Design guidelines for hands-free speech interaction

Conference Paper · September 2018

DOI: 10.1145/3236112.3236149

CITATIONS
7

READS
929

4 authors, including:

Christine Murad
University of Toronto
7 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Leigh Clark
Swansea University
25 PUBLICATIONS **79** CITATIONS
SEE PROFILE

Design Guidelines for Hands-Free Speech Interaction

Christine Murad
University of Toronto, TAGlab,
Department of Computer Science,
Toronto, ON, Canada
cmurad@taglab.ca

Cosmin Munteanu
University of Toronto, TAGlab and
University of Toronto Mississauga, ICCIT
Toronto, ON, Canada
cosmin@taglab.ca

Leigh Clark
School of Information &
Communication Studies,
University College Dublin
Dublin, Ireland
leigh.clark@ucd.ie

Benjamin R. Cowan
School of Information &
Communication Studies,
University College Dublin
Dublin, Ireland
Benjamin.cowan@ucd.ie

MobileHCI '18 Adjunct, September 3–6, 2018, Barcelona, Spain

ACM ISBN 978-1-4503-5941-2/18/09.
https://doi.org/10.1145/3236112.3236149

Abstract
As research on speech interfaces continues to grow in the field of HCI, there is a need to develop design guidelines that help solve usability and learnability issues that exist in hands-free speech interfaces. While several sets of established guidelines for GUIs exist, an equivalent set of principles for speech interfaces does not exist. This is critical as speech interfaces are so widely used in a mobile context, which in itself evolved with respect to design guidelines as the field matured. We explore design guidelines for GUIs and analyze how these are applicable to speech interfaces. For this we identified 21 papers that reflect on the challenges of designing (predominantly mobile) voice interfaces. We present an investigation of how GUI design principles apply to such hands-free interfaces. We discuss how this can serve as the foundation for a taxonomy of design guidelines for hands-free speech interfaces.

Author Keywords
Speech interaction; voice user interfaces; conversational interfaces; design guidelines.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces;

Conference/Journal	# of Papers
CHI	8
Communications of the ACM	1
CSCW	1
Mobile HCI	2
Behaviour and Information Technology	3
Interacting with Computers	2
International Journal of Human Computer Interaction	1
International Journal of Human-Computer Studies	2
Universal Access in the Information Society	1

Table 1: Number of papers from each conference/journal.

Introduction

Speech interaction is an increasingly popular field in the world of HCI. Speech has been described as a natural way of interacting with technology [12,15], and many devices are being released that advertise speech as a primary mode of interaction. However, there are still many usability issues with current commercial speech interfaces, such as Siri, Google Assistant, and Amazon Alexa [7,15]. There is a growing interest in Conversational Voice UIs, as they can truly enable people to be mobile and hands-free [15].

There are many currently established user interface design guidelines for graphical interfaces [17,19,25]. However, we do not currently have an equivalent set of principles for speech interfaces grounded in empirical and theoretical research, that have also been extensively validated and accepted by the community at large. Previous attempts to directlymap graphical user interface interactions to speech interfaces have not been successful [23,29].

Previous literature had expressed the need for design principles in speech interfaces. Suhm [26] created a database of "solutions" to existing design problems in telephone dialogue interfaces, and created 10 design guidelines based on the solution database. While his guidelines are more centered around the speech technology and production, it shows the need for creating principles to design better speech interfaces. Nielsen [18] developed design recommendations for touch-based interfaces such as the iphone, which was later reflected on, showing improvement in touch-based mobile interfaces [3]. By developing design principles for speech interfaces in a similar way, we can start working towards the better designing user interaction for these mobile voice interfaces, which can continue to be developed and built on.

In this work, we explore how discussions of design in current speech HCI literature align with existing design guidelines for graphical interfaces. We first describe a set of 10 design guideline categories based on three notable sets of established guidelines for graphical interfaces by Nielsen [17], Norman [19], and Shneiderman [25]. We then review current speech HCI literature in order to: 1) examine how speech HCI literature reflects on these existing design guidelines (either directly or indirectly), and 2) uncover design issues in speech interfaces that are not covered by existing design guidelines, which can serve as possible new design principles for speech interfaces.

