What’s The Talk on VUI Guidelines? A Meta-Analysis of Guidelines for Voice User Interface Design

Christine Murad  
University of Toronto  
Toronto, Canada  
christine.murad@mail.utoronto.ca

Heloisa Candello  
IBM Research  
San Paolo, Brazil  
heloisacandello@br.ibm.com

Cosmin Munteanu  
University of Waterloo  
Waterloo, Canada  
cosmin.munteanu@waterloo.ca

ABSTRACT
Over the past decade, voice user interface (VUI) design has been steadily growing, along with a growing VUI presence in consumer markets. However, there is currently a lack of widely-established guidelines for VUI design. While many sets of VUI guidelines have been proposed, they tend to be developed independently of each other, leading to a lack of consensus on appropriate guidelines for VUI design. This can hinder the wider adoption of practical VUI guidelines. To address this gap, we performed a large-scale meta-analysis of 336 VUI design guidelines that have been proposed in academic literature. Using thematic analysis, we present a unified and synthesized set of 14 guidelines, representing the most universally proposed principles of VUI design as captured by the 336 VUI guidelines identified in academic literature. We hope that this synthesized set can address several of the challenges to the adoption of VUI guidelines in design practice.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI); HCI design and evaluation methods; Interaction design; Interaction design process and methods;

KEYWORDS
Voice user interfaces, Design, User experience design, Speech interfaces, Design guidelines

ACM Reference Format:

1 INTRODUCTION
Over the past few decades, AI and speech technology has improved steadily, and more consumer devices have adopted voice interaction - such as smart speakers or smart IoT devices. While research on the design of Voice User Interfaces (VUIs) has increased greatly in the last decade, recent research suggests that commercial VUIs still suffer from usability issues that can deter adoption [24, 26, 62, 66, 72]. This may be due to the lack of foundational VUI design heuristics or guidelines [66, 68].

The field of Graphical User Interfaces (GUIs) has a collection of design guidelines that have been validated by the HCI community and are traditionally considered the “gold standard” [67, 75, 76]. These guidelines were the result of decades of study and refinement. However, we are still in the early days of establishing and validating VUI design heuristics, and many proposed VUI guidelines in literature are developed individually from each other, with little validation and lack of wide adoption [70, 71] - causing a lack of consensus on how to design for VUIs [70, 71]. While we can allow VUI guideline design to take its natural course as GUI guidelines did, the significant demand for VUIs requires a faster process. According to recent research by Murad et al [71] has shown, many GUI designers are now transitioning to VUIs and are finding that a lack of appropriately developed and validated guidelines are a large barrier to developing good VUIs.

To achieve this, we performed a large-scale meta-analysis on 336 design guidelines across 40 papers for voice-user interface design in academic literature. Using thematic analysis, we synthesized these guidelines to gather a unified representation of the numerous yet disparate guidelines that have been proposed for VUI design. From this meta-analysis, we present a synthesized set of 14 guidelines which represent the most accepted and validated aspects of VUI design proposed by guidelines in academic literature. In other fields (such as health sciences), similar problems in developing adoptable guidelines have been addressed through a meta-analytical approach to synthesizing and consolidating the disparate guidelines proposed in peer-reviewed papers into a cohesive set. We believe following this method of synthesis is the first step to consolidating a set of guidelines that can be practically used and refined for VUI design.

2 LITERATURE REVIEW
It has been suggested that the development and adoption of design guidelines is immediately necessary to help designers address VUI usability issues [66, 70, 71]. Below, we discuss relevant literature about the work on improving usability for VUIs, the work that has been currently done in developing guidelines for VUI design, and the adoption of guidelines in VUI industry.

2.1 Usability of Voice User Interfaces
While VUIs have grown in popularity over the past decade or two, commercially advertised “conversational” interfaces are still far from conversational. Conversational agents like Google Home and Amazon Echo employ command-based interaction that is usually learned through trial and error, rarely including functionality required for a realistic dialog. Yet users perceive these systems to have more human-like conversational abilities than they currently have [11, 26, 62, 65] not moving much farther from the capabilities...
of decades-old prototypes such as ELIZA [97]. Numerous usability challenges are still encountered [26], including difficulties recalling information [88], system feedback [26, 59, 62], recognition errors [26, 81] and learnability [39, 73, 102]. This causes users to often abandon VUls [26, 62, 66]. Many of these issues have been documented over the past 5 years, showing that VUI issues are still present, even with commercial VUIs existing for many years now. It has been suggested that having a widely-established set of VUI guidelines can help designers to address many of these issues [66, 70, 71].

2.2 Developing Design Guidelines for Voice User Interfaces

Several methods have been proposed for developing guidelines for a new domain. One method is exploring documented usability issues within a paradigm and generating guidelines that seek to resolve these issues. This method was used in creating video-game heuristics [80], and for telephone-dialogue heuristics [90]. Another method is to take established usability heuristics and adapt them to a new paradigm. These are often grounded in Nielsen’s [75] established heuristics for user interface design. This has been used for web pages [16], virtual reality applications [91], touch-screen mobile devices [49], and very recently for voice interfaces [57, 68]. The various methods add to the lack of consensus on designing VUI guidelines. Furthermore, many of these guidelines do not go through a validation process [70, 71], which can further hinder their adoption past paper publication.

2.3 Wider Adoption of VUI Design Guidelines

As the consumer market for VUIs has grown, so has the need for validated tools and practices for VUI design, with design guidelines being one such tool [36, 57, 70, 71, 96] (particularly in industry [69, 71]). This need is evident in major companies’ efforts to present their own guidelines, such as those from Amazon [2], Google [1], Apple [3], etc. Murad et al [71] found that, through a survey of over 100 industry designers, a lack of universal and appropriate design guidelines was one of the largest barriers to VUI design. However, guidelines proposed in literature rarely transfer over to practical industry design, despite their potential usefulness. Synthesizing these different guidelines could provide an avenue to improve validation and adoption for VUI design in industry. Similar work has been done by Branham & Roy [18], though they focused on synthesizing industry guidelines for VUI accessibility, while our focus is to perform an extensive meta-analysis on guidelines published in scholarly peer-reviewed venues, which has not been explored previously.

