CITE AS:
C. Murad, C. Munteanu, B. R. Cowan and L. Clark, "Revolution or Evolution? Speech Interaction and HCI Design Guidelines," in *IEEE Pervasive Computing*, vol. 18, no. 2, pp. 33-45, 1 April-June 2019, doi: 10.1109/MPRV.2019.2906991.

Revolution or Evolution? Speech Interaction and HCI Design Guidelines

Christine Murad
University of Toronto, Canada

Cosmin Munteanu
University of Toronto Mississauga, Canada

Benjamin R Cowan
University College Dublin, Ireland

Leigh Clark
University College Dublin, Ireland

The evolution of designing interactive interfaces has been rather incremental over the past few decades, largely focused on Graphical User Interfaces (GUIs), even as these extended from the desktop, to mobile or to wearables. Only recently can we engage in ubiquitous, ambient, and seamless interactions, as afforded by voice user interfaces (VUI) such as smart speakers. We posit here that recent speech engineering advances present an opportunity to revolutionize the design of interactions. Yet current design guidelines or heuristics are heavily oriented towards GUI interaction, and thus may not fully facilitate the design of VUIs. We survey current research revealing the challenges of applying GUI design principles to this space, as well as critique efforts to develop VUI-specific heuristics. We use these to argue that the path toward revolutionary new ubiquitous conversational voice interactions must be based on several evolutionary steps that build VUI heuristics off existing GUI design principles.

The design of interactive computing systems has experienced several evolutions over the past few decades. Yet what is striking is that, along with the evolution of device modalities and capabilities for multimodal interaction, we still conduct the majority of our day-to-day interactions under a one-to-one paradigm[1]. It is therefore unsurprising that, even if increasingly capable multimodal devices afford new interactions (from smartphones to smart speakers and to virtual reality), our approach to designing for these new interactions is still steeped into a mindset shaped by Graphical User Interfaces (GUI). Engineering advances are facilitating the evolution of these devices toward ubiquity and seamlessness; yet the design of interaction is not necessarily keeping up[2]. Particularly for speech modality, we have yet to design truly natural, ubiquitous interactions. The human voice can be a more natural and intuitive modality[3] (albeit with its own limitations, e.g, accessibility or social context). Yet it is not meaningfully employed when interacting with devices in a ubiquitous (almost "anytime, anywhere") way[2].

Researchers have long seen voice and multimodal interaction as fundamental to enabling device-transcendent forms of ubiquitous interaction, especially if this interaction is meant to be naturally conversational [4]. In fact, as argued by Roni Rosenfeld and colleagues, interacting with multiple "appliances" should become seamless and "universal" across devices, allowing interactions that are not focused on a single device ("ambient … rather than attentional")[2]. Jakob Nielsen, while skeptical at the time about the potential for speech to replace GUI interaction, offers a vision of voice as a way to enable seamless interactions with "information appliances" in a multimodal way [5]. Eric Chung and colleagues use the term "fluid" to describe such ubiquitous interactions [6].

Until recently, voice interfaces were mostly limited by engineering capabilities[3], resulting in voice being largely ignored as a modality by HCI research. Yet, new engineering advances in speech processing have reduced these limitations. However, interacting with voice-enabled devices (Google Home, Amazon Alexa, etc.) is still not seamless, reminiscent of issuing commands to control a terminal device, as under the paradigm of task-based dialogues[1]. Such design paradigms do not adequately support many emerging speech application areas (e.g. social companionship, fluid interactions with multiple embedded devices). Nor do they seem to avoid notorious usability issues, such as: lacking ability to interpret non-speech conversational cues (e.g. pauses)[7], users' lack of awareness about the utterances that can be articulated[3], difficulty in retaining information presented through a sequential-only audio channel[3], difficulty with back-and-forth navigation[7], etc.

Given voice interfaces' current state of usability, this proposed revolution in voice-based interactions may first require an evolution of the principles that guide us when designing such interactions. This has been seen in other domains. For instance, Chung and colleagues[6] proposed new design patterns in order to support their newly-envisioned ubiquitous fluid interactions. Similarly, Ulmer and Ishii propose approaches for the design of tangible interfaces[8]. We claim that, in order to truly leverage these new capabilities to help us revolutionize speech interactions, we should similarly focus our research efforts on developing new design guidelines. Very recent research has begun exploring the development of specific design heuristics – work such as that carried out by Wei and Landay[9], even if targeted at smart speakers, represent important first steps.

This raises the question of whether we, as interaction designers, have the right tools (and conceptually, the right frameworks), to revolutionize the design of voice interfaces. As such, we argue in this paper that:

1. The recent engineering advances in speech processing afford us the ability to create novel speech interfaces, representing advances in ubiquitous, fluid, conversational, or natural interactions
2. While our research efforts have been (justifiably so) extensively focused on the engineering of speech interfaces, envisioning new interactions has received less attention
3. Where efforts have been dedicated to designing speech interfaces, these have been largely informed by existing design paradigms for GUIs, producing less desirable outcomes from a usability perspective.