Exploring Existing Design Guidelines

The most notable sets of guidelines for graphical interfaces are those by Nielsen [17], Norman [19], and Shneiderman [25]. These guidelines are frequently used to evaluate the usability of graphical interfaces, particularly when conducting heuristic evaluations. The principles proposed have a distinctive amount of overlap between them. We discuss these guidelines, and synthesize them into ten guideline categories:

G1: Visibility/Feedback of System Status: User interfaces should make the system status visible, and provide informative feedback to the user [17,19,25].

G*2: Mapping Between System and Real World*: User interfaces should map symbols and controls from the system to the real world [17,19].

G3: User Control and Freedom: User interfaces should give the user control over a system's actions [17,19].

G4: Consistency throughout the Interface: Systems should strive for consistency by having similar actions cause similar outcomes in the interface [17,19,25].

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 [17,25].

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 [17,19].

G7: Flexibility and Efficiency of Use: User interfaces should be flexible and promote efficient interaction (such as through providing shortcuts to perform familiar actions) [17,25].

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 [17,25].

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 [17,25].

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 [17].

These guidelines are commonly used in HCI for GUIs, yet it is unclear how these fit with the design of speech interfaces. Below we conduct a literature review to see how these align with speech interfaces and identify the need for further guidelines in a speech context.

Method

We collected 21 papers composed of full papers published in 9 leading HCI conferences and journals. Publication venues were gathered using sources listed in Google Scholar, Thomson Reuters and Scimago rankings (see Table 1). The ACM Digital Library, ProQuest and Scopus databases were searched for relevant publications using the terms in Table 2, generated from keywords from existing VUI and speech literature and from a survey of 11 leading researchers in VUIs and speech technology. Only papers that primarily investigated speech input, output and/or dialogue alongside aspects of design and usability between 1980-2018 were included. Papers not written in English, not peer-reviewed or discussing embodied interfaces or robots were excluded.

We then used the 10 design guideline categories from the previous section and examined the extent to which each paper directly or indirectly aligned with any of these categories. The method for categorizing papers was based on the paper authors' own discussions and observations of user evaluations (if the paper proposed and evaluated a new system), or on the discussion of current usability issues in speech interfaces (if the paper was a review/position paper). If the paper discussed particular positive or negative usability aspects or system implementations, the paper was included in said category. By usability aspects, we are referring to the core principles as captured in the foundational HCI texts by Nielsen [17], Norman [19], and Shneiderman [25], as aggregated in the previous section. Many papers fit into multiple categories.

Findings

An overview of the papers in each category is shown in

Review Search Terms
Speech interface; voice user interface; voice system; human computer dialog*; human machine dialog*; natural language dialog* system; natural language interface; conversational interface; conversational agent; conversational system; conversational dialog* system; automated dialog* system; interactive voice response system; spoken dialog* system; spoken human machine interaction; human system dialog*; intelligent personal assistant; Siri; Alexa; Cortana

Table 2: Review search terms. Asterisks (*) denote truncation to account for alternative spellings e.g. *dialog* or *dialogue*.

Guideline Category	Papers
G1: Visibility/ Feedback	[2, 5, 12, 16, 22, 29]
G2: Mapping	[2, 6, 10, 12, 22, 28, 29]
G3: Control and Freedom	[2, 4, 11, 21, 28, 30]
G4: Consistency	None
G5: Preventing Errors	[16, 29]
G6: Recognition over Recall	[1, 4, 8, 9, 13, 16, 24, 27, 28, 29, 30]
G7: Flexibility and Efficiency	[2, 13]
G8: Minimalism	[9, 28, 29]
G9: Recovering from Errors	[2, 5, 16, 20, 22, 24, 28, 29, 30]
G10: Help and Documentation	[4, 29, 30]
A1: Transparency/ Privacy	[2, 5, 14, 30]
A2: Context	[2, 6]

Table 3: Papers in each guideline (G*) and additional (A*) category (by citation number).

Table 3. Below, we discuss the papers and occurring themes that we observed for the ten categories.

G1: Visibility/Feedback of System Status: Six papers reflected on the visibility and feedback of a speech interface's actions. One theme was the lack of visibility in when and how users could respond to speech interfaces. Yankelovich et al. [29] found that users often didn't know if it was their turn to speak. Luger & Sellen [12] also noted that a lack of system feedback and transparency made it difficult for users to know what their voice UI could do. Another issue was a lack of feedback of how user speech input was processed [12]. Meyers et al. [16] found that users misinterpreted recognition errors from speech interfaces due to a lack of feedback. Begany et al. [2] reported that users often did not know if speech interfaces understood their spoken input, while Porcheron et al. [22] and Cowan et al. [5] found that people still relied on visual feedback from voice UIs in interpreting their input as being understood.

