Voice Control Guideline Scratchpad

Voice Control Guideline Scratchpad

Meeting logistics

- Day and Time
    - Tuesday - 12pm EST (2nd hour of the regular AGWG meeting, if confirmed by Chairs)
    - Thursday - 9am EST (regular meeting time)
- Meeting zoom information: https://usablenet.zoom.us/j/83932526255
- IRC: #wcag3-[1 word from guideline name]
- Github page
- Subgroup participation handbook
- Members: [List current Subgroup members]
    - Giacomo
    - Rachael
    - Filippo
    - Nina
    - Joe
    - Jamie
    - MJ
    - Mike
    - Graham
    - Makoto
- Previous Members: [List previous subgroup members]
- Scope

Resources

- Writing for WCAG3
    - Guideline Writing Process by Maturity Levels - Q1 2023
    - Process for Writing Guidelines using Maturity Levels
- Prior work
    - Research
        - understanding and a wiki page with research

        - [Add direct link to relevant breakouts]
    - Categorization exercise
        - Folder of documents from the Categorization Exercise
        - Categorization exercise 2 Aug.xlsx - extract of the database from the categorization exercise used to analyze information. It includes all the success criteria, even the ones that do not have an individual document.
    - Links to relevant content from Making Content Usable
- Guidance from BBC on hot to test with Dragon Naturally Speaking: https://bbc.github.io/accessibility-news-and-you/assistive-technology/testing-steps/dragon-windows.html
- Other potentially useful research and documents (external to w3c work)
    - [List all articles here]

Week 1: Review all research

Instructions

- Group introductions
- Decided on how minutes will be kept (IRC, running google doc, etc)
- Divide up research locations among team. Completing relevant research list is homework.
- Note: Do not cite WCAG 2.x directly in research. Instead trace guidelines back to the research or guidance that led to the SC.

Relevant Research

- [Article title] [author(s)][URL] [A few words about what is relevant in the article]
- [JG] Automatic speech recognition for assistive technology devices (included because there seems to be research on using speech to control an AT – as opposed to the page itself. So one model seems to support multiple layers of intermediation: user>speech rec>at>page?)
    - Some of the steps they've taken could be used to create an open API for voice command. Certain features, such as the creation of a word bank for different languages, might be useful for a whole web API for voice command. Also, their research steps can be used to help create a more robust API that can account for variances in speech or accent.

- [pending access] Speech input to support universal access (Requested through ILL 3 December 24- RLB)
- [MJ] Machine learning assistive application for users with speech disorders (research specific to speech support for users with atypical speech)
    - The study aimed to develop a machine learning-based automatic speech recognition system tailored for users with dysarthria, a neuromotor speech impairment. The researchers focused on a keyword spotting task using a convolutional neural network model. They developed a mobile app, CapisciAMe, to collect speech samples from users with dysarthria to train the model. The study found that the global speech model configuration, which includes samples from multiple users, showed better performance in keyword recognition compared to personalized models. The study faced challenges such as the limited number of participants and the variability in speech among users with dysarthria. Collecting sufficient data was also time-consuming and tiring for participants.
- [GP] Why college students prefer typing over speech input: the dual perspective (capturing that for people who do not self-identify, there may be a preference for keyboard; what does that infer for PwDs?)
    - Short study description: This study investigates why college students resist adopting speech input. Despite the influence of easy of use and perceived usefulness on adoption intentions, operational challenges and "uncertainty" discourage transitioning from traditional typing method (focuses more on dictation aspects rather than voice control).
    - Study limits:
        - Does not distinguish between levels of experience
        - Homogeneous sample of college students (not generalisable)
        - Does not have actionable items to review
- [rachaellbradley@gmail.com if possible, we need access to this document] Foundational article:
    - Voice-input aids for the physically disabled - historical article from 1984
- Voice user interface (VUI) related articles:
    - Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study [Multiple authors]. Includes findings about how the most common issues for people using VUIs is lengthy commands. To prevent this issue, commands should be short and easy to speak.
    - Improving the Usability of Voice User Interfaces: A New Set of Ergonomic Criteria [Caroline Nowacki, Anna Gordeeva & Anne-Hélène Lizé]. Includes guidelines for VUIs. Some of the guidelines are relevant for usability with voice control software. In particular, the guidance for internal consistency is relevant and useful. A user need from this would be to ensure that controls that perform similar actions, can be activated consistently, with the same commands.
    - [NK] What can I say?: addressing user experience challenges of a mobile voice user interface for accessibility