3 METHODS

We conducted a large-scale meta-analysis of over 40 scientific (peer-reviewed) papers that proposed a total of 336 VUI guidelines. In many fields, meta-analytic synthesis has been used to develop universal guidelines that help with the consistent application of domain-specific principles by a wide range of practitioners [41]. We used an adaptation of the PRISMA process to conduct our database search and select appropriate papers based on specific eligibility criteria. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which was first introduced by Liberati et al. [60] and then updated in 2021 by Page et al. [77], provides a 27-item checklist and a flow diagram for transparent reporting of systematic reviews. We followed Tuibn et al.’s [94] process for the database search, eligibility criteria, and paper selection process, which used PRISMA to perform a systematic review of assessment methods for conversational agents. For extracting and synthesizing guidelines, we followed the method used by Branham & Roy [18], who synthesized guidelines for VUI accessibility design in industry using inductive thematic analysis [19].

3.1 Database Search and Paper Selection

3.1.1 Eligibility Criteria. We followed a rigorous process for selecting papers for this meta-analysis, in order to conduct a consistent analysis and synthesis for each paper and guideline. We therefore chose to include only peer-reviewed academic papers, as it was necessary to follow a consistent synthesis process from design, to analysis, to article presentation. Due to the extensive variety and modalities of design guidelines for conversation and voice interfaces, we also narrowed our scope to focus on voice-first interfaces, to maintain consistency in analysis. This meant excluding papers that dealt with embodied agents, gestures, eye-tracking, avatars, chatbots, and multi-modal interaction (where voice was not the primary form of interaction). While guidelines for conversational interfaces in these mediums are important, the varying mediums require uniquely independent analysis for each, which is outside the scope of this paper. We also chose to focus on guidelines about designing voice-first interaction. This meant excluding papers that proposed primarily speech model, dialogue-only, or persona-only guidelines, as these do not focus on interaction design and require their own separate analysis. We excluded guidelines from textbooks explicitly in this study due to the largely different format in which they are presented vs. how an academic paper is presented, and since not all textbooks may be peer-reviewed under similar processes. The finalized formal criteria that were used are listed below.

**Inclusion Criteria**

- Papers proposing or reflecting on guidelines for designing and evaluating voice-first interaction were included.
- Papers explicitly listing voice or speech-based guidelines were included.
- Only peer-reviewed papers that were published in scholarly venues were included.

**Exclusion Criteria**

- Papers involving embodied conversational interfaces were not included.
- Papers involving gestures, eye-tracking, emotion/facial features were not included.
- Papers that did not offer an explicit list of voice or speech-based guidelines (e.g. providing vague “recommendations”) were not included.
- Papers involving the design or testing of speech models were not included.
- Papers discussing multi-modal interaction where speech was not the primary form of interaction were not included.
- Textbooks were not included.
· Papers discussing guidelines for primarily dialogue or persona design were not included (if none of the guidelines involved interaction design).

3.1.2 Selection Process. We queried five databases: ACM Digital Library, IEEE Digital Library, ProQuest, Scopus, and Web of Science. The first two were selected as they are the major databases for computing and technology-based papers, and the last three cover a larger range of scientific domains, so we could include a large range of scientific domains within our search. We derived the keywords directly from similar papers which performed meta-analyses on speech-based and conversational interfaces – the most relevant being a best paper from CHI 2019 which performed a large-scale survey on speech and voice interface papers in HCI [23] (see Table 1). We queried all peer-reviewed scholarly publications published up to June 2021 (no lower limit), and field of study was not restricted across the five databases. Keywords were queried within the Title, Abstract, and/or Author Keywords. Plurals and alternative spellings were included. Papers were filtered based on whether they contained the words “guideline”, “heuristic”, “principle”, or “design”.

Table 1: General Database Query

<table>
<thead>
<tr>
<th>General Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Speech interface OR voice user interface OR voice system OR speech-based OR voice-based OR speech-mediated OR voice-mediated OR human computer dialog OR human machine dialog OR natural language dialog system OR natural language interface OR conversational interface OR conversational agent OR conversational system OR conversational dialog system OR automated dialog system OR interactive voice response system OR spoken dialog system OR spoken human machine OR intelligent personal assistant ] AND [ guideline? OR heuristic? OR principle? OR design? ]</td>
</tr>
</tbody>
</table>

After querying, papers were imported into Zotero with general demographic metadata (title, author, year, conference/journal), along with the abstract and author keywords. Figure 1 demonstrates the flow of information during screening, eligibility, and selection process, with how many and which papers were removed at each step of the search.

The final 40 papers selected for the meta-analyses ranged across 28 conferences and journals (see Figure 2), ranging from 1995 to 2021. The highest number of papers were from 2020, and the most published to conference/journal was CHI. The papers ranged across many domains – from Human-Computer Interaction to Behavior and Information, to Intelligent User Interfaces, to even Dialogue and Signal Processing.

3.2 Guideline Extraction

The first author went through each selected paper and manually extracted each guideline. For each guideline, a guideline title and description were extracted. All attempts were made to preserve the original text during extraction. Due to varying paper formats, some guidelines did not contain a description, and some came with very long descriptions, as guidelines were presented as subsections of a paper with accompanying paragraphs of text as its “description”.

3.3 Coding, Thematic Analysis, and Synthesis of Guidelines

After extracting all guidelines, inductive thematic analysis was used to synthesize all proposed guidelines. Each guideline and its accompanying description were coded using open coding with NVivo 12. A subset of 50 randomized guidelines and accompanying descriptions were used as a reliability set, that was coded by the second author. There was 97% average coding agreement across these 50 guidelines. The rest were then coded by the first author. Both coders have extensive experience in the space of designing and researching VUIs, both from a theoretical and practical perspective, with one author being based in academia, and the second in industry within the space of VUI design. The open coding resulted in 233 codes across all 336 guidelines. Using thematic analysis, the open codes were then grouped together into 31 axial codes. These 31 codes were synthesized to 14 guidelines.