Based on these arguments, we claim that:

1. Rethinking speech interactions requires us to envision, develop, and validate new design guidelines specific to speech interactions (the "revolution" we hope for)
2. Speech interaction is markedly different from GUI interaction; however, GUI guidelines provide a useful foundation to begin with, and are even preferable to developing VUI-specific guidelines from scratch
3. For VUI-specific guidelines or heuristics to be effective, they must be adoptable by designers. This requires them to be grounded in familiar frameworks such as GUI heuristics (the "evolution" that charts the path toward a VUI design "revolution" and onto truly ubiquitous interactions)

In the next sections, we present a critical analysis that supports the above arguments, and a meta-study illustrating a possible path to supporting these claims.

BARRIERS TO CONVERSATIONAL VOICE INTERACTIONS

Work has been done in the area of conversational speech interfaces that is sensitive to or driven by the desire for a seamless and natural user experience. However, many outstanding issues still exist that prevent these designs from fully unlocking the potential afforded by current engineering advances.

For example, the human-like nature of the voice can cause a misalignment between people's perceptions of a VUI's abilities and its true capability[3]. This is exacerbated by such interactions being slower than GUIs, due to users navigating by means of voice alone - such as for display-less VUIs[3]. The learnability of conversational VUIs is another barrier, with users often left guessing what they can say, having to remember long lists of commands, or knowing how to recover from errors made by the device or by users[3]. Additionally, VUIs may introduce obvious accessibility barriers for some user groups (e.g. hearing loss, speech impediments), which need to be taken into consideration when designing interactions where speech is the primary modality.

Heuristic evaluation is a popular method for designing and assessing user interfaces, due to its simplicity, effectiveness, ability to be used early in the design process[10]. Many designers are either directly familiar with the high-level heuristic design principles, or with specific guidelines derived from these [10]. As such, from the variety of available design toolsets, we choose here to focus on heuristic principles as one theoretical construct required to improve VUI design.

THE ROAD TOWARDS NEW DESIGN HEURISTICS FOR SPEECH

As VUIs become ubiquitous in the market, the need for proper speech interaction design grows, thus requiring designers to adapt to this space. This may present a challenge for designers trained under a GUI-focused body of practice. We postulate that what is required are guidelines that designers can relate to, and that could be used for VUI design much in the same way as GUI heuristics are nowadays used (even if only conceptually). Such evolutionary, yet familiar heuristics may help designers transition to a new domain. For example, transitioning from GUIs to accessible GUIs posed challenges due to lacking knowledge of guidelines[11]. Transitioning to designing VUIs may be even more challenging, with Yankelovich and Lai reporting how designers feel lost when adapting to this new space[12]. Training may ease the transition, although it is time-consuming.

We thus raise several questions: do we, as a design community, currently possess the right tools (i.e. design guidelines) to significantly encapsulate issues in speech interactions? Do we need a revolutionary redesign of these, or can we evolve incrementally on current tools and practices? And finally, do current proposals for speech design guidelines represent the much-needed step forward? We answer these through several arguments, grounded in existing research:

1. Through a meta-analysis of speech interface design literature, we show how current GUI guidelines can be interpreted as being (indirectly) applicable to VUI design
2. We show that most of the VUI-specific (heuristic) design guidelines in existing literature are in fact closely aligned with existing GUI guidelines
3. We identify barriers in the applicability of these VUI-specific guidelines
4. We propose exploring established GUI design heuristics as a starting point for the development of new design heuristics for VUIs, and outline a proposal for such development

Existing Design Guidelines

Before we begin our analysis of the VUI/GUI design research space, we need to synthesize the available guidelines into unified sets.

The most notable sets of guidelines for graphical interfaces are those by Nielsen[13], Norman [4], and Shneiderman & Plaisant[14], used in heuristic evaluations of GUI usability. We have grouped these overlapping principles into ten guideline categories, as described in our previous research [15]:

G1. Visibility/Feedback of System Status

G2. Mapping Between System and Real World

G3. User Control and Freedom

G4. Consistency throughout the Interface

G5. Helping to Prevent User Errors

G6. Recognition Rather than Recall

G7. Flexibility and Efficiency of Use

G8. Minimalism in Design and Dialogue

G9. Allowing Users to Recognize and Recover from Errors

G10. Providing Help and Documentation

Earlier research on voice interfaces, conducted before these became widely available, has shown that GUI heuristics don't directly apply well to VUIs[16]. However, more recent empirical work, such as the usability inspections conducted by Whitenton[17] or the user studies by Budiu[7], suggest that some of the existing heuristics could be applied to certain aspects of VUIs and help us identify usability issues with speech interfaces. We thus investigated more broadly the applicability of GUI heuristics to the design of conversational voice interfaces. This investigation is detailed extensively in our prior work [15] and summarized in the next section.

