFi network and Bluetooth connections, which are already leveraged for non-speech features such as disabling a phone's lock when at home.

Context from both the user and the setting provides key information to prepare for longer interactions. In this case, the device can keep actively listening, so the user will not need an additional wake word. For example, turning on the lights is likely a single interaction (e.g., an IoT statement), while looking for nearby restaurants is likely going to take several turns (e.g., several InfoSeek questions)

5.2.2 Use Context for Streamlined Interactions

The type of interaction and domain of use provide key context that clarifies intended use and can inform how the computer should respond. First, the structure of a user's query can indicate the intended domain for the interaction: questions are usually InfoSeek tasks and statements are mostly IoT tasks (F4: Grammar provides context). Interaction type also indicates the appropriate computer response with commands and statements receiving more actions and questions, more verbal responses (F5: Grammar infers response). The determined domain provides further indication to the response, Entertainment and IoT spaces mostly only need an action, while InfoSeek likely expects a spoken response (F6: Domain infers response).

This knowledge can be used to inform the desired outcome based on the interaction's structure. For example, when a user states "lights", with no more context given, a VUI can guess that the intended query is "turn on the lights in this room". The simple statement suggests an IoT or Entertainment task, all of which points to an action response. The question "Is my garage door open?" is likely an InfoSeek task, which likely expects a spoken answer, and not an IoT interaction such as the computer immediately opening the door.

Commands have no consistent primary domain, but are often used as the initial set up to be followed by either a question or statement, based on the domain. Based on this, VUIs can rely on other context clues to determine what the likely domain will be when given a command, and, from that, what type of answer is expected.

This type of context awareness is a major difference between the real and imagined interactions that is not severely held back by technical limitations. For example, smart speakers can be aware of the devices that share a room with them, allowing them to infer defaults for less detailed statements, such as "lights".

5.2.3 Leverage Multimodality

Much of what allows for the quick and easy use seen in this dataset is that the computer can respond by completing an action rather than continuing verbally. Design of VUIs should leverage task context and multimodality in order to provide visual or other non-verbal cues, as appropriate. For example, "Play 'Happy Birthday'" should simply play the music without further response, and "Dim lights" should not only dim the lights without context, but further, if there are multiple rooms with smart lights, should assume the user is referring to their current room (e.g., where this speaker is located).

Using the context provided is key to determining this, as seen in the previous recommendation, but multimodality goes beyond the type of response given by the computer. For example, when a person is using a holodeck and asks a question of the computer, the computer uses what is currently displayed and can guess that is the context of the query. Designs can use information available to other modalities to support more intuitive interactions. For example, if a phone screen is showing a restaurant menu and a person ask, "Are they open tomorrow?", the speech interaction can use the known information to complete this task without need for clarification.

5.2.4 It Is Not a Conversation

Despite the continuing push for human-like conversation between a user and a voice agent, our findings show that speech with the *Star Trek* computer, an idealized interaction, is not treated as a conversation by users. Two thirds of these interactions are a sole adjacent pair, and over half of those interactions have no verbal interaction from the computer. The computer's consistent use of actions to cue a successful interaction demonstrates the extent to which we do not expect a full conversation with the computer. In fact, the only space in which we see longer interactions and more conversation, is Analysis, the only domain not seen in modern data as it relies on futuristic computing power.

Rather than focusing on creating a truly conversational interaction, VUIs should expect most interactions to be functional and brief. Further, the domain of the initial query can be used to understand whether the interaction is likely to be longer and respond accordingly. The rise in user interest around home IoT tools (e.g., smart fridges) aligns with this pattern of brief, contextual use.

All of the previous design recommendations point towards this as well. Across the data, both user and system context are used to foster quick and simple interactions, and the computer avoids longer conversations unless there is a need, such as during problem solving. Ultimately, this shows us that our desire for true human-like conversation does not align with our expectations for what can be done with VUIs.

5.3 Discussion of Outliers

Speech interaction data is always messy and includes some unexpected interactions. These corner cases can represent important uses that are not generally considered by designers. Here we highlight two such important corner cases and explore how these are relevant to modern VUI use.

5.3.1 An Actual Conversation

As discussed in the Methods section, though 95% of interactions are under 10 turns, one has 34. This outlier comes from a plot in which Doctor Crusher is alone on the *Enterprise*, as the rest of the crew has seemingly vanished because of an experiment gone wrong in engineering [6]. As the chief medial officer, she is not expected to fly a ship or troubleshoot engineering mishaps, so she relies on a conversation with the computer to deduce the problem.