4 META-ANALYSIS FINDINGS

Figure 3 below illustrates the final synthesized guidelines, which were created by grouping the 31 axial codes (that were derived from the 233 open codes) into categories, and then creating a formalized guideline for that category.

We were interested in what the most discussed principles in VUI guidelines from academic literature were. We quantified the frequency of guidelines (out of the original 336) that each final synthesized guideline was composed of (see Table 2). We found the most discussed principles were around providing an interactive user experience with a large amount of user control (158 guidelines) and designing natural conversations that map to real-world norms.
and patterns (130 guidelines). Others of note were designing clear and informative feedback (99 guidelines) and error recognition and handling (72 guidelines). The least addressed principle was designing for accessibility and diversity (7 guidelines).

Next, we discuss the detailed findings of the thematic analysis. We discuss each guideline one by one, starting with the title, a brief description of the guideline, and a discussion of the thematic analysis findings for each axial code that made up said guideline (from Figure 3).

4.1 Design conversation based on the task domain of the voice interface.

*Description:* Conversations should be tailored to the task domain that the voice interface has been made for, and it should contain an extensive coverage of knowledge in that task domain.

#### 4.1.1 Tasks

Considering the task domain when designing interaction was very important. Being “capable of handling a wide range of topics in the task domain” [56] can improve usefulness and allow the system to manage different types of tasks, depending on user preference. At the same time, the system should have sufficient knowledge in the task domains that it deals with [13, 14, 31]: “Full task-domain coverage within specified limits is necessary in order to satisfy all relevant user needs in context. Otherwise, users will become frustrated …” [14]. If a task cannot be completed, the system should communicate the reason for failure immediately [101]. Inferring the task domain from users and considering user inferences can help: “Take into account possible (and possibly erroneous) user inferences by analogy from knowledge related task domain” [31].

4.2 Personalize the user experience to each user based on context.

*Description:* A VUI should use context and background knowledge from the user and the task domain in order to personalize the user experience for every user, and to allow users to personalize their own user experience.

#### 4.2.1 Context

Several guidelines noted the importance of establishing a common ground throughout the interaction and using this common ground as context to guide interaction with the user. One way to do this is through prompts [27], so that the user may also understand the language and capabilities of the system. This
4.3 Be transparent about system intelligence, capabilities, and constraints.

Description: A VUI should be transparent about what its capabilities and its constraints are. It should clearly state what it can, and cannot do, to properly manage a user’s expectations about the system’s intelligence and abilities.

4.3.1 System Expectations/Capabilities. The need for transparency of the capabilities and limitations of a VUI was often addressed. This can help establish appropriate mental models and expectations for the user throughout the interaction: “…anthropomorphism set unrealistic expectations that framed user perceptions of what constituted system failure… leading them to question the ‘intelligence’ of the system, indicating that user expectations of CAs should be scaffold through more considered revelation of system intelligence through design” [62]. A mismatch of expectations during interaction can lead to general frustration for the user [14], so presenting capabilities at the beginning of interaction is suggested. Using prompts to convey capabilities is one way that was suggested to achieve this [27], by prompting the user for actions they can perform with the system.

4.3.2 System Intelligence. Conveying whether the system has understood a request was also an important aspect. One way of doing this is to provide transparent feedback immediately when system recognition has failed [14], to prevent user confusion [27, 62]. This can help users understand the intelligence of the VUI [62], which can help users manage their own mental models and avoid unreasonable user expectations [62, 83]: “The system should not mislead the user into thinking that it is more intelligent or capable than it is, since that will only encourage the user to make requests that it can’t understand or accomplish” [83].

4.4 Break tasks up to allow for less cognitive load and better recall.

Description: A VUI should break up tasks into more manageable sub-tasks, to require less cognitive load, and allow users to recall information and functionality more easily. Menus should be short, and dialogue prompts should be minimal in nature.

can help prevent communication misunderstandings and failures [14]. As one guideline suggested: “We propose that future apps provide conversational scaffolding, which may adjust the conversation based on the child’s responses. For example, an app could rephrase the prompt with more accessible language if a child has difficulty responding to the prompt” [100].

4.2.2 Personalization. Guidelines also advised for VUIs to take users’ background knowledge and inferences into account, to tailor the conversation to user’s needs and capabilities [14]. Ways of doing this varied from learning from and adjusting conversation based on user responses [100], to using the task domain to establish a common ground [14]. As one noted: ‘Need for adjustment of system responses to users’ relevant background knowledge and inferences … to prevent the user from not understanding the system’s utterances or making unpredicted remarks such as, for example, questions of clarification, which the system cannot understand or answer” [14].

Once a common ground is established, guidelines advocated for voice systems to adapt to each user. Interaction should be personalized to users by learning about their behaviors [45, 101], their language [96], and their needs [56]: “Adapt agent style to who the user is, how they speak, and how they are feeling” [96]. Creating user profiles was one way to do this [56].

4.3 Be transparent about system intelligence, capabilities, and constraints.

Description: A VUI should be transparent about what its capabilities and its constraints are. It should clearly state what it can, and cannot do, to properly manage a user’s expectations about the system’s intelligence and abilities.
### 4.4.1 Cognitive Load/Resources.

Many guidelines emphasized breaking tasks up during interaction, to allow for better recall of previous information and reduce cognitive resources needed to perform tasks. Chunking tasks is one suggested way to do this [46]. Some suggest reducing the number of options shown at one time (4 to 9 options on average) [42, 46, 90], to make it easier for users to process, and to avoid memorization [87]. It should be easy to recall system information and functions “... through affordances and visibility of system functionality” [68]. Other guidelines advocate for creating minimalistic dialogue, to reduce short term memory load [68]. Voice interfaces require larger amounts of short-term and working memory than due to using speech as the primary medium of presenting information and have less visible cues [46]. Making prompts and system dialogues short [36, 46] is one way to reduce short-term memory load. “System dialogues should be short, straightforward and intuitive, incorporating familiar words and phrases” [36]. Guidelines note that users should not be required to remember large amounts of previous information: “System should not expect users to recall details from earlier parts of the conversation” [36].