FROM GUI TO VUI HEURISTICS

We conducted a meta-analysis identifying how issues in VUI research in leading HCI publications align with the guidelines identified. We systematically gather VUI usability issues, and then reflect on how GUI guidelines may apply to these. We also identify additional guidelines that may be required to fit with speech interaction.

Method

A comprehensive literature search was conducted across several bibliographic indices, as described in our recent work [1,15]. In brief, we have performed an extensive search using several academic bibliographic databases since 1980. Search terms were manually generated in a snowball sampling manner and manually reviewed by the authors and independently validated by speech

interface design experts. The search queries included terms (and their lexical variations) related to voice interfaces, such as human computer dialog, conversational interface/agent, interactive voice response system, intelligent personal assistant, Siri, Alexa. 21 papers in 9 leading HCI publications were collected[15]. We then used the 10 GUI design guideline categories from the previous section and conducted a critical analysis of the 21 papers, examining the extent to which each paper directly or indirectly aligned with any of these categories. The analysis was based on the cited authors' own discussions and observations of user evaluations or usability issues.

Insights

While an extensive synthesis is included in our recent study[15], we highlight here the insights we gained from our meta-analysis of VUI design research. Table 1 summarizes the themes emerging from these insights. In the next section, we use these themes as the foundation for analyzing VUI-specific guidelines.

Table 1. Themes emerging from the meta-analysis of HCI research on designing speech interfaces. The themes are clustered as challenges to the design of VUIs under either existing GUI guidelines (G1 to G10) or additional categories (N1 and N2). These additional categories are yet to be validated as guidelines; however, these emerged from the meta-analysis as candidates.

Guideline/ Category	*Number of Papers*	*Themes*
G1: Visibility/ Feedback	6	• Lack of visibility in when and how users could respond to speech interfaces (not knowing when to speak) • Difficult for users to know what their voice UI could do • Misinterpretation of speech recognition errors • Unsure if the system understood users and reliance on visual feedback for such confirmation
G2: Mapping	7	• Users have their own mental models of how a conversational VUI should reflect real-life interactions • Exposing systems' operating schemas helps users develop a more accurate mental model of the VUI interface
G3: Control and Freedom	6	• Frustration with the lack of control of the interface (feeling rushed, worried about missing parts of the interaction), especially when compared to other input interfaces • Lack of control leading to user performance issues (task completion, errors, etc.)
G4: Consistency	None	
G5: Preventing Errors	2	• Loss of trust in system if no mechanisms for preventing user errors • Actual speech recognition errors do not directly impair users' interaction
G6: Recognition over Recall	11	• High cognitive load to remember speech commands (especially in audio-only interfaces) • Load increases with the number of acceptable voice commands • Lack of guidance on how to structure speech when interacting with a VUI, resulting in many guesses and eventually abandonment of task
G7: Flexibility and Efficiency	2	• Text input interfaces evolved to include efficiency-focused affordances (e.g. shortcuts); this is lacking in speech interfaces although free-form speech may provide an alternative

G8: Minimalism	3	• Unclear how the number of available options facilitate navigation of VUI (the 7+/-2 may have a different "sweet spot" for VUIs)
G9: Recovering from Errors	9	• Correcting automatic speech recognition errors may lead to users' introducing more errors (e.g. speaking louder, attempting to correct recognized text with more speech) • Speech understating/communication errors are also present more frequently in VUIs than in GUIs • Lack of undo • Inability to edit issued commands
G10: Help and Documentation	3	• Usability of VUI increased through use of a tutorial, progressive help, or contextual help
N1: Transparency/ Privacy	4	• Concerns with the privacy of data gathered by the speech processing engine behind VUIs • Privacy concerns are increased when the interaction is in public
N2: Social Context	2	• Uncomfortable speaking (loudly) in public – departure from social norms

Discussion

As illustrated in the previous section, a number of issues identified in the speech interface literature seem to echo issues highlighted in current GUI guidelines. In particular, System Visibility/status (G1), Mapping between System and the Real World (G2), User Control and Freedom (G3), Recognition rather than Recall (G6), and Recognition and Recovering from Errors (G9) were all found in the literature. G6 was the most comprehensively covered – unsurprisingly since VUIs are often both displayless and present information using a single output modality (audio). G1 also received considerable coverage, likely due to the importance of guiding users during interaction with a VUI through audio channels only (lacking visual tools), where they may also need to remember large amounts of information such as possible commands. This is in contrast to GUIs where such information does not need to be displayed in a serial manner.

Surprisingly, Consistency throughout the Interface (G4) had little coverage in the literature. A possible explanation for this may be that for GUIs, being predominantly visual, it is easier to conceptually follow the principle of consistency during design. Its application to VUIs has not yet been widely investigated – the way we speak naturally in conversations can be rather inconsistent, which translates into users often not knowing what they can or cannot say[3], and it is unclear if a VUI that is highly predictable (so as to be consistent) may be perceived as less natural. This may become quite important as we look towards long-term relationship development between users and agents, as well as when the same agent is used across a number of devices.