This extended back-and-forth is an extreme example of the predictable analysis and problem-solving interactions. This allows her to Think Aloud with a digital support to encourage her thought process. This specific use case is unsupported by modern VUIs, and should be further explored to understand how a virtual conversation partner could support users' thought processes in specific situations, such as during brainstorming.

5.3.2 Working with Children

Another episode, in which crewmembers had been accidentally changed into children, shows a brief interaction with a children's computer within the *Enterprise* system [12]. As mentioned earlier, this interaction was excluded from the dataset as it was so drastically different from interactions with the main computer. This computer uses a different voice than the main computer, and responds differently, providing a more conversational interaction and suggestions of what a user can do. For example:

Picard: Computer, display an internal schematic diagram.

Computer: I'm sorry, but I can't do that. Would you like to see some interesting animals?

We also see that the computer gives a verbal response even when an action is requested (and performed) in order to engage the user further:

Guinan: Computer, can you show me a picture of the inside of the *Enterprise*?

Computer: The *Enterprise* is a Galaxy-class starship. Do you know how to spell *Enterprise*? E - N - T - E…

Several previous works have explored how children are using Alexa and other VUIs [11,43,48], and how that differs from adult use. When Alexa added a politeness skill to respond to "thank you" with "you're welcome", they cited the motivation that children should have proper manners reinforced by these interactions [39]. Similarly, the *TNG* computer never apologizes to a user, but the children's computer does. Though, within the plot this is not an ideal interaction for the adults, the design still recognizes what these recent works have: the design of children's voice interfaces should be distinct from those for adults. The domains of use, allowed interactions, and use of language have little overlap and should not try to fit into the same system. This is further indicated to the users by using a recognizably different voice making it quick to recognize the change in system. Having a separate version of technology designed specifically for children is common, and speech design should consider how this should apply to this space.

5.4 Future Work

As discussed above, though Star Trek has long been a source of positive design innovation, due to the fictional source of our data, these recommendations must be explored thoroughly in practical settings before being broadly adopted. This is an initial review of this rich dataset and there is much more to be explored in future works. The full open dataset has

been made available[2]. We invite researchers to use this dataset to pursue further new insights into current, future, and futuristic voice interactions. In particular, the use of multimodality, presentation of negative interactions, error handling, and how users respond to unexpected interactions are important areas to be further explored in order to better understand the potential design implications for modern VUIs.

6 CONCLUSION

In this paper we seek out user design expectations for VUIs using the idealized and fictional voice-based interactions between people and the computer on the *Enterprise* from *Star Trek: The Next Generation*, based in the long history of science fiction influencing user expectations and technology design. As this show presents a consistently designed interface within a near-utopian future, we are able to create design recommendations that reveal how today's VUIs can be brought closer to this ideal. We find that the technology of modern tools is not significantly different from those seen in *TNG*, but the style of interactions, and their design, are mismatched.

Star Trek has been shown to both inspire technology's progress and influence users' expectations of how technology should work. We build on this and draw inspiration from *Star Trek* to inform VUI design to better match users' media-influenced expectations.

Actual conversation with the computer is rare. Brief, functional interactions are most prevalent, with the computer only verbally responded when necessary. User's physical and spoken context is heavily leveraged to determine, for example, the implied meaning in a simplified statement, allowing users to use simple statements without necessary overhead, such as multiple commands or excessive use of wake words. Multimodality is key to enabling these interactions through the simple leveraging of screens and other surrounding context. These recommendations can be designed for today, without the futuristic capabilities of the *Enterprise*. Modern VUIs must take these guiding ideas and move us towards our imagined future.

ACKNOWLEDGMENTS

This work was supported by AGE-WELL NCE Inc., a member of the Networks of Centres of Excellence (NCE), a Government of Canada program supporting research, networking, commercialization, knowledge mobilization and capacity building activities in technology and ageing to improve the quality of lives of Canadians. The authors also wish to acknowledge the sacred lands on which the University of Toronto operates. These lands are the traditional territories of the Huron-Wendat and Petun First Nations, the Seneca, the Haudenosaunee, and most recently, the Mississaugas of the Credit River. Today, the meeting place of Tkaronto is still the home to many Indigenous people from across Turtle Island, and we are grateful to have the opportunity to work in the community, on this territory.

REFERENCES

1. Joachim Allgaier. 2018. "Ready To Beam Up": Star Trek and its Interactions with Science, Research and Technology. In *Set Phasers to Teach!: Star Trek in Research and Teaching*, Stefan Rabitsch, Martin Gabriel, Wilfried Elmenreich and John N.A. Brown (eds.). Springer International Publishing, Cham, 83–93. https://doi.org/10.1007/978-3-319-73776-8_8
2. Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3: 1–28.
3. Isaac Asimov. 1950. *I, Robot*.