- **Description:** Study about visibility and learnability of voice commands of a M-VUI application being developed on the Android platform
    - Participants with a variety of motor impairments, unpublished software VoiceNavigator tested on a Nexus 5, operating system Android KitKat
- **Issues & learnings:**
    - Software didn't offer a continuous listening mode (participants expected to be heard and understood at any point in time and be listened to as long as they talk)
    - They didn't like the visual clutter (every touch target came with a box and an associated number)
    - Some parts of the visual interface were even blocked by the overlay which resulted in missing information
    - The participants were unsure when their input was perceived as dictated text or navigation mode
    - Users were guessing for the most part when trying to figure out what to say to archive a certain action to happen
    - It gets harder to guess what to say when there are many visual elements without obvious names/labels on the screen
    - The numbers became popular shortcuts when users didn't know what to call a certain touch target
    - Some users thought about how they would perform a certain physical interaction and translate that into a voice command
    - Providing a tutorial with commands for onboarding was very effective

- Accessible Voice Interfaces. Brewer, et. al. 2018. [need full text]
- [GP] ~~Interacting with mobile devices via VoiceOver: usability and accessibility issues B Leporini, MC Buzzi, M Buzzi - … of the 24th Australian computer-human …, 2012 - dl.acm.org~~
    - Short study description: this is entirely related to the usage of VoiceOver (output, not input).
    - This is out of scope for our activity
- [GP] Voice Games: Investigation Into the Use of Non-speech Voice Input for Making Computer Games More Accessible Susumu Harada, Jacob O. Wobbrock, James A. Landay.
    - Short study description: This study tested a new input device for hands-free control of computer games using non-speech voice commands. The voice game controller combines non-speech and speech input to improve gaming performances, especially for users with motor disabilities. The findings suggest that voice-driven input can broaden interactive gaming, offering new possibilities for all users, especially people with disabilities.

        - Study limits: The voice game controller translates voice commands into mouse/keyboard commands, which is more about hardware rather than authoring.
    - [RLB] The state of speech in HCI: Trends, themes and challenges L Clark, P Doyle, D Garaialde, E Gilmartin… - Interacting with …, 2019 - academic.oup.com **(RLB)**
        - "malberti et al. (1993) showed that people adapt their language choices according to their partner models but noted that differences between human and computer speech choices decreased as people got more familiar with the interaction. People also tended to use fewer fillers (e.g. "um", "err"), request confirmation and repetition more, and use fewer topic shifts in computer compared to human interaction. Kumar, Paek & Lee (2012) compared existing dictation with "Voice Typing" - a speech interaction model that transcribes users' utterances as they are produced, allowing for error identification in real-time. In using this, their study showed a reduction in error rate and certain cognitive demands compared to dictation. Another paper explored the impact of spoken translation software on cross-lingual dialogues (Hara & Iqbal, 2015). During experiments, participants were observed adapting their speech and comprehension due to imperfections in system- produced translations, and the authors accordingly formulated a set of design guidelines for such systems." RB note: Should a training program be available to help users prepare for interaction?
        - "
    - [GR] Patterns for how users overcome obstacles in voice user interfaces C Myers, A Furqan, J Nebolsky, K Caro… - Proceedings of the 2018 …, 2018 - dl.acm.org **(GR)**
        - **Specific use case / software (limited example / use case)**
        - **Errors**
            - Majority of errors were voice recognition misunderstanding intent
            - Next most often encountered error was speech recognition of accents, pronunciation etc.
            - Additional secondary complaints are ambiguity in what the system can actually do (what commands it can recognise / process)
            - Confusion in feedback / users missing feedback / expected feedback not given
        - **Techniques Employed to overcome errors**
            - Hyper-articulation
            - Simplification of commands
            - Rephrasing / New utterance
            - Adding additional info to commands
            - Use GUI to work out what command to say
            - "Settling" - errors in speech recognition ("effects" and "FX") that cannot be corrected are simply left.

            - Restarting - if a barrier is hit that cannot be overcome within a process they start the whole process again.
            -
    - [GP] Older adults and voice interaction: A pilot study with google home J Kowalski, A Jaskulska, K Skorupska… - Extended Abstracts of …, 2019 - dl.acm.org
        - Context: older adults in the context of Smart Home Technology
        - Short study description: A group of older adults participated in this study, where they were asked to use a Google Home Speaker to control various devices. The results highlighted positive feedback, with participants expressing their amazement at the range of possibilities offered by the technology.
        - Benefits:
            - Intuitive interaction via natural language
                - Improvements: more feedback allowing to create a step-by-step guide.
            - Voice control, no motor function involved
                - Improvements: understanding more natural utterances, including context and metaphors, as well as tackling and explaining the problem of voice priority to prevent conflicts that may arise.
            - Friendly manner, friendly voice and patience
                - Improvements: building up on the voice recognition functions to initiate friendly conversations with reminders.
            - No handling of devices, no device that has to be found and turned on
                - Improvements: solving concerns about the range of effective voice interaction in their home.
            - Granting independence, the tech can do some things which may require assistance.
                - Improvements: ensuring the existence of fail-safes to resolve concerns about the reliability of VIUs in executing commands.
        - Barriers:
            - Time consuming
                - Note: although this was mentioned, it was not a drawback for older adults as they say that they have the time
            - Lacks sensors and cameras, which would allow it to better assist with some tasks
                - Note: connecting a camera, to give the user hints
            - Lack of a screen to give feedback and context
                - Note: introduction of companion screens to see context, status or key information searched for, as it is hard to store it all in memory