Our meta-analysis revealed that some of the largest usability issues involve 1) the cognitive load required in interaction, 2) the need for users to have control over interaction and 3) the need to deal effectively with errors. These map into guidelines such as Recognition over Recall (G6), Control and Freedom (G3), and Recovering from Errors (G9). There is also an apparent usability challenge for users pertaining to Matching from System to Real World (G2), whereby the current metaphor of natural conversation as a model of speech interaction may be inappropriate. We also see that two new guidelines need to be considered in a speech context. These revolve around the need to ensure transparency and privacy (N1) and to consider the social context and how this may affect speech interaction (N2). It should be noted that these are only two potential VUI new design guidelines.

Recognition over Recall (G6) is brought up frequently when discussing VUI design issues. The ability to recall how to interact with a speech interface, using the device itself, provides affordances for more seamless and less segmented interaction. The principle of Mapping Between

System and Real World (G2), another well-discussed heuristic, may allow us to support multi-modal and ubiquitous speech interaction as we experience it naturally in daily interactions. Moreover, since free-form speech does not naturally follow a strict command-based format, a reasonable amount of freedom should be provided to the user (User Control and Freedom - G3).

The meta-study has also revealed new usability problems that are not currently encompassed by existing GUI heuristics. For example, interacting with a VUI in public requires the user to speak (possibly loudly) due to voice being the primary operating modality. This raises issues not encountered within GUIs, suggesting heuristics such as Privacy (N1) and Social Context (N2). These apply to users' discomfort when talking to a device in public (not aligned with social norms), lack of transparency regarding data perceived as private (speech), and the worry of publicly exposing private information when speaking through a device (more easily controllable through GUIs).

It is important to emphasize that the applicability and importance of these design guidelines may differ depending on the nature and context of interaction. Many of the papers reviewed in the meta-analysis discuss task-oriented interaction. Designing for prevention of user errors, for example, may be markedly different for more traditional question-answer exchanges than with social conversations. However, this foundational discussion of guidelines based primarily on task-based speech interactions presents the first step in creating further guidelines for varying spoken exchanges with machines.

Moving Forward

While heuristics have been developed in the past[9,18], they have taken a more top-down approach, by collecting VUI usability issues and using them to develop usability heuristics. Therefore, these heuristics are mostly usable by speech designers. As argued before, what is needed are guidelines that are grounded in a conceptual framework that a wider range of designers can use. GUIs represent such a framework, as evidenced by how various attempts at developing new frameworks, guidelines, or heuristics for new domains have anchored these efforts either explicitly or implicitly in GUI principles[6,8,19,20]. Our meta-analysis shows that even current discussions in speech HCI literature are implicitly grounded in existing GUI principles. We therefore propose to formally investigate adapting GUI heuristics to VUI design – small evolutionary steps toward a more revolutionary shift from GUIs to VUI with respect to both technological capabilities and user/consumer preferences.

THE SAME, ONLY DIFFERENT?

In comparison to GUI design principles, the space of speech-specific heuristics is significantly scarcer. The most relevant set of such heuristics has been proposed by Suhm [18] to assist designers of telephone-based dialogue systems (one of the earliest and most widespread application of voice interaction). The heuristics were generated from usability issues observed in telephone dialogue interfaces, along with design solutions for each of the issues. These solutions were then distilled into 10 design guidelines by experts.

More recently, as prompted by the emergence of consumer smart speakers, Wei and Landay [9] proposed a set of 17 VUI design guidelines that expand on previous efforts such as Suhm's [18]. These aimed to represent a significant departure from existing GUI guidelines. Experts evaluated the guidelines in order to explore what usability issues were identified for home-based voice assistants.

The above VUI guidelines are, in our assessment, the most comprehensive heuristics currently available, as well as covering some of the most common types of voice interactions (phone-based dialogue systems and smart speakers). As such, for the purpose of our analysis, we have merged them into a unified set as described below (Wei and Landay's heuristics are marked by the letter "A", with Suhm's by the letter "B"):