### 4.5 Design clear and informative feedback.

**Description:** VUIs should design feedback to be clear and easy to understand, while also providing all the necessary information required from said feedback — such as when the system is processing information, when errors occurred, etc. Feedback should be provided quickly and efficiently to the user.

#### 4.5.1 Feedback.

Providing informative feedback to users throughout an interaction was often mentioned. Non-verbal feedback can also be assistive in user interaction [33, 85, 96]. Feedback should be provided with speed and efficiency [14, 38]: Immediate feedback on user commitments serves to remove users’ uncertainty as to what the system has understood and done in response to their utterances” [14]. The system should inform the user about any delays in feedback [38, 96]. This can be done through confirmations from the system about the information the user has provided or what the system is doing [36]. On system failure, clear feedback should be provided about what went wrong and why the system cannot complete a task [84, 101]. If user’s have conveyed information that the system cannot immediately process or breaks constraints, guidelines suggested using feedback in order to confirm whether the
system has appropriately understood what the user conveyed [47]: “Provide relevant feedback if analysis of user inputs suggests that these have violated recognition constraints (e.g., too loud, too quiet, too fast etc.).” [47].

4.5.2 Visibility. Along with this, guidelines advocated for always making the status of the system visible throughout user interaction. If the user is interacting with different “skills” or functions in the voice interface, the system should always make aware which function the user is using and interacting with [45]. As mentioned earlier, providing information upon system failure, and why the failure has occurred, is one way to do this [84, 101]. Notifying the user that it is “listening” or “processing” is vital as well [82, 83]: “If an operation takes more than a few seconds, the system should indicate that the operation is in progress. Ideally, the progress indication should be specific (e.g., printing...) rather than generic (e.g., working...) to indicate that the correct operation is in progress” [83]. All of this feedback should be unambiguous and transparent to the user [14, 48], as it is important to instill confidence that the information they have conveyed was processed by the system correctly or not [84].

4.6 Design appropriately for specific user and application types.

Description: A VUI should be designed based on the user that will be interacting with it, and the type of application it will be used in.

4.6.1 User Types. Designing for the type of user that will be interacting with the interface was heavily advocated for. The age of the user is one aspect that should be considered. Creating a voice interface with children will come with several specifications, such as allowing for more verbal engagement in interaction [100], through using open-ended prompts [99, 100]: “...relying too heavily on restrictive prompts fails to elicit children’s responses at the upper limit of their language competence and may result in responses that are semantically and/or syntactically simplistic.” [99]. Creating prompts and interactions that are less likely to break down can also help prevent frustration among children [99, 100]. On the other hand, older adults may require designing voice interfaces in a way that assists them in managing their mental models about what a voice interface is and how it interacts [55]. Another user aspect that was advocated for was novice vs. expert users [13, 14, 31, 62, 84] – particularly, the fact that expert users already have more knowledge about the functionality of voice interfaces: “...technically skilled participants were better able to see beyond artificial humanlike qualities to devise their own mental models of interaction” [62].

4.6.2 Application Types. Designing appropriately for the type of application with which the VUI to be used was also important. VUIs can be incorporated into many types of application settings. Among those mentioned in the guidelines we assessed were gaming [41], health [87], aviation [33], automotive [92], general crowdsourcing [46], and Interactive Voice Response systems [85]. Each application type requires its own considerations, and guidelines note to keep these application types in mind during design.

4.7 Protect the users’ privacy and security throughout conversational interaction.

Description: Interactions and information shared with a VUI should always be kept private, and security measures should be built throughout a conversational interaction (authorization, authentication, etc.).

4.7.1 Security. Guidelines coded here advocated for providing a secure interaction experience. One aspect that requires security is authentication of users that interact with the voice interface – authorizing the user before continuing the interaction or providing personal information is important [48]: “…imagine a guest asking for today’s appointments. Without authentication and authorization, the PVA will inform the guest about the owner’s appointments even though they may contain sensitive information such as doctor appointments.” [48]. Protection of privacy of personal information must always be maintained during interaction [56]. Providing transparency of privacy settings and functions is one way to do this and can help create trust between the user and the VUI [50, 54, 57]: “For example, “The only one who knows about your story is me. If you want me to forget your story, I can delete my memory.” This clarification will put teenagers at ease regarding the risk of the spread of rumors” [54].

4.7.2 Data. Allowing users access to view and manage their own data is an important aspect that guidelines note [50, 85, 101]. This can help improve trustworthiness with the application: “The system should convey trustworthiness by ensuring privacy of user data, and by being transparent and truthful with the user” [57]. Confirmations should be provided when performing functions that would edit the data: “Require the user to confirm destructive commands unless the entry is incomplete; destructive commands are those that result in deletion or erasure of user data” [85]. When providing non-personal data, being clear about where that data has been curated from can help reduce mental load: “Let me check weather information online. According to AccuWeather, tomorrow’s temperature will be...Do you want to hear more about how I instantly retrieved this information from AccuWeather?” [55].

4.8 Design conversational interaction that maps to real-world conversational norms and dialogue patterns.

Description: A VUI should use well-known and familiar conversational norms and dialogue patterns when interacting with a user and allow users to use natural speech and language to interact with it. A VUI should work to match realistic mental models of conversation, where appropriate (depending on the task domain).