- Need to have compatible devices, fear of losing diversity and individuality
    - Note: working towards compatibility between manufacturers
- Fear of malfunction, something may not turn off, even if the signal was sent
    - Note: making clear what backup security measures are in place
- Fear of too much reliance, afraid of a possible loss of creativity, and lack of mental and physical exercise
    - Note: serve as an assistant, verifying their cognitive health and reminding them about some elements of a healthy lifestyle.
- Danger of entering a "search bubble", without the text interaction with a lot of context sometimes it is hard to find exactly what we are looking for, or to remember what it was
    - Note: a companion screen could mitigate this effect, if the user could glance at it and request to be read a specific result.

- [MU] Voice Coding Experiences for Developers with Physical Impairments **(MU)**
    - 6 participants are utilising Talon whilst 1 has previous experience in using Dragon NaturallySpeaking
    - Participants were aged between 20 and 41 with coding experience ranging between 2.5 and 27 years
    - **Talon Voice Experiences:**
        - Using tools such as Talon requires some technical experience (e.g., to write Talon scripts that map voice commands to specific desired actions).
        - Custom voice commands need configuring when working with a new programming language to ensure that voice coders can operate at a similar level to non-disabled developers.
        - The need to know a large vocabulary of commands which can present a learning curve for those new to voice coding.
        - Felt that after a year of voice coding someone can be as efficient as developers using a mouse and keyboard
        - Code faster now using voice compared to when previously using traditional input devices.
        - Voice RSI (repetitive strain injury) can present a problem if users are new voice interaction.Common accessibility issues with VS Code when using Talon where voice coders cannot easily access the sidebars and do not know what has focus within the interface.

        - Modifying code is challenging via voice tools such as Talon, as well as navigating through syntax and performing standard actions such as copy and paste.
- **Voice Recognition Accuracy:**
    - Accidental actions triggered by misrecognition could cause frustration (e.g., when chaining multiple voice commands together)
    - Longer commands within Talon are typically recognised more accurately
    - Out-of-vocabulary words can be difficult to input using voice (these need to be entered letter-by-letter)
    - It was felt to not be intuitive due to the high number of voice commands that need to be learnt
    - Four participants (P3, P4, P6, P7) also highlighted Rango (2024) as a useful browser extension for supporting web accessibility and wider development tasks (through attaching labels to all links on a web page which users can verbalise to perform a selection)
- **Additional Tools:**
    - Applications which require the use of a mouse (e.g., features based within hover states) can cause issues for voice coders
        - Emphasised the need for gaze interaction to address problems associated with these types of scenarios
        - Utilise a dense mouse grid for performing clicks
    - Use foot switches to facilitate stable input (e.g. to enable or disable eye gaze interaction)
    - Can be problematic when used with voice assistive devices and requires filtering of voice commands, normal speech, and environment noises
- **Voice Coding Research Prototype**
    - Autocompletion was also supported, so that when users verbalised characters a pop-up was displayed that showed completion options.
    - Had a number associated with it that users could verbalise to action their desired completion
    - Numbers were placed next to files located in the file explorer to support users in opening files via vocally specifying the numbers associated with them (e.g., “open file 1”)
        - Implemented to address the potential challenges in verbalising the names of files which may in some scenarios present recognition issues
- **Speech recognition architecture**
    - Consists of three elements:
        - (1) a speech recognition component to convert speech into text (using the Web Speech API),
        - (2) a command interpreter to process user commands, and

        - (3) an execution routine that processes actions based on the commands issued by users.
    - **Verbose Voice Commands**
        - Use of two words in some commands was too verbose (e.g., “command space” to add a space)
            - Prefer monosyllabic commands
            - Downside of monosyllables is it is easy for them to overlap and can result in misrecognitions
        - Have to experiment with synonyms and find what works optimally for them (especially for different accents)
    -
- [MU][blog] Accessibility and me: Rani Nayyar **(MU)**
    - Use Dragon Naturally Speaking software
    - Use the Dragon commands that have been written and I have had to train it for specific commands
    - Navigating around some of the pages has been straightforward so far, by saying any of the following commands:
        - “go to top”
        - “go to bottom”
        - “page down”
        - “page up”
        - and a good one is “start scrolling up/down”
    - Struggle to use Dragon around the internet, like for online banking
    - To verify my security questions. I end up having to manually use an iPad for general shopping, online banking and reading emails.Take a lot of time and effort to learn how to use spreadsheet developed in-house
    - Have problems navigating around our corporate applications
    -
- [GR][YouTube video] Voice Navigation 101 with Matthew Putland (October 2024) **(GR)**
    - **Nearly all of the points here directly reference existing WCAG best practices:**
        - Accessibility Issues with Labels - labels must be programmatically associated (i.e. 2.5.3)
        -
- [GP] Designing SpeechActs: Issues in Speech User Interfaces **Should have design guidance.
    - Short study description: the document highlights the potential of speech-based interfaces named SpeechActs for mobile professionals, offering hands-free, intuitive access to applications such as email and calendar via natural language.
- [JH] Moran, S., Pantidi, N., Bachour, K., Fischer, J. E., Flintham, M., Rodden, T., … Johnson, S. (2013).Team reactions to voiced agent instructions in a pervasive game (p. 371). ACM Press. https://doi.org/10.1145/2449396.2449445