A1. Give the agent a persona through language, sounds, and other styles

A2. Make the system status clear

A3. Speak the user's language

A4. Start and stop conversations

A5. Pay attention to what the user said and respect the user's context

A6. Use spoken language characteristics

A7. Make conversation a back-and-forth exchange

A8. Adapt agent style to who users are, how they speak, and how they are feeling

A9. Guide users through a conversation so they are not easily lost

A10. Use responses to help users discover what is possible

A11. Keep feedback and prompts short

A12. Confirm input intelligently

A13. Use speech-recognition system confidence to drive feedback style

A14. Use multimodal feedback when available

A15. Avoid cascading correction errors

A16. Use normal language communication errors

A17. Allow users to exit from errors or a mistaken conversation

B1. Keep it simple

B2. Carefully control the amount of spoken output

B3. Word options the way users think

B4. Minimize acoustic confusability of vocabulary

B5. Provide carefully designed feedback

B6. Abide by natural turn-taking protocol

B7. Coach a little at a time

B8. Offer alternative input modalities

B9. Yes/no queries can be very robust

B10. Carefully select the appropriate persona

As discussed earlier, prior research is rather inconclusive with respect to applicability of GUI heuristics to VUIs, with older research suggesting a lack of applicability[16], while more recent research bringing (indirect) evidence that such a mapping may in fact be possible in certain cases[17]. This prompts us to investigate whether there is a concrete mapping of current GUI heuristics into the recent VUI guidelines developed by Suhm[18] and by Wei & Landay[9]. Based on the detailed descriptions provided in the papers where these VUI guidelines are introduced, we analyzed each of these guidelines, identified the usability and interface design situations it refers to, and considered whether these situations (specific to speech interfaces) would be covered under existing GUI guidelines. The results of this critical appraisal are captured in Table 2.

Table 2. Mapping of recently proposed VUI heuristics into "traditional" GUI design principles.

Guideline	**Description**	**Mapped VUI HEs**	**Explanation for Mapping**
G1: Visibility/Feedback of System Status	User Interfaces should make the system status visible, and provide informative feedback to the user	A2: Make the system status clear A11: Keep feedback and prompts short A12: Confirm input intelligently	The VUI heuristics mapped to this GUI Guideline advocate providing good feedback or exposing the system status of an interface clearly. In particular, all of the A* heuristics -

		A13: Use speech-recognition system confidence to drive feedback style A14: Use multimodal feedback when available B5: Provide carefully designed feedback	except A2 - are categorized under "Feedbacks and Prompts" by Wei & Landay[9]. A2's wording is almost a direct rephrasing of G1. Heuristics such as A12 and B5 show that for VUIs, it is not as simple as just displaying feedback for system actions. Much like Ullmer and Ishii[8] identify with Tangible User Interfaces (TUIs), both control of interface and representation of information are often coupled through the same channel (audio in the case of VUIs.)
G2: Mapping Between System and Real World	User interfaces should map symbols and controls from the system to the real world	A1: Give the agent a persona through language, sounds, and other styles A3: Speak the user's language A5: Pay attention to what the user said and respect the user's context A6: Use spoken language characteristics A7: Make conversation a back-and-forth exchange A8: Adapt agent style to who users are, how they speak, and how they are feeling A16: Use normal language in communication errors B3: Word options the way users think B6: Abide by natural turn-taking protocol B10: Carefully select the appropriate persona	The VUI heuristics mapped to this GUI Guideline all advocate the idea of matching natural spoken language characteristics in interface interaction, or the creation of a persona that match how people view human personas in real life. In particular, all of the B* heuristics are categorized as "spoken language" by Suhm [18]. As well, A6-A8 are categorized under "Conversational Style" by Wei & Landay. While A1, A3, and A5 are categorized under "General" by Wei & Landay, they advocate as well for mapping conversational interaction and personas to what humans are typically familiar with [9]. This includes the natural flow of conversation that maintains context – where interlocutors easily reference elements from the conversation's prior turns (A5).
G3: User Control and Freedom	User interfaces should give the user control over a system's actions	A4: Start and stop conversations B8: Offer alternative input modalities	The VUI heuristics mapped to this GUI Guideline all advocate for giving users control over both interaction and modality. Heuristics such as B8 show that a VUI may require multiple types of input methods to interact with the interface and the information along with audio - some of which can be tangible, such as gestures. The justification for this is similar to that discussed under G1 with respect to the coupling of control and information in the same channel – a characteristic common for TUIs[8] but also applicable to VUIs.
G4: Consistency throughout the Interface	Systems should strive for consistency by having similar actions cause similar outcomes in the interface	NONE	None of the VUI heuristics mapped to this GUI Guideline.