G2: Mapping Between System and Real World: Seven papers reflected on how users mapped their perception of speech interfaces to things in the real world. Five of the papers discussed how users map their own mental models of conversational interaction to speech interfaces [10,12,22,28,29]. Another theme discussed was the use of schemas to improve ease of use and understanding of a speech interface. Howell et al. [6] and Wolters et al. [28] found that giving users the schema of a familiar interaction in the world increased task performance and ease of use. Begany et al. [2] discussed that unfamiliarity with the interaction style can decrease usability.

G3: User Control and Freedom: Six papers mentioned the need for user control and freedom in speech interfaces. A common theme was frustration with the lack of control of the interface. Corbett & Weber [4] found that users felt rushed when interacting with a voice UI, and worried about missing parts of the interaction. Both Begany et al. [2] and Limerick et al. [11] found that interacting with a keyboard interface made users feel more in control than when using voice. Wolters et al. [28] expressed a more mixed view, finding that older users sometimes prefer to take the initiative, but at other times are content with letting the system be in control. Conversely, providing users with control over interaction can improve performance and user satisfaction. Perugini et al. [21] implemented an "out-of-turn interaction" system, which allowed users to provide information that would be needed later in dialogue, which reduced task completion time and was more usable. Zajicek et al. [30] allowed older adults to interrupt dialogue by saying "help" or "home", letting them take control of the interaction.

G4: Consistency throughout the Interface: No papers that we examined discussed consistency throughout the interface. This may be due to speech interfaces being a developing field, and no understanding of what consistency would entail in a speech interface currently exists.

G5: Preventing User Errors: Two papers discussed the need to prevent user errors in user interface interaction. Yankelovich et al. [29] state the importance to design to prevent user errors to increase trust in the interface. However, Meyers et al. [16] noted that while errors with natural language processing (NLP) were

very common, they did not always impair a user's interaction with the speech interface.

G6: Recognition Rather than Recall: Eleven papers reflected on the need to be able to recognize information or actions for interaction. A common theme is the level of cognitive load required to remember speech commands. Both Shneiderman [24] and Aylett et al. [1] note that using audio as the only output modality increases cognitive load, and requires users to remember long pieces of information. Four papers found that providing many options to a user increases the struggle of recalling them [9,13,28,30]. Knutsen et al. [8] report that information presented by a speech interface can be more difficult to recall than information a user has presented themselves. Another theme is the difficulty in recognizing how to interact with a speech interface. Meyers et al. [16] and Yankelovich et al. [29] discuss that users often are not aware of how to structure their speech to a voice UI. Corbett & Weber [4] found that users often make attempts at guessing what they can say. Wilke et al. [27] found that some participants did not complete a particular task because they couldn't understand or remember how to.

G7: Flexibility and Efficiency of Use: Two papers highlighted the need for user interfaces to be flexible and efficient. Molnar & Keltke [13] highlight that a lack of flexibility causes a reduction in productivity and satisfaction with the interface. Begany et al. [2], when comparing a textual vs. speech search interface, found that the use of keyboard shortcuts are a useful feature that improves task efficiency. However, the authors also note that speech interaction can improve efficiency because users are able to just say their requests, instead of searching through a GUI [2].

G8: Minimalism in Design and Dialogue: Three papers reflected on minimizing the amount of information presented to the user. One theme focused on the issues with maintaining large pieces of information presented by speech interfaces. Le Bigot et al. [9] found that, when exploring the primacy and recency effects in presentation of menu options, five or more options affected the ability to remember earlier options. In contrast, Wolters et al. [28] found that fewer options did not help in remembering all the options presented. Another issue was one of feedback statements to confirm that the speech interface understood the user correctly. Yankelovich et al. [29] found that implicit confirmations within dialogue are more preferred over repetitive explicit confirmations.

G9: Allowing Users to Recognize and Recover from Errors: Nine papers discussed allowing users to recognize and recover from errors. Often, these errors were speech recognition errors, which Shneiderman [24] notes are difficult for novice users to correct without creating more errors. Wolters et al. [28] also state the need to develop ways to recover from communication errors. Cowan et al. [5] noted the frustration of not being able to edit speech queries when users were misunderstood. Different ways were observed in how to correct these errors. Porcheron et al. [22] and Yankelovich et al. [29] found that users often repeat and refine misunderstood requests. Oviatt et al. [20] found that users spoke more loudly to the interface when it did not recognize them. Meyers et al. [16] found that if a user cannot fix an error, they will accept the error and move on. Another struggle is the ability to go back to a previous menu or undo an action. Begany et al. [2] noted participants' requests to

edit speech queries, while Zajicek et al. [30] highlight the difficulty of returning to previous menus.