- [JH] Sammon, M. J., Brotman, L. S., Peebles, E., & Seligmann, D. D. (2006). MACCS: Enabling Communications for Mobile Workers Within Healthcare Environments. In Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (pp. 41–44). New York, NY, USA: ACM. https://doi.org/10.1145/1152215.1152224.. [Rachael had PDF]
    - Short study description: an industrial user study in a real world healthcare environment over 2 months with 35 users of Mobile Access to Converged Communications System (MACCS) equipping mobile workers (staff whose jobs require moving like nurses and retail stockers) with a hands-free voice interface to manage their communication. In addition they also discuss the design, implementation and deployment of MACCS.
    - WOZ testing method (Wizard of Oz Testing) usability testing method in which a human operator simulates the behavior of a computer system behind the scenes.)
    - Study participants used a headset with attached microphone to make commands and do tasks like message or locate another employee, set availability to “busy”, and query the system for specific questions, and accept incoming contact from others.
    - Challenges: participants complained the tool did not hear them well to do the command- ambient sound and variations on how the user asked (alternate terms or tone/volume) primarily contributed to this. The users did not regularly use some of the features, citing either preference for familiar tool (nurse locating system) or not understanding value for them
    - All found it improved paging process (100%) and hands-free was an advantage (93%); results on whether it improved productivity and job functions was around 70% agreed
- Tutor design for speech-based interfaces. J Hakulinen, M Turunen, EP Salonen, KJ Räihä
- [JH] Derriks, B., & Willems, D. (1998). Negative feedback in information dialogues: identification, classification and problem-solving procedures. International Journal of Human-Computer Studies, 48(5), 577–604. https://doi.org/10.1006/ijhc.1997.0182. [Rachael had PDF]
    - Study: Dense research article on nuances of negative feedback (the phrases uttered when trying to get a task done like on a call, the response of “sorry, I don’t understand.” Vs more positive or specific alternatives)
    - WOZ testing
    - Four basic communicative functions are distinguished: contact (willingness and ability to continue the interaction), perception (willingness and ability to perceive the message), understanding (willingness and ability to understand the message) and reaction (willingness and ability to adequately respond to the message, in particular the acceptance or rejection of the message)
    - whatever the problem, a resolution by repetition works best. If the problem is identified (Q), new or correct information is given as the

resolution. If the problem remains, a systematic return to the main aspects of the request until the problem is correctly recognized by both speakers is often the only solution. This complex procedure can take many tries
        - exchange of information concerning particular names or numbers often remains a source of difficulties.
    - Wilke, J., McInnes, F., Jack, M. A., & Littlewood, P. (2007). Hidden menu options in automated human – computer telephone dialogues: dissonance in the user's mental model. Behaviour & Information Technology, 26(6), 517–534. https://doi.org/10.1080/01449290600717783.
        - In speech-enabled applications, menus may facilitate the interaction for novice or infrequent users by promoting a step-by-step interaction, but can also render the interaction in speech-driven applications un-necessarily stilted and long. The challenge to designers of such applications is to strike a balance between restricting the user inputs and at the same time conveying to the user a conceptual model that allows them to fully exploit the strength and flexibility of the speech recognition technology. With the automated telephone services becoming ever more ubiquitous in society, and with the increased application of speech recognition in such services, this is a research domain well worth exploring.
    - [FZ] Wilkie, J., Jack, M. A., & Littlewood, P. J. (2005). System-initiated digressive proposals in automated human–computer telephone dialogues: the use of contrasting politeness strategies. International Journal of Human-Computer Studies, 62(1), 41–71. https://doi.org/10.1016/j.ijhcs.2004.08.001.
        - This article explains how human-computer telephone call interactions can be improved using politeness. Full access to the articles is needed
    - [JH] Wolters, M., Georgila, K., Moore, J. D., Logie, R. H., MacPherson, S. E., &b Watson, M. (2009). Reducing working memory load in spoken dialogue systems. Interacting with Computers, 21(4), 276–287. https://doi.org/10.1016/j.intcom.2009.05.009.