G5: Helping to Prevent User Errors	User interfaces should have error prevention mechanisms and constraints built in place to help users not to come across errors as they use the interface	A15: Avoid cascading correction errors B4: Minimize acoustic confusability of vocabulary B9: Yes/no queries can be very robust	The VUI heuristics mapped to this GUI Guideline advocate for helping users to not trigger errors through interacting with an interface. A15 advocates this by designing the interface so that it is difficult for a user to get into a situation where they would have cascading errors. B4 advocates for this by minimizing the amount of shared syllables between prompt options so that system would not misrecognize what the user said. B9 advocates implementing yes/no queries to help users from causing errors, as yes or no are hard for speech recognition systems to misinterpret or for users to say incorrectly.
G6: Recognition Rather than Recall	Users should be able to recognize user functions and options just through interaction, through affordances and visibility of system functionality	A10: Use responses to help users discover what is possible	A10 advocates for teaching users how they can use an interface through interacting with it. In particular, the interaction should be in a natural manner, vs. just stating what kinds of commands someone can do as a bulleted list.
G7: Flexibility and Efficiency	User interfaces should be flexible and promote efficient interaction (such as through providing shortcuts to perform familiar actions)	A11: Keep feedback and prompts short A14: Use multimodal feedback when available B1: Keep it simple B2: Carefully control the amount of spoken output B8: Offer alternative input modalities B9: Yes/no queries can be very robust	The VUI heuristics mapped to this GUI Guideline advocate for making the interaction as efficient and flexible as possible - either by not making interaction too long, or by making the type of input or feedback flexible. A14 and B8 illustrate how the flexibility principle is translated not only to VUIs but also to other types of non-GUIs. As Ullmer and Ishii[8] state, in GUIs, input and output are separated. However, much like for TUIs, VUIs may require multiple methods for both input interaction and representation of information – such as graphical, audio, gesture, etc. – in order to afford the same flexibility of interaction that is more naturally available in GUIs.
G8: Minimalism in Design and Dialogue	User interfaces should be designed to be minimalistic in their design and dialogue. Only necessary information should be provided, to reduce short-term memory load	A11: Keep feedback and prompts short B1: Keep it simple B2: Carefully control the amount of spoken output	The VUI heuristics mapped to this GUI Guideline advocate for minimizing the amount of feedback so that it doesn't overload people's cognition. It should be noted that all the VUI heuristics mapped to G8 overlap with G7, because "Efficiency" and "Minimalism" can be thought of as overlapping concepts.

G9: Allowing Users to Recognize and Recover from Errors	User interfaces should help users recognize and recover from errors, by providing simpler error handling and the ability to reverse actions	A15: Avoid cascading correction errors A16: Use normal language in communication errors A17: Allow users to exit from errors or a mistaken conversation B7: Coach a little at a time	The VUI heuristics mapped to this GUI Guideline advocate for helping to helping users understand and fix recognition and communication errors. In particular, all the A* heuristics were categorized under "Errors" by Wei & Landay.[9] B7 is also particularly described as an error recovery technique by Suhm[18]
G10: Providing Help and Documentation	User interfaces should provide assistance and documentation to the user when interacting with a speech interface to guide them through the interaction	A9: Guide users through a conversation so they are not easily lost A10: Use responses to help users discover what is possible B7: Coach a little at a time	The VUI heuristics mapped to this GUI Guideline all advocate guiding users and providing help through interaction. In particular, all the A* heuristics are categorized under "Guiding, Teaching, and Offering Help" by Wei & Landay[9].

The above table identifies overlaps between existing GUI guidelines (using the categories from our previous research[15]) and the preliminary VUI heuristics that have been proposed in literature. As can be seen, most of these VUI heuristics overlap with established GUI principles. Only G4 (Consistency throughout the Interface) does not overlap with a VUI heuristic. In addition, this table draws some parallels between the GUI-VUI mapping (e.g. G1, G3, G7) and similar considerations for other interfaces such as TUIs. Although papers such as Ullmer and Ishii's[8] do not explicitly propose actionable heuristics for such interfaces, these parallels provide additional perspectives and justification for extending GUI heuristics to VUIs (and others such as TUIs). These justifications are grounded in MCRpd (Model Control Representation Physical and Digital[8]) – a theoretical model proposed to explain how such non-graphical interfaces pose the challenge of having both the input and the output mapped into a single channel.

Some of the most notable GUI Guidelines in this table are G1, G2, and G7. These overlap extensively with the recently-proposed VUI heuristics[9,18], suggesting that these three GUI principles may be some of the most applicable ones to VUIs. This overlap also suggests that, although Wei & Landay[9] and Suhm's[18] heuristics were not explicitly derived from established GUI principles, there are strong theoretical and practical connections between these types of heuristics – even if the VUI-specific ones are more specific to their domain. Based on this, the central contribution of this paper is the argument that VUI heuristics need not be grounded in different theoretical frameworks than fundamental GUI principles. This is illustrated by the mappings in Table 1, showing that efforts to develop VUI heuristics "from scratch" may produce guidelines that still overlap with GUI principles.

In the next section we present a justification as to why deriving VUI heuristics from GUI principles may be preferable, and why this path is feasible.

The Trouble with Speech (Heuristics)

As shown in Wei and Landay's study[9], although their proposed heuristics are the only VUI ones (to our knowledge) to have been evaluated with design experts, this did not ensure unanimous acceptance. In particular, these heuristics did not fully resonate with designers not familiar with voice interfaces, who had difficulties using the VUI-specific heuristics, and identified significantly less usability issues than the speech experts did. This raises the question of what is the best path toward ensuring that designers incorporate VUI heuristics into their practice? This is particularly critical as designers, many not previously familiar with VUIs, may now be asked to consider such conversational interfaces (e.g. as evidenced by the multitude of Alexa skills promoted by many companies).