G10: Providing Help and Documentation: 3 papers reflected on help and documentation in a voice UI. Corbett & Weber [4] found that users performed better with a speech interface after an interactive tutorial. Yankelovich et al. [29] identified that providing help progressively throughout interaction guides the user in performing tasks efficiently. Zajicek et al. [30] and Corbett & Weber [4] both found that help that was provided contextually within the interaction was particularly useful. This both decreased cognitive load and provided assistance to users only when required.

Discussion

As seen in Table 3, a number of current design guidelines seem to be applicable to speech interfaces: G1, G2, G3, G6, and G9. ***Recognition over Recall (G6)*** is the most comprehensively covered in the research reviewed. As speech interfaces are often both displayless and presenting information using a single output modality (audio), it is understandable that this would be one of the most common occurring HCI issues in the literature. With no visual tools to help guide users through the interaction, it is important that users are given a way to learn how to interact with the interface through primarily audio channels (***Visibility/ Feedback (G1)***). This is a considerable challenge. This also means that users can only receive one piece of information at a time, compared to a graphical interface, which can use multiple means of instruction giving (e.g. pictures, text). This requires users to remember significant amounts of information, like the options and commands needed.

The principle of ***Consistency Throughout the Interface (G8)*** had little coverage in the literature. Though there may be little discussion this as of yet, this may be something that changes as the work on design of speech interaction develops.

It is clear from our review that some of the largest usability hurdles in hands-free voice UIs involve the cognitive load required in interaction, and the need to have control over interaction and dealing with errors. Guidelines such as ***Recognition over Recall (G6)***, ***Control and Freedom (G3)***, and ***Recovering from Errors (G9)*** stand out in this respect. There is also an apparent need to map interaction to familiar interactions in the world (***Matching from System to Real World (G2)***). These, then, are essential guidelines we must keep in mind when designers are designing hands-free speech interfaces.

Additional Guidelines

While existing guidelines are a good starting point, the nature of speech interaction leads to new problems that one would have not considered in traditional graphical interfaces. We noted some common issues that several papers had discussed in their evaluations or discussions. From the literature, we propose that the following guidelines be added to those highlighted:

A1: Ensure Transparency/Privacy: 4 papers discussed the issues of privacy with speech interfaces. Both Cowan et al. [5] and Zajicek et al. [30] found that users are unsure what information is being collected, due to a lack of transparency. Begany et al. [2] note the particular issue in public, when the information being shared can be heard by anyone that is within earshot of the user. Moorthy & Vu [14] also found that

users are concerned with privacy of their information when using speech interfaces in public.

A2: Considering How Context Affects Speech Interaction: 2 papers discussed issues with comfort level of speaking out loud to a speech interface. Begany et al. [2] and Howell, Love & Turner [6] found that users often feel uncomfortable speaking to their phone in front of other people, as this is not a common type of interaction. Begany et al. [2] report that people are more likely to use a textual interface in public, reserving speech interaction for when in private.

Conclusion

This paper explores how current speech literature in HCI has reflected on currently established design guidelines for graphical interfaces. We take three sets of guidelines and create 10 aggregated guideline categories. Based on a review of 21 HCI papers, we found that guidelines such as recognition over recall, supporting user freedom, and matching system with the real world are commonly discussed in speech literature in HCI. We also found that guidelines such as minimalistic design and consistency are not as discussed in the current literature. We identified two issues that are not represented by current guidelines for graphical interfaces and propose that these need to be considered when designing for speech interfaces.

As mentioned earlier, Nielsen [18] spoke about the significant usability issues the iPhone contained. An updated review was then written, talking about the improvement in usability [3]. At this point in time, we may be in the same situation Mobile UIs were a decade ago. In order to take advantage of the truly hands-free and mobile capability of voice UIs, these issues must be explored and addressed. The HCI community, particularly within Mobile HCI, needs to engage in the same research that Nielsen did, in order to improve the usability of hands-free voice UIs. 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 the HCI community in building more usable and intuitive speech interfaces.