Possible Additional content from 2017-2024 CSUN Journal on Technology and Persons with Disabilities (volume 4 to 12): (posting but please weed out any unrelated to current scope) https://scholarworks.calstate.edu/catalog?utf8=%E2%9C%93&search_field=all_fields&q=The+Journal+on+Technology+and+Persons+with+Disabilities%2C

- [FZ] Vol 12 (2024): User Experience of Voice Assistants by People with Visual Disabilities, p. 133 Hyung Nam Kim, http://hdl.handle.net/20.500.12680/z029pc91k
    - Bad user experience can cause more problems to people with visual disabilities
    - This is out of scope for our activity

- [FZ] Vol 12 (2024)AAC Social Communication Group Photovoice, p. 307 Samuel Sennott, Linda Akagi, Sam Vranizan, Laura Moeller, Lateef McLeod, http://hdl.handle.net/20.500.12680/p2677364s
    - This is out of scope for our activity
- [FZ] Vol 11 (2023) Disability Bias & New Frontiers in Artificial Intelligence, p. 28, Christopher W. Land http://hdl.handle.net/10211.3/225163
    - This is out of scope for our activity
    - (pag. 19) Task Assisted Browsing consists in navigating a page using AI and a chat: an algorithm reads the webpage and identifies common elements. The user chats with the AI which gives the user a list of commands to navigate the page. The main problem is the inconsistency of components (like "search" and "filter" components) and the relationship between elements
- Vol 11 (2023) Empathy Talk with the Visually Impaired in Design Thinking, p. 153, Hyung Nam Kim http://hdl.handle.net/10211.3/225171
    - This is out of scope for our activity
- Vol 10 (2022): Creating Accessible XR Technologies: Rehabilitation for TBI, p. 1, Jesse D. Flint, Jennifer M. Riley, Caitlin J. Lang http://hdl.handle.net/10211.3/223462
- Vol 10 (2022):Online Learning & COVID-19: Exploring Digital Accessibility, p. 82, Justin Brown, Ruchi Permvattana, Scott Hollier, Jason McKee http://hdl.handle.net/10211.3/223467
- Vol 10 (2022):The Decentralized Education of Digital Accessibility for Technologists, p. 206, Dana Frayne http://hdl.handle.net/10211.3/223475
- Vol 10 (2022):Video Game Trends Over Time for People with Disabilities, p. 232, Sarah Mosely, Raeda Anderson, George Usmanov, John Morris, Ben Lippincott http://hdl.handle.net/10211.3/223477
- Vol 10 (2022):People with Disabilities Online Engagement During COVID-19, p. 266, Raeda Anderson, George Usmanov, Nicole Thompson http://hdl.handle.net/10211.3/223479
- [JH] Vol 9 (2021) Non-Verbal Interaction with Virtual Home Assistants for People with Dysarthria, p. 71, Aisha Jaddoh, Fernando Loizides, Omer Rana http://hdl.handle.net/10211.3/219936
- Vol 9 (2021) Haptic Paradigms for Multimodal Interactive, p. 110, Jennifer Tennison, Jesse Greenberg, Emily Moore, Jenna Gorlewicz http://hdl.handle.net/10211.3/219938

- Vol 9 (2021) Survey of User Needs: eGaming and People with Disabilities, p. 157, Nicole Thompson, Nicholas Ehrhardt, Ben Lippincott, Raeda Anderson, John Morris http://hdl.handle.net/10211.3/219941
- Vol 8 (2020) Smart Home Stress Assist: A Real-Time Intervention for PTSD, p. 40, Leighanne Jarvis, Tracey Wallace, John Morris, Kevin Caves http://hdl.handle.net/10211.3/215978
- Vol 8 (2020) "Alexa, Can You See Me?" Making Individual Personal Assistants for the Home Accessible to Deaf Customers, p. 130, Gabriella Wojtanowski, Colleen Gilmore, Barbra Seravalli, Kristen Fargas, Christian Vogler, Raja Kushalnagar http://hdl.handle.net/10211.3/215984
- Vol 8 (2020) Wireless Device Use by Individuals with Disabilities: Findings from a National Survey, p. 196, Nathan W. Moon, Patricia C. Griffiths, Salimah LaForce, Maureen Linden http://hdl.handle.net/10211.3/215988
- Vol 8 (2020) How WCAG 2.1 Relates to Online User Experience with Switch-Based Tools, p. 223, Sambhavi Chandrashekar, Lindsay McCardle http://hdl.handle.net/10211.3/215990
- Vol 8 (2020) User Personas: Smart Speakers, Home Automation and People with Disabilities, p. 237, John T. Morris, Nicole A. Thompson http://hdl.handle.net/10211.3/215991
- Vol 8 (2020) Digital Tech for Inclusive Aging: Usability, Design and Policy, p. 257, Alexander H. Denker, Paul M.A. Baker http://hdl.handle.net/10211.3/215992
- Vol 7 (2019) Global Atlas of People with Profound Intellectual and Multiple Disabilities, p. 106, Meike Engelhardt, Bartosz Gluszak, Michal Kosiedowski, Torsten Krämer, Jaroslaw Urbanski http://hdl.handle.net/10211.3/210394
- Vol 7 (2019) Accessibility of Voice-Activated Agents for People Who are Deaf or Hard of Hearing, p. 144, Jason Rodolitz, Evan Gambill, Brittany Willis, Christian Vogler, Raja Kushalnagar http://hdl.handle.net/10211.3/210397
- Vol 6 (2018) A Toolkit for User-Centered Design of Assistive Technology Solutions, p. 43, Christoph Veigl, Klaus Miesenberger http://hdl.handle.net/10211.3/202985
- Vol 6 (2018) RingBoard - A Dynamic Virtual Keyboard for Fist Based Text Entry, p. 83, Bradley Wojcik, Tony Morelli, Bryan Hoeft http://hdl.handle.net/10211.3/202987
- Vol 6 (2018) Smart Speaker Usability by Military Service Members with mTBI and PTSD, p. 127, Tracey Wallace, John T. Morris http://hdl.handle.net/10211.3/202991