[2] http://www.speechinteraction.org/TNG/

4. Benett Axtell, Christine Murad, Benjamin R. Cowan, Cosmin Munteanu, Leigh Clark, and Phillip Doyle. 2018. Hey Computer, Can We Hit the Reset Button on Speech? In *Proc. of SIGCHI 2018 Extended Abstracts*.
5. Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.
6. Cliff Bole. 1990. Remember Me. *Star Trek: The Next Generation*.
7. Richard A. Bolt. 1980. *"Put-that-there": Voice and gesture at the graphics interface*. ACM.
8. Rob Bowman. 1988. Datalore. *Star Trek: The Next Generation*.
9. Rob Bowman. 1989. Manhunt. *Star Trek: The Next Generation*.
10. Sergey Brin and Denis Diderot. 2017. Building the Star Trek Computer. In *What Algorithms Want: Imagination in the Age of Computing*. 57.
11. Fabio Catania, Micol Spitale, Giulia Cosentino, and Franca Garzotto. 2020. What is the Best Action for Children to" Wake Up" and" Put to Sleep" a Conversational Agent? A Multi-Criteria Decision Analysis Approach. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–10.
12. José Coelho, Carlos Duarte, Pradipta Biswas, and Patrick Langdon. 2011. Developing Accessible TV Applications. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '11), 131–138. https://doi.org/10.1145/2049536.2049561
13. Eric Corbett and Astrid Weber. 2016. What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '16*: 72–82. https://doi.org/10.1145/2935334.2935386
14. Abide Coskun-Setirek and Sona Mardikyan. 2017. Understanding the Adoption of Voice Activated Personal Assistants. *Int. J. E-Services Mob. Appl.* 9, 3: 1–21. https://doi.org/10.4018/IJESMA.2017070101
15. Benjamin R. Cowan. 2014. Understanding speech and language interactions in HCI: The importance of theory-based human-human dialogue research. In *Designing speech and language interactions workshop, ACM conference on human factors in computing systems, CHI*.
16. Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '17), 43:1-43:12. https://doi.org/10.1145/3098279.3098539
17. Joshua Cuneo. 2011. "Hello Computer": The Interplay of Star Trek and Modern Computer. In *Science Fiction and Computing: Essays on Interlinked Domains*. McFarland.
18. Paul Dourish and Genevieve Bell. 2014. "Resistance is futile": reading science fiction alongside ubiquitous computing. *Personal and Ubiquitous Computing* 18, 4: 769–778. https://doi.org/10.1007/s00779-013-0678-7
19. Jens Edlund, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson. 2008. Towards human-like spoken dialogue systems. *Speech communication* 50, 8–9: 630–645.
20. Jonathon Frakes. 1992. The Quality of Life. *Star Trek: The Next Generation*.
21. Anne Galloway. 2013. Emergent Media Technologies, Speculation, Expectation, and Human/Nonhuman Relations. *Journal of Broadcasting & Electronic Media* 57, 1: 53–65. https://doi.org/10.1080/08838151.2012.761705
22. Emer Gilmartin, Marine Collery, Ketong Su, Yuyun Huang, Christy Elias, Benjamin R. Cowan, and Nick Campbell. 2017. Social talk: making conversation with people and machine. In *Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents*, 31–32.
23. Emer Gilmartin, Benjamin R. Cowan, Carl Vogel, and Nick Campbell. 2017. Exploring Multiparty Casual Talk for Social Human-Machine Dialogue. In *Speech and Computer* (Lecture Notes in Computer Science), 370–378. https://doi.org/10.1007/978-3-319-66429-3_36
24. Ido Guy. 2016. Searching by Talking: Analysis of Voice Queries on Mobile Web Search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (SIGIR '16), 35–44. https://doi.org/10.1145/2911451.2911525
25. Philipp Jordan and Brent Auernheimer. 2017. The fiction in computer science: a qualitative data analysis of the ACM digital library for traces of star trek. In *International Conference on Applied Human Factors and Ergonomics*, 508–520.
26. Philipp Kirschthaler, Martin Porcheron, and Joel E. Fischer. 2020. What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–9.
27. Stanley Kubrick. 1968. *2001: A Space Odyssey*.