Acknowledgements

This work was supported by AGE-WELL NCE Inc., a member of the Networks of Centres of Excellence (NCE) program funded by the Government of Canada.

References

1. Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd. *Proc. of CHI EA '14*: 749–760.
2. Grace M. Begany, Ning Sa, and Xiaojun Yuan. 2015. Factors Affecting User Perception of a Spoken Language vs. Textual Search Interface. *Interacting with Computers* 28, 2: 170–180.
3. Raluca Budiu. Progress in Mobile User Experience. *Nielsen Norman Group*. Retrieved May 16, 2018 from www.nngroup.com.
4. Eric Corbett and Astrid Weber. 2016. What can I say? *Proc. of MobileHCI '16*: 72–82.
5. Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can I help you with?" *Proc. MobileHCI '17*: 1–12.
6. M. Howell, S. Love, and M. Turner. 2005. The impact of interface metaphor and context of use on the usability of a speech-based mobile city guide service. *Behaviour and Information Tech* 24, 1.
7. Joseph Kaye, Joel Fischer, Jason Hong, Frank Bentley, Cosmin Munteanu, Alexis Hiniker, Janice

Tsai, and Tawfiq Ammari. 2018. Voice Assistants, UX Design and Research. In *Proc of CHI EA '18*.
8. Dominique Knutsen, Ludovic Le Bigot, and Christine Ros. 2017. Explicit feedback from users attenuates memory biases in human-system dialogue. *J of Human Computer Studies* 97: 77–87.
9. Ludovic Le Bigot, Loïc Caroux, Christine Ros, Agnès Lacroix, and Valérie Botherel. 2013. Investigating memory constraints on recall of options in interactive voice response system messages. *Behaviour and Information Technology* 32, 2.
10. Lucian Leahu, Marisa Cohn, and Wendy March. 2013. How categories come to matter. *Proc. CHI '13*.
11. Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces. *Proc CHI '15*.
12. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA". *Proc. CHI '16*.
13. Kathleen K. Molnar and Marilyn G. Kletke. 1996. The impacts on user performance and satisfaction of a voice-based front-end interface. *J Human-Computer Studies* 45, 3.
14. Aarthi Easwara Moorthy and Kim Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *J Human-Computer Interaction* 31, 4.
15. Cosmin Munteanu and Gerald Penn. 2017. Speech and Hands-free Interaction. *Proc. CHI EA '17*.
16. Chelsea Myers, Jessica Nebolsky, Karina Caro. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. *Proc. CHI '18*.
17. Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. *Proc. CHI '94*.
18. Jakob Nielsen. Mobile Usability, First Findings. *Nielsen Norman Group*. Retrieved May 16, 2018 from https://www.nngroup.com.
19. Donald Norman. 1988. The design of everyday things. *Doubled Currency*.
20. Sharon Oviatt, Colin Swindells and Alex Arthur. 2008. Implicit user-adaptive system engagement in speech and pen interfaces. *Proc. of CHI '08*.
21. Saverio Perugini, Taylor Anderson and William Moroney. 2007. A study of out-of-turn interaction in menu-based, IVR, voicemail systems. *Proc. CHI '07*.
22. Martin Porcheron, Joel Fischer and Sarah Sharples. 2017. "Do animals have accents?" *Proc. CSCW '17*.
23. Jahanzeb Sherwani, Dong Yu, and Tim Paek. 2007. Voicepedia: towards speech-based access to unstructured information. *Interspeech*: 2–5.
24. Ben Shneiderman. 2000. The limits of speech recognition. *CACM* 43, 9.
25. Ben Shneiderman and Catherine Plaisant. 2010. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*.
26. Bernhard Suhm. 2003. Towards Best Practices for Speech User Interface Design.
27. J. Wilke, F. McInnes, M. A. Jack, and P. Littlewood. 2007. Hidden menu options in automated human-computer telephone dialogues. *J Behaviour and Information Technology* 26, 6.
28. Maria Wolters, Kallirroi Georgila, Johanna Moore, Robert Logie, Sarah MacPherson, and Matthew Watson. 2009. Reducing working memory load in spoken dialogue systems. *Interacting with Computers* 21, 4.
29. Nicole Yankelovich, Gina-Anne Levow, and Matt Marx. 1995. Designing SpeechActs. *Proc CHI '95*.
30. Mary Zajicek, Richard Wales, and Andrew Lee. 2004. Speech interaction for older adults. *J Universal Access in the Information Society* 3, 2.