- Vol 6 (2018) Mobile Health Apps and Needs of People with Disabilities: A National Survey, p. 149, Frank DeRuyter, Michael L. Jones, John T. Morris http://hdl.handle.net/10211.3/202993
- Vol 6 (2018) Development of a Servious Gaming App for Individuals with Spinal Cord Injury, p. 162, Michelle A. Mead
- Vol 6 (2018) Virtual Reality Based Scalable Framework for Travel Planning and Training, p. 219, Loren Abdulezer, Jason DaSilva http://hdl.handle.net/10211.3/202997
- Vol 6 (2018) Survey of User Needs for ICT - Community Living by People with Disabilities, p. 230, Michael L. Jones, Frank DeRuyter, Nicole A. Thompson, Jenna Norelli, John T. Morris http://hdl.handle.net/10211.3/202998
- Vol 6 (2018) A Pilot Study of Computer Auto-Personalization at American Job Centers, p. 247, J. Bern Jordan, Gregg C. Vanderheiden, Maureen Kaine-Krolak, Vera Roberts http://hdl.handle.net/10211.3/202999
- Vol 6 (2018) Media Player Accessibility: Summary of Insights from Interviews & Focus Groups, p. 325, Terrill Thompson http://hdl.handle.net/10211.3/203005

Non-academic articles about voice activated devices

- https://www.orcam.com/en-us/blog/voice-activated-devices
-

Week 2-3: User Needs

Instructions
Make a list of the user needs from research. Clearly reference the research. It does not have to be peer-reviewed papers, but should have credibility within the accessibility community. The list needs to include:

- The barriers encountered by people with disabilities
- The common user needs that apply to all the disability groups
- The unique needs (if any) that only apply to a specific group of functional needs.

When a user need is related but not necessarily within scope, note it anyway and add a note, “may be out of scope”

Capture user needs, even when they are not definitely testable. Assertions (using a process to improve accessibility instead of testing a result) are an option.

User Needs with Referenced Research or Gaps in Research

User Need	**Research**
I need to have Real-time text alternative to use a one-on-one communication equivalent to one-on-one speech interface	● [Article title from Relevant research list] ● The group is unclear on what exactly this means, but thinks it is out of scope for this outcome
~~I need Speech to never be the only way to control a page / site~~	A legitimate user need, but 'an alternative to voice control' seems out of scope?
I need Speech to never be the only way to interact / chat on a site	Same as prior; non-reliance on speech for interaction seems to be out of scope
If there is multi-way discussion (more that two parties) then I need all speech to be changed to text so I can view it	This seems to be speech to text (the opposite of the wording of another break-out group). Out of scope?
If there is multiway discussion (more that two party) then I need all text to be changed to speech so everyone else can 'hear' it	This seems to be text-to-speech (which is a separate breakout group). out of scope for Voice control
A user has another way of interacting (mode) that does not involve input or output by voice	out of scope for Voice control
A user has the ability to provide input using voice/voice alone	"dictation"
Users have the ability to manipulate/control objects via voice	
A user has the ability to receive output in speech	out of scope for Voice control
Users have the ability to use the keyboard API via voice/speech/audio input (assuming already in place via 2.1.1 in WCAG 2)	out of scope for Voice control
A comment more than a user need: text to speech is about 'translation/transposition' of context from speech to text. Not about interaction	

User Need	Research
I need controls to be labelled by their action first. For example a link to a company's YouTube channel should be labelled "YouTube for <company>" not "<company> YouTube channel" - this is important for voice control to ensure consistent selection when saying commands like "Click YouTube".	
I need to be able to use information in the accessibility tree to be available via a speech API	Low level, but potential way to meet outcome?
I need to be able to access what voice control commands are available for me to use.	https://chelmyers.com/papers/CameraReady_PatternsforHowUsersOvercomeObstaclesinVoiceUserInterfaces_v8.pdf There is a bit in this paper on how most users relied on guesswork rather than recall or visual aids.
I need icons to have a way of exposing their alternative text so I know what to say to click them when using voice control.	
An ability to operate the pointer API via voice	
I need a way to distinguishing between two controls with similar/overlapping names	
I need components that have the same functionality within a web page to be identified consistently.	
I need components on page that have different functions to be uniquely identified.	
I need controls to not be accidentally activated by my speech when I am dictating	
I need my voice control software (or voice control embedded into a page or voice controlled interface) to allow me to train it on my voice so I do not need to hyper-articulate	