4.8.1 Speech. Speech is a very important aspect in VUI interaction. Speech interaction can be a difficult form of interaction, as it comes with its share of expectations from users [25, 62] and involves managing different mental models [36, 68]. Using appropriate prosody [42, 56, 96] and low latency [87, 103] in system responses can help avoid user confusion and frustration: “Eleven participants were aware that a natural VUI should convey non-verbal meaning with the appropriate prosody, including intonations, pauses, and stress.” [56]. Clear pronunciation of words is also noted [35, 42, 44]. Being
able to control and tailor certain aspects of speech on the system is also suggested [44, 87]: “Users should be able to accelerate or decelerate the speed of the utterance and choose their preferred voice tone.” Some guidelines also advocate for using non-verbal forms of auditory feedbacks, such as beeps and chimes, throughout the interaction [33]. Controlling the amount of verbal output [90] can also help manage cognitive load.

4.8.2 Reality. One aspect of conversational interaction that many guidelines considered was realism of the interaction, and how interaction mapped to similar types of interaction in the real world. Matching existing mental models of conversation [36], and matching the user’s own language structure [57, 87] are important for improving interaction with voice interfaces – this can make interaction more natural: “VUI should be able to interpret the context of the ongoing conversation and answer in line with the user’s existing conversation models” [36]. Avoiding technical jargon and using more familiar terms and phrases can help improve natural interaction as well [96]. Incorporating norms such as greetings, compliments, etc., can improve the realism of the conversation [56]: “These words make the conversations appear ‘real-ish’ … and make the user feel their request is acknowledged” [56]. On the other hand, some guidelines argue against making conversation too realistic, particularly if the system’s use case is transactional [56]: “This is where the designers’ conceptions of naturalness in VUIs depart from what it means to be natural in human-to-human conversations. Our designers described natural human speech as often being indirect and inefficient, so these aspects of human conversation should be left out when designing for a natural VUI” [56]. As mentioned in G1, keeping in mind the task domain while incorporating realism is necessary.

4.8.3 Conversation. Conversation is at the center of VUI interaction, and a concept that many guidelines address. Some guidelines advocate for abiding by natural turn-taking protocols [90], while acknowledging users immediately once it is the system’s turn to speak [103]. The VUI’s responses should be brief in nature [14, 46]. Conversation should be used to guide the interaction [27]: “…what is needed are reliable techniques to guide the coordination of spoken output, and reliable techniques for using spoken outputs to help the user know what to say” [27]. When conversation breaks down, the VUI should attempt to repair said breakdown collaboratively with the user [56]. Incorporating questions into the conversation can help do this [53]. Questions can also help guide the interaction [99], establish common ground [27], or to help users understand the system’s request more easily [99]. Questions should be formulated in a similar way throughout the interaction [31].

4.8.4 Dialogue. Designing appropriate dialogue is a large part of designing useful conversational interaction in VUIs. Many guidelines advocate for designing a dialogue tree to classify responses from users and respond appropriately [99]. In dialogue trees, there should be a variation of responses [45]: “Skills are expected to provide several variations of opening prompts including one for first-time use, one for return and personalized prompts” [45]. Rephrasing prompts during interaction can help achieve this, especially if the user has difficulty understanding initial dialogue [45, 100]: “For example, an app could rephrase the prompt with more accessible language if a child has difficulty responding to the prompt” [100]. Being able to accept a variety of input responses from users can greatly improve dialogue [45]. As human language contains variations, so should the language that the system uses in its dialogue [56]. These variations should be familiar and easy to understand [36, 42, 96], and the type of language used may vary depending on user and application type [33, 83]. Allowing for interruptions on the user’s end throughout the dialogue tree is also suggested [27, 42, 83, 85]: “The user can interrupt the system, except in highest-priority situations. If the system is talking when the user speaks, it should stop.” [83]. On the other hand, the system should not interrupt the user while they are speaking [38, 83]: “Never interrupt the user while input is given… Even when the system is already knowing the answer to the currently asked question, it creates a better usability if the system waits for the user to finish the request” [38]. As mentioned earlier, the VUI system should provide short dialogue responses, to reduce cognitive and mental load on users [13, 36, 68].

4.9 Make conversation consistent and efficient across the user interaction experience.

Description: Conversations should be designed to be consistent across the entire interface. Prompts should use the same type of formulation and language across the interface, and similar commands should use similar language to invoke them. The VUI should also be efficient and minimalist in conveying information to the user.

4.9.1 Minimalism. Minimalism is heavily advocated for in VUI design and has been mentioned in several of the other previous sections. Presenting information in a short and concise way is essential to smooth VUI interaction [42, 101], and the system should only present what is necessary at time of the request [57, 68, 87]: “Dialogues should not contain information which is irrelevant or rarely needed. Provide interactional elements that are necessary to engage the user and fit within the goal of the system” [57]. Menus should only present a few options at a time, and then prompt the user if they want to hear more [57, 96]. This focus on minimalism is primarily for reducing cognitive load on the user [36, 45, 56]. If a large amount of information must be presented, one way to present it is to provide it via a different (graphical) platform [45]: “When this skill needs to tell users information that is not suitable through voice interaction, such as an URL, instead of saying it aloud, it sends the information to a user’s mobile app and explains to the user … ‘Please use the link we just sent to your app’. This practice eliminates the need for users to listen and remember long text.” [45].

4.9.2 Consistency. Consistency is often emphasized in the assessed guidelines. Using the same language, types of words, and terminologies throughout the entire interface was noted as important [45, 57]: “Users should not have to wonder whether different words, options, or actions mean the same thing. Follow platform conventions for the design of visual and interaction elements. Users should also be able to receive consistent responses even if they communicate the same function in multiple ways (and modalities)” [57]. Similar actions and commands should lead to similar outcomes [68, 83]. Many guidelines advocate for following conversational platform
conventions [36, 57, 87], though some note that due to the early adoption of commercial voice interfaces, platform conventions may not be fully available yet [87]. Using consistent voice output across the interface can also help users model their own verbal input to that of the interface, which can help prevent recognition failures and provide smoother interaction [104]: “Consistently worded output of an NLI system not only gives users a sense of familiarity, but also encourage users to model the output. A consistent pattern in user input would in turn simplify the task of the NL” [104].