We have thus argued here that new heuristics for VUIs need to be based in something familiar to current usability experts. While directly applying GUI heuristics to voice interfaces may not be fruitful, GUI heuristics can serve as a baseline to aid usability experts in making the transition from GUI to VUI design.

As described earlier, any current methods of developing design approaches, from frameworks to design patterns and to design heuristics, have been inspired by or grounded in established methods for GUIs. Ullmer & Ishii[8] developed a new model derived from the well-established Model-View-Controller framework used for GUI interfaces, that emphasizes the coupling of representation and control of digital information in tangible interfaces. Chung et al.[6] proposed 45 design patterns for ubiquitous and fluid interfaces, meant to complement design heuristics. Other researchers have also used established GUI design heuristics (like Nielsen's[13]) to propose modality-specific heuristics – from mobile touch applications[19] to virtual reality[20]. Many designers are familiar with high-level design principles or with specific guidelines derived from GUI principles[10].

As research from Tangible Usable Interfaces to Ubiquitous User Interfaces shows, developing design approaches for specific modalities can be grounded in GUI heuristics. This is largely achieved by initially identifying the differences in usability of a specific modality vs. GUI, that requires an adjustment in existing design methods. Then, an existing design approach is taken – whether that be an existing framework, a set of principles, etc. – and particular changes are identified that must be made to allow adapting the GUI design method to a new modality. In this paper, we have followed a similar approach for VUI, by first exploring the usability and design issues that are discussed in current HCI speech literature. We then identified how these issues currently map into existing GUI heuristics (Table 1), and how GUI heuristics may be adjusted to map into a VUI design space. For the latter mapping, we conducted a critical analysis of how two of the most prominent VUI-specific heuristics[9,18] also align with existing GUI guidelines (Table 2), which revealed conceptual overlaps between the proposed VUI heuristics and established GUI heuristics.

This critical analysis suggests a promising path for current GUI guidelines, to at least inform the development of broader-reaching principles that will more radically transform the design of speech interactions. This approach may facilitate a smoother transition for designers from GUI to VUI – a challenge when designers not familiar with voice interfaces evaluated recently-proposed VUI guidelines[9].

TOWARD DESIGNING UBIQUITOUS SPEECH INTERFACES

As we have claimed earlier, speech represents an opportunity to truly revolutionize our interactions, by making them fluid, natural, conversational, and ubiquitous. The design heuristics developed over time are now serving as guidance for designers of GUIs – that is, these provide a rule-like framework that helps designers set targets with respect to (primarily) the usability of their designs. We have only recently begun reflecting on how such guidelines may be developed for voice interfaces. As we have illustrated in this paper, we have yet to embrace a widely-adopted set of guidelines that may help us move speech interaction closer to being seamless, device-independent, conversational, and ubiquitous. We have presented an argument supporting an evolutionary approach to developing such guidelines, by deriving and adapting established GUI heuristics to the VUI domain.

The analysis we have presented here also exposed the need for a more engaged reflection on not only the applicability of these design guidelines to speech-based interactions, but on broader considerations about what kind of design guidelines or rule frameworks may help us move speech interaction closer to being seamless, device-independent, conversational, and ubiquitous. While we have argued for grounding these considerations into establishing GUI heuristics, any VUI-specific guidelines emerging from these need to be empirically validated. We thus conclude this paper with a proposal for a three-phase approach to the development and validation of such guidelines:

1. Engagement of primarily non-speech design experts in usability walkthroughs of representative conversational voice interfaces (e.g. Alexa), combined with participatory workshops, leading to refinement of the heuristic guidelines listed in Table 1. These may be complemented by further refinement though usability analyses of sessions capturing users' interactions with VUIs. As argued earlier in the paper, non-speech designers may struggle with interpreting and employing VUI heuristics[9]. Together with the growing ubiquity of commercial VUIs - which may lead to many such designers being involved in VUI design - these represent arguments for focusing the development of VUI heuristics on non-speech designers.
2. Validation of the design heuristics in an experimental setting, using wizard-of-oz prototypes that will be crafted in order to evaluate and validate the importance of each specific design heuristic in isolation. Such prototypes may embody practical applications that currently make use of speech as an interaction technique, such as a modified voice-controlled music player. In addition to qualitative evaluations such as using cognitive walk-throughs, new speech heuristics can be validated through more quantifiable usability inspections. For example, this can be achieved by measuring how many usability problems are identified in crafted prototypes that employ any of the proposed heuristics, in comparison to an equivalent prototype (or design) that does not employ said heuristic – similar to Wei & Landay's[9] evaluation of their proposed VUI heuristics.
3. Exploration of the ecological validity of these heuristics in realistic settings, using complete and functional applications designed by experts following the new validated heuristics. The applications may cover different embodiments and modalities of conversational voice interfaces, such as mobile digital assistants, voice-only dedicated devices, hybrid voice/graphical display devices, etc. These applications may be tested using cognitive walkthroughs with usability experts or usability evaluations across different representative user samples. Employing fully-functional prototypes that are designed following the proposed heuristics also offers the opportunity to validate these heuristics in a quantitative manner, e.g. through controlled experiments employing user-centric performance metrics (e.g. task completion time, success rate, error rate).