User Need	Research
I need to converse naturally. Having to remember specific command structures / orders is a cognitive drain	https://www.smashingmagazine.com/2022/02/voice-user-interfaces-guide/#2-natural-conversation Partially covered here in point 2.
I need responses to be succinct and my options to be short and clear. For example, if I ask for a list of possible actions to perform, if you provide me with more than 3 or 4 I will forget some of the options.	
I want my voice control system to provide me with a simple cue that it has heard my command when I stop talking and while it is thinking. A simple beep for example. This means I am not sat wondering if my system has heard me and start repeating myself.	AT related
Confirmation of actions. If I give a command to do something I need it to be reversible the first time (in case the system misinterprets / mishears my command). For example “transfer $1000 to Graham” - the system should respond with a confirmation question also stating any important information such as “you want to transfer $1000 to Graham, is this correct”. This ensures mistakes are reversible. A better example that is outside of current WCAG guidance might be “order me the blue slippers” and the system repeats the colour, brand etc. to ensure it is correct.	

Week 4: Tests

Instructions

- Briefly describe or sketch out the test that would be needed to determine if the user need has been met. This can be a structured (traditional computational or guided) accessibility test, a group of tests, an evaluation (fail, good, excellent) or an assertion. An assertion is a declaration that a process was followed instead of a result was tested.
- Aim for covering the user need not for perfectly written tests
- See Writing Process Tests for Goals

List or Table of Tests

- [User Need 1]
    - [Test 1 needed to support user need]
    - [Test 2 needed to support user need]
- [User Need 2]
    - etc.

Week 5: Guidelines written as Outcome Statements

Instructions

- Using the tests list, write plain language outcomes for the user needs. This may mean reworking user needs. See examples

Outcomes

- [User Need 1]
    - [Outcome 1]
        - [Test 1]
        - [Test 2]
    - [Outcome 2]
        - [Test 1]
        - [Test 2}
- [User Need 2]
    - Etc..

Week 6-8 Iterate and Write Pull Request

Instructions

- Revisit research and revise the Guideline(s), User Needs, Outcomes and Tests until you are satisfied with them.
- Remember these are at the exploratory level. They do not need to be perfect, just get us going in the right direction.
- Clearly note where additional research is needed
- Call out which outcomes would be difficult to incorporate in WCAG 2.
- Create a pull request in https://github.com/w3c/wcag3 with the new content

Notes:

Key Questions/Status About Name/ Scope

- Does this include Speech to text dictation?
  *The group thought 'no', but did have questions about key commands used as part of dictation correction (i.e., "**select** [text string]"; "**delete that**")*

- Should we rename this to "Direct Voice Input/Output (to/from Webpage/App)
  Note if we do this we are excluding the following as means that that people need to
  - Voice (not what you say but what your voice is)
    - Voice ID?
  - Vocal (sounds made by person but not necessarily speech)
  - Audio - any sounds

  This is only about Voice Input, not Output

- We need to differentiate between Voice Input/output to the PLATFORM (device, OS, browser) and Voice input/output from the actual webpage/app
  We are assuming that all these will be AT mediated

- OUT OF SCOPE NEEDS… If we don't include voice input and output in our discussion where else it covered?
  Text to speech guideline (text content is converted into speech)
  User need underserved by current requirements: Speech output has a text output alternative (i.e., captioning of spoken content from OS, etc., beyond time-based media)
  May need a "speech output" guideline?

- Although good to allow voice input/commands/dialogWithAI to the webpage/app shouldn't we always require that everything can also be done without voice input/control etc. (or else people who can't speak or speak clearly enough will not be able to use the webpage/app.)
  For things like phone based interactions which ask for voice input, e.g. "ask your question" or "say your phone number".
  Assuming it covered by keyboard requirements, but need to keep an eye out for gaps.

- Ditto for speech/voice coming FROM the website. Good but can't be only?
  Audio seems to be a gap, where audio-only media alternatives is specific to particular media. Apps/sites could include speech-audio output without an alternative.
  Suggest we have a guideline that requires alternatives to speech output. (do not rely on audio-only output)

- How many people using speech recognition are trying to rely entirely on speech to control input? How successful are they with it?
  The answer is going to be "some", we may not know how many, but that doesn't make it less of a requirement.

- What can authors do to improve this?
  We have the WCAG 2 SCs, but need more info on how people interact now, and what causes problems for them.

- How are things different with mobile interactions where speech is more baked into the OS (i.e., siri on ios)?

