COGA TF review of NAUR working draft

https://www.w3.org/TR/naur/
APA would like to receive feedback by December 10, 2021.

General thoughts:

- [rain] Lumping “cognitive and learning disabilities” together in this document feels problematic to me because the range of needs may be so different. I’m curious if we need to suggest a different approach, such as focusing on what the user has challenges with. Supporting an aging individual with dementia will be very different from supporting a young adult with down syndrome in this context.
- [rain] Also the continual repetition of “especially those with cognitive and learning disabilities” in all of the user needs really isolates…
- [lisa] need a section : Reduce the reliance of memory and executive function
- [lisa] need a section : follow best practices (see
- [lisa] need a section : avoid distractions and unnecessary content
- [lisa] need a section : Provide 1 step access to help at any point
- {lisa] general comment need to check against the work done in: https://w3c.github.io/coga/issue-papers/#voice-systems , https://docs.google.com/document/d/1B1vCqlU1IF5UmqxhJAy8Khdi-kRQNPaIVX8f3lCMr7w/edit#heading=h.9j8rnhzb9sv8, and https://www.w3.org/TR/coga-usable/ all objectives but especially https://www.w3.org/TR/coga-usable/#objective-6-ensure-processes-do-not-rely-on-memory
- Is there a requirement to support TTS for any visual output for individuals with reading difficulties or who are blind/low vision?

Abstract:

- [julie] Second paragraph begins: “This document is most explicitly not a collection of baseline requirements.” But the sentence after this one is about another topic. Is there more to say about why this isn’t a collection of baseline requirements? And/or what this User Requirements document provides *instead* of baseline requirements?

1.1 What is a Natural Language Interface?

- [from a conversational designer that rain works with] I think it would be helpful to explicitly state that typed language is a common input method. In the item describing chatbots, it's not clear if it's referring to text-based chatbots (which I think is the most common interpretation of a chatbot for most people) or text- and speech-based chatbots. (Could list both as options.) (And good callout later noting that even text-based chatbots can let people use speech via a keyboard dictation function)

1.2 final paragraph:

- [rain] “The design of the application should support the cognitive needs of users, including those who have learning or cognitive disabilities. Discoverability, simplicity, and affordances for example, are important considerations in the design of the natural language interaction.”
- [rain] Does not include anything about supporting memory or task completion. Simplicity is probably the wrong word.
- [julie] This sentence is very jargony: “Discoverability, simplicity, and affordances for example, are important considerations in the design of the natural language interaction.” Is this what it’s trying to say? “The natural-language application must be easy to find and use.”??
- [julie] This final paragraph mentions users with learning and cognitive disabilities. Should we suggest also mentioning accessibility for users who are speaking non-native language? (This question also applies to the next section about scope.)

2.1 User Identification and Authentication

- [from a conversational designer that rain works with]
    - "A user who is deaf or who has a speech disability" --> requires biometric authentication other than speech. I would hesitate to make this a declarative statement; there are deaf users who use voice assistants (for things that don't require TTS output, like turning on the lights) and I would not want to imply they wouldn't use voice authentication in some scenarios (like displaying a home security camera feed on screen).
    - Which leads to another recommendation for accessibility, which is to make devices like smart speakers have additional affordances for people who are hard of hearing or deaf (like more signals with lights). Some interesting proposals in this paper: "It Didn't Sound Good with My Cochlear Implants: Understanding the Challenges of Using Smart Assistants for Deaf and Hard of Hearing Users"

2.2

- [rain] “A smart speaker only supports speech output” -- Maybe change this to audio/sound output? There are other forms of communication aside from speech, and as an interaction designer I could see other possibilities with sound. Also, many smart speakers also have haptic output that could be employed
- [rain] Love that REQ 9a and 9b are here to “encourage further research and development efforts”
- [rain] Languages and different languages…. Perhaps also add that people may frame words differently?

REQ 6

- [rain] “Support a mode of operation in which the user can speak to the system, and the system's natural language output is presented textually (e.g., displayed visually or conveyed via the user's assistive technology).” – [note from Rain - “textually” should probably be “visually” instead of “textually” so as not to limit possibilities in ways that hinder those with reading or text disabilities

2.3 Communicating in a Language that the User Needs

- [julie] “User need 11: A user with a learning or cognitive disability needs to communicate with the system in a symbol set supported by a particular augmentative and alternative communication (AAC) assistive technology.” This user need is not followed by an REQ—looks like REQ 11 is missing from this doc?
- [from a conversational designer that rain works with] As is noted, automatic sign language processing is challenging. I think it would also be worth calling out that sign language is not just about the hand signs, but is really a whole-body language, with a strong reliance on accompanying facial expressions as well. So while being able to interpret signs is great, it's not the whole solution.

User need 11

- [rain] Mention of cognitive only focuses on symbols, which is only at the extreme end. What about memory supports, word-finding, echolalia, needing more time to speak, etc.?