4.10 Design a persona that evokes positive emotions with users.

Description: VUIs should provide a pleasant persona that provides a positive experience for users. The persona should use things like encouragement, engagement, and humor to evoke positive emotions with users, and help users feel more comfortable with being personal with the VUI.

4.10.1 Persona. The persona is one of the key aspects of a VUI, that must be designed with great care. Creating a persona that users can personally connect to was suggested, by exchanging greetings and kind words: “... VUIs can even make users feel as if they have personal connections to the applications by providing daily greetings or feedback on the user’s actions” [56]. Presenting the appropriate persona based on the task was also noted: “The tone of voice should match the application’s purpose to increase user trust and elicit proper user responses. For example … the VA in financial applications should sound serious so as to portray a reliable persona” [56]. Several guidelines advocated for having a persona that is polite, has proper social etiquette [93, 101], and accepts polite language from users [83]. Presenting a humorous persona can also help provide smoother interaction, especially with children [56, 62]. Finally, several guidelines encouraged designing an engaging persona [14, 30, 100]: “We recommend that more apps seek to promote a higher level of verbal engagement by incorporating open-ended prompts” [100].

4.10.2 Emotion. The type of emotion that a persona evokes during interaction is an aspect that many guidelines addressed. Some guidelines encouraged designing a persona that provides praise to the user: “All of us respond to praise, even when it isn’t warranted” [93]. Others encouraged the persona to be sympathetic to the user, to help soothe negative sentiment: “Ten participants mentioned the importance of providing sympathetic responses to users’ sentiments to maintain harmonious interactions…when the user experiences negative sentiments” [56]. Trying to match the user’s emotions was also suggested [96]. Expressing an interest in users can also help foster positive interactions [56]. Encouraging users throughout the interaction can help with this, particularly for certain user groups such as children: “Encouragement should provide sufficient scaffolding to draw children’s attention to the task at hand in an enjoyable way and help children participate in the dialogue. For example, apps could encourage a child to respond by using a more friendly tone” [99]. This can particularly help avoid frustration when there are recognition or communication failures [14, 45, 62, 100].

4.11 Help prevent, recognize, and correct errors.

Description: A VUI should work to prevent the user from encountering any errors, by providing appropriate help and documentation, and guiding the interaction to assist users in providing the correct input to be recognized. If an error occurs, the VUI should communicate the error quickly and efficiently, allowing the user to recognize what the error is, and allow the user mechanisms to correct errors.

4.11.1 Errors. There are several areas where errors must be handled by a VUI. The first, is preventing errors from happening during interaction [36, 68, 100]. Communicating system functionality [14] and potential limitations and errors that could happen [47] can help users adjust their interaction to help prevent them. Prompting for clarification in case of recognition or communication errors can also help prevent further breakdown [31, 84, 96]. The second is communicating when an error has been encountered [47, 87]: “Provide relevant feedback if analysis of user inputs suggests that these have violated recognition constraints (e.g., too loud, too quiet, too fast etc.)” [47]. Proper error communication can help prevent cascading errors [96]. Error feedback help users recognize what the error is so that they can correct it [45, 57], or provide a suggestion on how to fix it [14, 57, 90]: “Use examples in error/timeout reprompt, especially after open-ended prompts” [90]. Using simple, concise language is suggested [96, 104]. Finally, VUIs should provide functions for conversation repair. Providing the option to undo actions is one way to do this [37, 57, 87, 96]. The VUI should help initiate the repair process: “In case of system understanding failure, the system should initiate repair meta communication rather than leave the initiative with the user.” [14].

4.11.2 Help. A key aspect addressed in the guidelines were providing proper help and documentation to help promote proper user interaction. As mentioned earlier, communicating about how to interact with the system upfront can help avoid user frustration and interaction breakdown [13, 14, 57, 85]: “The system should guide the user throughout the dialogue by clarifying system capabilities. Help features should be easy to retrieve and search, focused on the user’s task, list concrete steps to be carried out, and not be too large” [57]. This can be done through an initial onboarding interaction: “When users dial into the interactive voice response system they should hear an opening message” [85]. After initial onboarding, the VUI should help guide users through the interaction [36, 68, 96] to help avoid users becoming lost [96]. Help should be accessible at any point of the interaction: “VUI should offer proactive help through guided on-boarding and contextual assistance. It should enable the user to easily access help whenever they need.” [36]. Having a “Help” that can be used at any time is one way to achieve this [45, 57]. This command should be clear and visible to the user [45, 57]. The VUI should always work on coaching the user on how to complete their task [25, 90].

4.12 Create diverse and accessible speech interaction.

Description: A VUI should be designed to be accessible to many types of users (those with hearing/mobility impairments, for example), and be usable by a diverse range of users.
4.12.1 Accessibility/Diversity. Accessibility is important to the design of any interface, and that is no different with VUIs. VUI interaction with deaf users is one thing guidelines addressed and creating interaction that is user-friendly for both hearing and deaf users [103]. Providing an alternative way to view messages (using multi-modal methods) is one way to do this [103]. Creating VUIs that are accessible to different populations, and on different platforms, was also recommended: "Compatibility of IPAs with a range of platforms/ applications and development of accessible IPAs for new user groups and new contexts" [50]. Finally, considering the needs of users who may have mobility impairments was important as well [25]: "Compatibility of IPAs with a range of platforms/ applications and development of accessible IPAs for new user groups and new contexts" [25]. This consideration can be important when implementing touch or gesture features for smart voice devices.

4.13 Provide open-ended prompts to establish common ground, and to request or clarify information.

Description: Open-ended prompts should be used to establish a common ground with the user, and to obtain background knowledge of the user. Open-ended prompts should also be used to request input from users, with more restrictive prompts being later used to clarify information.