CONCLUSION

At this point in time, we may be in the same situation mobile UIs were a decade ago or where website design was in the early 1990s. In order to take advantage of the capabilities of these speech devices, these issues must be explored and addressed, such as through the development of heuristics and design guidelines, dedicated to voice interfaces. We have presented here evidence from literature and an argument as to why grounding these in existing GUI guidelines may be the methodologically evolutionary step toward revolutionizing the design of voice interactions. Our hope is that by exploring established guidelines as a baseline, we will be in a position to identify and develop a taxonomy of design guidelines to assist in building more usable and intuitive speech interfaces.

BIOGRAPHIES

Christine Murad is a PhD student at the University of Toronto, researching the usability and design of conversational voice interfaces, particularly the development of design heuristics for conversational voice interactions.

Cosmin Munteanu is an Assistant Professor at University of Toronto Mississauga, investigating the human factors of interacting with information-rich media and intelligent technologies, such as speech interfaces, for several applications: mobile devices, mixed reality systems, and learning and assistive technologies for marginalized users.

Benjamin Cowan is an Assistant Professor at University College Dublin. His research lies at the juncture between psychology, human-computer interaction and speech technology,

investigating how the design of speech interfaces impacts user experience and user language choices in interaction.

Leigh Clark is a postdoctoral researcher at University College Dublin. His research focuses on the communicative and user experience aspects of human-computer interactions with speech technology, as well as understanding how linguistic theories can be applied in understanding these interactions.

REFERENCES

1. Clark, L. *et al.* The State of Speech in HCI: Trends, Themes and Challenges. *ArXiv181006828 Cs* (2018).
2. Rosenfeld, R., Olsen, D. & Rudnicky, A. Universal speech interfaces. *interactions* **8**, 34–44 (2001).
3. Cowan, B. R. *et al.* 'What can i help you with?': Infrequent Users' Experiences of Intelligent Personal Assistants. *Proc. 19th Int. Conf. Hum. Comput. Interact. Mob. Devices Serv. - MobileHCI 17* 1–12 (2017).
4. Norman, D. 1998. The design of everyday things. *Doubled Curr.*
5. Nielsen, J. 2003. Voice Interfaces: Assessing the Potential. *Nielsen Norman Group.* Available at: https://www.nngroup.com/articles/voice-interfaces-assessing-the-potential/.
6. Chung, E. S. *et al.* Development and Evaluation of Emerging Design Patterns for Ubiquitous Computing.
7. Budiu, R. & Laubheimer, P. Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri. *Nielsen Norman Group,* Available at: https://www.nngroup.com/articles/intelligent-assistant-usability/.
8. Ullmer, B. & Ishii, H. Emerging frameworks for tangible user interfaces. *IBM Syst. J.* **39**, 915–931 (2000).
9. Wei, Z. & Landay, J. A. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Comput.* **17**, 84–96 (2018).
10. Lodhi, A. Usability Heuristics as an assessment parameter: For performing Usability Testing. in *2010 2nd International Conference on Software Technology and Engineering* **2**, V2-256-V2-259 (2010).
11. Analysis of the ENABLED Web Developer Survey. (2005).
12. Yankelovich, N. Designing speech user interfaces. in *Proceedings of ACM CHI 98* 18–23 (1998).
13. Nielsen, J. Enhancing the explanatory power of usability heuristics. *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. Celebr. Interdepend. - CHI 94* 152–158 (1994).
14. Shneiderman, B. & Plaisant, C. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. **5th**, (2010).
15. Murad, C., Munteanu, C., Clark, L. & Cowan, B. R. Design guidelines for hands-free speech interaction. in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18* 269–276 (ACM Press, 2018).
16. Yankelovich, N., Levow, G.-A. & Marx, M. Designing SpeechActs. *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI 95* 369–376 (1995).
17. Whitenton, K. Voice Interaction UX: Brave New World...Same Old Story. *Nielsen Norman Group* (2016). Available at: https://www.nngroup.com/articles/voice-interaction-ux/.
18. Suhm, B. 2003. Towards Best Practices for Speech User Interface Design. 2217–2220.
19. Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C. & Rusu, V. Usability Heuristics for Touchscreen-based Mobile Devices. in *2012 Ninth International Conference on Information Technology - New Generations* 662–667 (IEEE, 2012).
20. Sutcliffe, A. & Gault, B. Heuristic evaluation of virtual reality applications. *Interact. Comput.* **16**, 831–849 (2004).