2.4 Speech Recognition and Speech Production

REQs 12

- [rain] is it worth giving examples of atypical speech so that people don't just think of the extreme edge cases? Possible examples:
    - Describing a word rather than using the word
    - Using a similar word that is technically correct, but contextually not correct
    - Stutter and echolalia
    - Unnatural (and potentially prolonged) pauses that may be caused by either word finding, or discomfort speaking
    - More pronounced and progressive slurring such as that being worked on by go/euphonia in the context of ALS
- [julie] I agree with Rain and this point applies to this entire 2.4 section, which mentions "atypical speech characteristics" and "atypical speech varieties" several times, but never gives any examples. Should our team suggest specific examples to include?
- [from a conversational designer that rain works with] I am curious about how strongly these requirements are meant to be. Like for this example... Google has been working for years on this problem and we're still not there yet. So I can't imagine most builders of tech to be able to meet this kind of requirement. I know later on it says to provide alternatives, but perhaps it's also worth calling out that some of these requirements are more challenging than others, and focus instead on making sure there are alternative ways to provide input in these cases?
- [rain] where possible, AI might be a tool to determine contextual accuracy and closeness to try to help with meaning, especially with "real word errors" (errors that are technically correct, but given the larger context and purpose, are wrong)
-

REQ 13b

- [rain] Suggest removing "repeat" and just focus on confirm… maybe "confirm understood correctly" – for some individuals with speech differences, it takes a lot of effort to talk, so can this be made easier.
    - asking the user to repeat could be a big ask. It may be helpful to prompt in some way to only ask for what may not have been clear, or instead to say "I heard…" and "is that correct?" to help the user with working memory or reduction of speaking effort

REQ 13c

- [rain] Wondering if we can add clarity that other sounds are also an option? E.g., grunts, claps, etc.

2.6.1 Understanding How to Interact with the Interface

REQs 17:

- [rain] keep in mind that often the individuals who may need this most may be the least likely to be able to find (or comfortable changing) accessibility settings
- [rain] consider that working memory is absolutely essential to navigating voice menus, and this is something that can be difficult for many individuals for a broad range of reasons, including situational. Enable the user to ask questions out of the flow to try to help them through this process, or give them a way to access reminders.
- [from a conversational designer that rain works with]
    - I think this one is tricky. With a personal assistant, there is really no practical way to do something like this, when there are hundreds or potentially thousands of things the system can do. And we struggle mightily with feature discovery.
    - Rather than try and force a menu structure onto broad-range assistants (which can also be cognitively overwhelming), perhaps we can instead focus on the second section, about ensuring robust grammars/NLU, so that users can ask for what they need in a natural way, and not have to memorize commands. In addition, "just in time" help--where we provide the right option at the right moment in time, when it's useful for the context the user is in--is another recommendation.
    - And of course, making sure the system can answer "how to" questions, so people can ASK for things like "how do I change my font settings" or "make the font bigger" rather than having to navigate challenging menu structures etc. or be forced to switch to another device (e.g. go from a smart speaker to a phone) to complete a task.
    - Another good requirement--allow users to complete the task on the same device they started it on

2.6.2 Giving Users Enough Time to Interact

User Need 20:

- [rain] In addition to ample time to figure out how to respond, consider ample time while producing the speech response
- [rain] The note under here is problematic (“it should be optional”) because individuals who need this support are far less likely to find the “option” – is there another way to frame this? Such as, if the user appears to need more time, provide a way for them to communicate their needs without invoking stress…

- [from a conversational designer that rain works with]
    - While it's really important to allow people enough time to respond via voice, practically speaking and for privacy reasons, we can't leave the microphone on indefinitely. The requirement to allow the user to adjust this timeout is a good one; some people will want a longer timeout, some people will prefer to keep it shorter.
    - One related recc: allow an audio option to signal when the microphone is on or off, in addition to a visual one.

REQ 21a

- [rain] “use language… that is no more complex than necessary for clear communication” – not sure what my reaction is here, but want to make sure that we aren’t dumbing this down…. Cognitive and learning disabilities do not necessarily imply the user isn’t intelligent

REQ 21c

- [rain] “simpler language than the default” – this feels strange to me. What is the goal? Is it simpler language, or more direct language like removing the personality?
- [rain] thinking about individuals who may have difficulty with numerical concepts (dyslcalculia), may be worth adding that the user might want to select a different way of representing numerical concepts that isn’t strictly number based. Example, instead of stating the temperature, state that it is “comfortably warm, typically considered t-shirt weather.”
-

REQ 21d

- [rain] “including users who may have learning or cognitive disabilities” – this addition feels problematic, like it is specifically calling out these individuals. Suggest just ending with “users of the system”

2.6.6 Note:

- [rain] “The purpose of this multimodal presentation of text is to enhance comprehension of the material, especially by people with learning disabilities that affect reading.” -- How about just “ people with reading disabilities.” since reading can be affected by many disabilities…

User need 27 Note:

- [rain] “Information presented graphically must also be available as text. See '§ 2.2 Means of Input and Output' above.” Specify that if the text is already an alternate or supporting presentation should not be read out as text because that would end up with redundant information for those using text only
- [julie] For User Need 27, should we suggest additional REQ or example about how familiar icons can help, but unfamiliar icons can be confusing/distracting?