4.13.1 Input. VUIs often must manage many types of (or lack of) input from users. Allowing for open answer inputs from users was generally recommended [46, 99, 100], however some guidelines warned of providing too much openness without much direction: “A lot of people make a mistake in the design by saying ‘Welcome way to communicate that the system has processed input properly” [83]. These can also help if the system detects a lack of input from the user: “Almost half of the apps we examined simply terminated the dialogue if a child did not respond, despite this type of breakdown being easily prevented by providing the child with additional opportunities to respond (e.g., repeating the question, nudging like ‘Try again’).” [99]. Open-ended prompts can also be used to encourage more engagement with users [100]. Clarification and confirmation prompts can be included after user input is detected, particularly if the input was presented in a way the system was not expecting: 

4.13.2 Prompts. As mentioned previously, using prompts is an important way to guide users on system interaction, and how input is processed [27]. Open-ended prompts can also be used to encourage follow-up restricted prompts resulted in the VUI’s higher rate of intent detection than open-ended prompts . . . restrictive questions reduced the likelihood of a child providing unanticipated answers” [99]. If there is a decent chance that the task will fail, a clarification prompt should be initiated [56, 92], especially before taking critical actions (such as modifying data) [104]: “. . . if the consequence of failing the task is considerable, a natural VUI should ask the user to confirm” [56]. The system should deliver prompts in a format similar to what the system expects from the user [27, 104]: “. . . the choice of what words to use at any point in a conversation is affected in part by how frequently and how recently the words have been used in the preceding conversation” [104]. Prompts should not be long in nature, to help reduce cognitive load [90, 96]. The VUI should allow for natural turn-taking [96], and avoid interrupting the user with a prompt in the middle of speaking [86, 96].

4.14 Provide an interactive and intuitive user experience with a large amount of user control.

Description: A VUI should be interactive and provide the user with as much control as possible over the user experience and its functions. A VUI should provide control over navigating through the interface as well.

4.14.1 User Control. Providing user control is an important aspect of good user experience. Control should be given to the user wherever possible [14, 68]. It should be easy to identify different areas of the interface and at what part of the interaction you are at [45, 101]: “Make it convenient for the user to discover and access relevant capabilities” [101]. It should be easy to interrupt the VUI when the user wants to make a choice [85] or return to the main menu at any point [42]. Allowing customizable commands was also suggested [101]. One way a user may be given control to customize actions is through creating shortcuts: “Allowing users to add tailored shortcuts of their choice, or identifying and calling the most frequently used content from each Skill, as well as across the Skills, can help.” [87]. VUIs should also introduce and encourage the usage of shortcuts developed into the VUI should they so choose [36].

4.14.2 Usability. Usability is a large concept to manage when developing any interface. Part of good usability is that the interface should be useful in the task domain it was designed for: “It is clear that the majority of users engage with the system only up to the point that it ceases to provide utility” [62, 103]. A system must also be easy to learn to use. Prompts and spoken output should help guide the user on where to go and how they can interact with the system [13, 14, 27, 48, 103]: “Use responses as a way to help users discover what is possible” [96]. The VUI should help reduce the gulf of execution [62, 100]. While these guidelines were focused on voice-based and voice-first interfaces, many guidelines suggested using alternative mediums of interaction to support usability where necessary [27, 30, 33, 44, 57, 84, 90, 96]: “Support flexible interactions depending on the use context by providing users with the appropriate (or preferred) input and output modality and hardware” [57].

4.14.3 Interaction. Designing good interaction is essential in a VUI. There should be clear communication from the VUI to the user in order to facilitate the interaction [13, 14, 99]: “Provide clear and comprehensible communication of what the system can and cannot do” [14]. Having proper opening and ending interactions was also
important, by having trigger words to start [45] and stop interaction [45, 83, 85, 96]: “To the extent possible, the user should be able to cancel a command or question in progress” [83]. A VUI should be able to listen to the user and show that it is listening: “[The VUI] can react to users by responding with “Uh-huh” or, “Really?” so that users feel as though there is someone there who is listening attentively.” [53].

4.14.4 User Experience. The user experience of a VUI must be tailored to facilitate a smooth experience [25, 62]. Users may come with different expectations of how the experience will look like [62, 93, 101]. Not managing these expectations may cause considerable user frustration: “Where users were not able to draw from a technical frame of reference, they tended to find blame in themselves, and often abandoned particular types of tasks, a behavior seen where systems present a gulf of execution” [62]. The user experience should directly help users achieve their workflow and goal [42, 62] and avoid confusing the user as much as possible [14, 27]: “The design commitment is to reduce the possibilities of evoking wrong associations in users, which in their turn may cause the users to adopt wrong courses of action or ask questions the system cannot understand” [14].

4.14.5 User Interface. While a voice-first interface does not have the same visual and aesthetic requirements a graphical interface would, many guidelines advocated for aesthetic design of voice output [50, 57, 87, 93, 103]: “While it doesn’t take much to elicit a social response, people are accustomed to high quality output.” [93]. This included having minimalistic dialogue [57, 87, 103]. The user interface should provide a pleasant experience, that adjusts itself as the interaction continues: “… at the beginning of a conversation, longer explanatory prompts will be tolerated, even expected, because that is what we typically do in conversations. And longer, more informative responses from the user can be expected. As the conversation progresses, the prompts should be shortened and briefer responses from the user expected.” [27].

5 DISCUSSION

In this paper, we aimed to synthesize and consolidate the vast amount of VUI guidelines proposed in literature into a set that represents the most common principles discussed in VUI design literature, to help promote their adoption in practical VUI design. Through this analysis we had several findings that can also help direct new VUI guidelines in the future.

We found that many of the principles that appeared in a higher frequency of guidelines were strongly similar to widely established GUI guidelines. This supports prior research suggesting that designers use previous experience from GUI design and map it onto VUI design [70, 71]. This connection suggests that GUI guidelines are already being used as a framework for developing new VUI guidelines, as has been previously suggested in research [68, 70, 98]. At the same time, there are challenges unique to VUI design that were identified - such as designing conversations, designing a persona, and managing cognitive resources when audio/speech is the primary mode of interaction. This supports prior research that has noted the differences between designing for VUIs and designing for GUIs [68, 70, 86, 102]. Much research has focused on solving localized usability issues that are specific to VUI-based interface design [25, 38, 39, 59, 73, 102], that we believe are also important to translate into VUI design guidelines (as several of our synthesized guidelines show).