28. Michael R. Levin. 2016. Amazon Echo - What We Know Now. *HuffPost*. Retrieved September 19, 2019 from https://www.huffpost.com/entry/amazon-echo--what-we-know_b_9802834
29. Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2018. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*: 0961000618759414. https://doi.org/10.1177/0961000618759414
30. Silvia Lovato and Anne Marie Piper. 2015. "Siri, is This You?": Understanding Young Children's Interactions with Voice Input Systems. In *Proceedings of the 14th International Conference on Interaction Design and Children* (IDC '15), 335–338. https://doi.org/10.1145/2771839.2771910
31. Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.
32. Teena Maddox. 2017. Tech leaders share how Star Trek inspired them to pursue a career in technology. *TechRepublic*. Retrieved September 20, 2019 from https://www.techrepublic.com/article/tech-leaders-share-how-star-trek-inspired-them-to-pursue-a-career-in-technology/
33. Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. 2019. Designing with Gaze: Tama -- a Gaze Activated Smart-Speaker. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 176:1-176:26. https://doi.org/10.1145/3359278
34. T. Daniel Midgley, Shelly Harrison, and Cara MacNish. 2006. Empirical Verification of Adjacency Pairs Using Dialogue Segmentation. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue* (SigDIAL '06), 104–108. Retrieved September 20, 2019 from http://dl.acm.org/citation.cfm?id=1654595.1654615
35. Christine Murad. 2019. Tools to Support Voice User Interface Design. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '19), 1–5. https://doi.org/10.1145/3338286.3344424
36. Jakob Nielsen. 2003. Voice Interfaces: Assessing the Potential. *Nielsen Norman Group*. Retrieved September 20, 2019 from https://www.nngroup.com/articles/voice-interfaces-assessing-the-potential/
37. Adam Nimoy. 1992. Rascals. *Star Trek: The Next Generation*.
38. Steve North. 2019. Imaginary Studies: A Science Fiction Autoethnography Concerning the Design, Implementation and Evaluation of a Fictional Quantitative Study to Evaluate the Umamimi Robotic Horse Ears. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, alt03.
39. Amy Packham. 2018. Amazon's Alexa Will Reward Kids For Saying "Please" And "Thank You" | HuffPost Canada. *HuffPost*. Retrieved September 20, 2019 from https://www.huffingtonpost.ca/entry/amazon-alexa-kids-please-thank-you_uk_5ae19610e4b02baed1b6dada
40. Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 640:1-640:12. https://doi.org/10.1145/3173574.3174214
41. Daniel M. Russell and Svetlana Yarosh. 2018. Can we look to science fiction for innovation in HCI? *interactions* 25, 2: 36–40.
42. Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8, 4: 289–327.
43. Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. Hey Alexa, What's Up?: A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868.
44. Petra-Maria Strauß, Holger Hoffmann, Wolfgang Minker, Heiko Neumann, Günther Palm, Stefan Scherer, Friedhelm Schwenker, Harald C. Traue, Welf Walter, and Ulrich Weidenbacher. 2006. Wizard-of-Oz Data Collection for Perception and Interaction in Multi-User Environments. In *LREC*, 2014–2017.
45. Aaron Suplizio, Cherian Abraham, and Ben Bajarin. 2016. Unpacking the Breakout Success of the Amazon Echo. *Experian*. Retrieved September 19, 2019 from https://www.experian.com/innovation/thought-leadership/amazon-echo-consumer-survey.jsp
46. Janice Y Tsai, Jofish Kaye, Tawfiq Ammari, and Abraham Wallin. 2018. Alexa, play some music: Categorization of Alexa Commands. *Voice-based Conversational UX Studies and Design Wokrshop at CHI*: 4.
47. Gokhan Tur, Dilek Hakkani-Tür, and Larry Heck. 2010. What is left to be understood in ATIS? In *2010 IEEE Spoken Language Technology Workshop*, 19–24.
48. Brenda K. Wiederhold. 2018. *"Alexa, Are You My Mom?" The Role of Artificial Intelligence in Child Development*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

49. Stacy L Wood. 2002. Future fantasies: a social change perspective of retailing in the 21st century☆ ☆Stacy L. Wood is assistant professor of marketing at The Moore School of Business, University of South Carolina, Columbia, SC, 29208; Phone (803) 777-4920; Email wood@darla.badm.sc.edu. The author would like to thank Bruce Money, Sherry Roberts, Danny Wadden, and Georgiana Craciun for their assistance with this research. *Journal of Retailing* 78, 1: 77–83. https://doi.org/10.1016/S0022-4359(01)00069-0
50. Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands Free - Care Free: Elderly People Taking Advantage of Speech-only Interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14), 203–206. https://doi.org/10.1145/2639189.2639251
51. Victor Zue, Stephanie Seneff, James R. Glass, Joseph Polifroni, Christine Pao, Timothy J. Hazen, and Lee Hetherington. 2000. JUPlTER: a telephone-based conversational interface for weather information. *IEEE Transactions on speech and audio processing* 8, 1: 85–96.
52. The hottest thing in technology is your voice - Technology & Science - CBC News. Retrieved February 1, 2018 from http://www.cbc.ca/news/technology/brunhuber-ces-voice-activated-1.4483912
53. Memory Alpha. Retrieved from https://memory-alpha.fandom.com/wiki/Portal:Main. Retrieved September 20, 2019 from https://memory-alpha.fandom.com/wiki/Portal:Main