Our findings also suggest that certain topics may not be currently addressed adequately enough in VUI guidelines that may require more focus as VUI design matures, such as making VUIs accessible and diverse [10, 18, 22, 25, 42, 74, 103], and maintaining user privacy [7, 32, 43, 58, 64, 95]. While they did not appear frequently in assessed guidelines, research shows these are growing concerns that should be translated into future VUI guidelines.

6 OPEN RESEARCH PROBLEMS AND FUTURE DIRECTION IN VUI GUIDELINE DEVELOPMENT

The resulting meta-analysis and synthesized guidelines present many open research problems that still require addressing and discussion. This analysis reveals the state of VUI guidelines discussions and proposals in academic literature as of very recently (April 2022), and we can see that even amongst recent work in academia, there are many issues that are still omitted or not discussed in great detail. One of the biggest ones is the lack of guidelines addressing accessibility and inclusivity. A lot of recent work has emphasized the importance of including accessibility and inclusivity considerations in the design process [8, 20, 34, 35, 40, 63]. However, based on our analysis, these concerns have not transferred over a great amount to actual guideline proposals – while there are guidelines that have been proposed, it was the smallest amount, with only 7 out of 336 guidelines in our analysis that addressed accessibility and inclusivity in some way. This would include cultural issues related to diversity and racism, of which recent research has discussed as an issue with current VUIs [22, 28, 51, 61]. Privacy is another concern that is of particular importance in the current digital climate, particularly with voice interfaces [7, 32, 43, 58, 64, 95], and yet also appeared at a low frequency (25 out of 336 guidelines). As VUI design has matured and the VUI design community is better understanding initial key interaction issues with voice user interfaces, extra focus is required for these areas where guideline proposals in literature may currently be low, but are issues that are key in the promotion of large-scale adoptability of VUIs in everyday consumer life.

It is worth noting that the guidelines presented in this analysis are primarily a synthesis of existing guidelines in academic literature and are not at a stage where they are finalized and validated for full practical usage. This work is an initial step to gather an understanding of the collection of knowledge and proposals of VUI guidelines that have already been presented in academic literature, that can be iterated on and reified by the VUI community, and eventually be adopted into practical VUI usage. As previous research has shown, what is currently out in the field may not be being used, and practical VUI designers find a lack of guidelines that they can use as one of the largest barriers to VUI design [71] – which may be surprising given the number of guidelines that were identified from this meta-analysis alone. This shows that their current state may not promote adoption and practical usage, whether it is a lack of usability or just a lack of knowledge of what has been proposed.
We believe this meta-analysis to be an intermediary step to the development of refined and validated guidelines that will be truly adoptable by the VUI design industry. Several aspects were not explored in this synthesis, due to the need to maintain a consistent data set and method of analysis, that we believe are important to be explored in future research. Our development of guidelines follows a very rigorous meta-synthesis process that is used across several fields (e.g. health sciences) where such rigor is necessary when synthesizing guidelines (e.g. health sciences). Such synthesis processes rely on peer-validated research, thus our decision to only include guidelines from peer-reviewed studies, following the similar process encountered in such other disciplines. We acknowledge that, in the field of interaction design, there are other sources of guidelines that may be employed in practice (e.g. books, tech blogs, industry-published guidelines). Given the breadth of peer-reviewed research we have captured, we consider that the guidelines we have synthesized are representative of a large majority of VUI design challenges.

This does mean that certain methods and sources of guidelines were omitted from this analysis. In future work, we plan to investigate other methods to incorporate guidelines from various sources (e.g. books) or other formats (e.g. personas). This synthesis can therefore be refined by incorporating additional, non-peer reviewed sources, including those developed in industry, from textbooks [79], to industry design docs [1–5], to further unify our understanding of how to design usable VUIs. Given the wide range of approaches, peer review, and formal validation under which industry guidelines may be developed, criteria must first be established on how to include such guidelines in future analyses. This would also involve recruiting designers in industry to be a key part of the refining and validation process.

This also meant that certain modalities were also omitted from this analysis, due to the complexity that adding more types of modalities adds to the analysis, development, and interpretation of guidelines. These should also be considered as well in future work, such as multi-modal interfaces [4, 8, 12, 17, 65], embodied voice agents [5, 6, 9, 21], gesture interfaces [5, 15, 29, 89], etc. – and it may be that guidelines may need to be specific to each modality. While for similar reasons of controlling for type of voice interface to allow for rigorous analysis, we focused on voice-first interfaces in this meta-analysis - as voice-first interfaces currently dominate the consumer market and require immediate focus – other modalities should not be forgotten and would be an immediate next step in this line of research. Due to the complexity and varying natures of interaction that other modalities add to interaction, there’s an argument to be made that we will need several sets of guidelines particularly tailored to different modality spaces, and even for multi-party and ubiquitous interaction – this is something that also requires further research.

7 CONCLUSION

This paper presented a thematic analysis synthesizing 336 VUI design guidelines proposed across 40 papers, presenting a unified set of 14 guidelines, representing the most commonly discussed principles VUI academic design literature. We believe that conducting this meta-analysis was a necessary first step to developing a consistent synthesized set of VUI heuristics that can be used in voice-first interfaces. We can also further assess how the broader VUI design community engages with the 14 unified guidelines presented in this paper. We hope that this work can help address the lack of consistency in VUI design guideline literature and help improve the adoption of future VUI guidelines.

ACKNOWLEDGMENTS

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). This work is also 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.

REFERENCES

What’s The Talk on VUI Guidelines? A Meta-Analysis of Guidelines for Voice User Interface Design
CUI ’23, July 19–21, 2023, Eindhoven, Netherlands