A APPENDICES

Below are the long text descriptions for each figure

A.1 Long Text Descriptions of Figure 1

A pie chart showing the domains of the interactions. InfoSeek is most common with 26%, followed closely by Entertainment with 24%, then IoT with 21%, Analysis with 15%, and finally None with 8% and Other with 6%.

A.3 Long Text Descriptions of Figure 2

A bar chart showing use of wake words by people in AP1 and AP2-N, across domains and overall. Overall, wake words are used in 69% of AP1 compared to 26% of AP2-N. In Entertainment, wake words are used in 68% of AP1 compared to 33% of AP2-N. In IoT, wake words are used in 44% of AP1 compared to 19% of AP2-N. In InfoSeek, wake words are used in 87% of AP1 compared to 27% of AP2-N. In Analysis, wake words are used in 86% of AP1 compared to 16% of AP2-N.

A.4 Long Text Descriptions of Figure 3

A bar chart showing use of interaction types by people in AP1 and AP2-N, across domains and overall. Overall, commands are used in 65% of AP1 compared to 53% of AP2-N, questions are used in 10% of AP1 compared to 25% of AP2-N, and statements are used in 27% of AP1 compared to 21% of AP2-N. In Entertainment, commands are used in 79% of AP1 compared to 60% of AP2-N, questions are used in 5% of AP1 compared to 6% of AP2-N, and statements are used in 16% of AP1 compared to 34% of AP2-N. In IoT, commands are used in 44% of AP1 compared to 62% of AP2-N, questions are used in 3% of AP1 compared to 10% of AP2-N, and statements are used in 57% of AP1 compared to 29% of AP2-N. In InfoSeek, commands are used in 57% of AP1 compared to 38% of AP2-N, questions are used in 29% of AP1 compared to 49% of AP2-N, and statements are used in 14% of AP1 compared to 13% of AP2-N. In Analysis, commands are used in 75% of AP1 compared to 46% of AP2-N, questions are used in 15% of AP1 compared to 48% of AP2-N, and statements are used in 13% of AP1 compared to 6% of AP2-N.

A.5 Long Text Descriptions of Figure 4

A bar chart showing use of responses and actions by computers in AP1 and AP2-N, overall and in response to different interaction types. Overall, AP1 sees responses in 40% of interactions compared to actions at 56%, and AP2-N sees responses in 42% compared to actions at 36%. After commands, AP1 sees responses in 35% of interactions compared to actions at 56%, and AP2-N sees both responses and actions in 36%. After questions, AP1 sees responses in 80% of interactions compared to actions at 4%, and AP2-N sees responses in 91% compared to actions at 0%. After statements, AP1 sees responses in 20% of interactions compared to actions at 73%, and AP2-N sees responses in 31% compared to actions at 36%.

A.6 Long Text Descriptions of Figure 5

A bar chart showing use of responses and actions by computers in AP1 and AP2-N, across domains and overall. Overall, responses are used in 40% of AP1 compared to 42% of AP2-N and actions are used in 56% of AP1 compared to 36% of AP2-N. In Entertainment, responses are used in 25% of AP1 compared to 33% of AP2-N and actions are used in 71% of AP1 compared to 58% of AP2-N. In IoT, responses are used in 23% of AP1 compared to 53% of AP2-N and actions are used in 73% of AP1 compared to 41% of AP2-N. In InfoSeek, responses are used in 71% of AP1 compared to 79% of AP2-N and actions are used in 17% of AP1 compared to 16% of AP2-N. In Analysis, responses are used in 52% of AP1 compared to 72% of AP2-N and actions are used in 48% of AP1 compared to 21% of AP2-N.

A.7 Long Text Descriptions of Figure 6

A bar chart showing frequency of most common pairs of interaction types in AP1 and AP2-N. In AP1, Command to Action is most common with 36%, followed by Command to Response with 22%, then Statement to Action with 18%, and Question to Response with 8%. In AP2, Question to Response is most common with 23%, followed by Command to Response and Command to Action with 20% each, and Statement to Action with 7%.