- Known products: Jamie began listing some known tools, such as UtterlyVoice, scribe buddy. She is going to populate a list
- We need to create a section for definitions that is easy for the group to understand, with an example and a use case…. and which can help refine scope and research. (possibly from WCAG or WCAG2ICT key terms, but maybe expanding where those descriptions are not the full picture of what we need for this group:
  - User agent:
  - Voice/Speech input:
  - Voice/Speech output:
  - Dictation:
  - Text to speech:
  - Assistive technology (AT)
-

List of known "tools"

During the research phase, "tools" is used loosely, but we want to include any and all relevant named devices, tools, assistive technologies, software etc from research and/or personal experience. Anything that proports to meet user needs where users are intended to use their voice to interact. This will include formal authoring tools (link to definition) but also other known devices, software, or assistive technologies (link to definitions) either specific to voice input or that include voice input as a feature, possibly in a limited capacity like dictation.

- Software
  - Dragon Naturally Speaking
  -
- Smart Speakers
  - Alexa
  - Amazon Echo
  - Google Home

- Voice Enabled virtual assistants
    - Apple Siri
    - Samsung Bixby
    - Google Assistant
- Voice-related assistive technology
    - Apple Voice Control
    - Android Voice Access
    - OrCam MyEye
- Home automation
- Security systems
-

Minutes

Minutes:

12 November

- Gregg suggested to focus on web pages functionalities to recognize voice rather than focusing on User Agents ability to translate voice into inputs.
- Mike:
    - Being able to interact with speech voice inputs
    - Being able to interact not only with speech inputs

19 November

- The group agreed to meet on Thursday at 9 AM Boston time.
- The group refined the user needs, excluding those that seem to fall outside the scope of the activity.

21 November

Participants

- Giacomo
- Rachael
- Filippo
- Nina
- Joe

Minutes

- Rachael clarified the group's intent: to address user needs related to voice-based interaction with digital interfaces. The focus is on identifying the requirements that are necessary but currently lacking for users who engage with content specifically through voice commands.
- The group adhered to the exploratory process and timeline, beginning with the research phase. The instructions are to compile a list of the articles or studies found, including their titles, links, authors, and, if possible, a few key points summarizing the outcomes of each article.
- Joe discovered interesting articles regarding the verbosity of commands, emphasizing the need to avoid overly lengthy commands, such as potentially long accessible names (Not about assistive technology commands themselves, as these are more related to

assistive technology aspects, which are outside the group's author-focused scope). Additionally, Joe found another article advocating for internal consistency in labeling controls, suggesting that controls performing the same or similar actions should consistently use the same name.
- The other participants didn’t come across anything particularly noteworthy. Most of the articles they found focused on methods and optimizations for recognizing users’ voices, particularly for improving recognition of individuals with speech impairments.
- The final 10 minutes were dedicated to demonstrating how a speech recognition assistive technology, specifically Voice Control, functions and how users can interact with it.

Async tasks to prepare for the next meeting

- Search for research, articles, or studies related to the use of voice control assistive technologies.

5 December

Participants

- Giacomo
- Filippo
- Nina
- MJ

Research Needed

What we have found from existing Research:

Most of the existing research on voice control focuses on the ability of assistive technologies to convert user commands into inputs, recognise commands (such as those from individuals with speech impairments), and related features of these technologies.
The group didn't find anything about user needs and possible authoring responsibilities.

What we need from Research:

- We need to understand how individuals using voice control assistive technologies interact with the system (e.g., labels vs show numbers vs show grid)
- We need to understand the extent to which individuals using voice control assistive technologies rely on voice input, whether for dictation vs navigation, other/all tasks
- We need to understand and identify the challenges faced by individuals using voice control assistive technologies when navigating content/views (in general), categorised into challenges vs complete blockers
- A very specific use case: How does a blind individual using voice control interact with unlabeled buttons and other elements without a grid or visual reference?
- If we conduct tasks or surveys, we need to gather the following information (preferably through a recording of the live session):
    - The reasons they use voice control assistive technologies
    - Their level of expertise
    - The technology they are using and how they use it (what they like, types of navigation they use and in which conditions, how they feel in using this navigation, ...)
- Gather information to define the baseline that voice control assistive technologies should provide to the user
- A list of the combination of assistive technologies, devices, and operating systems used, along with their usage percentages, and an assessment of availability based on language (e.g., Dragon is not translated for Japanese users).

Suggested questions:

- To what extent do you use or rely on voice input?
    - Only for dictation
    - For some navigation (specify)
    - For everything
    - Other (specify)
- What strategies do people use to navigate?
    - Grid navigation
    - Saying the visible label
    - Asking to display acc names
    - Asking to display numbers close to each control

    - Other (specify)
- What issues do people encounter? (list all the issues)
- How does the sequence of words impact the effectiveness of your navigation? For example, how does saying “YouTube for <company>” differ from “<company> YouTube channel,” or what happens if you say “Click YouTube” because you see the YouTube icon but aren’t sure which is the acc name or label?
- How do you discover custom commands in an app to make navigation and interaction easier and faster (e.g., Google Doc shortcuts menu)
- When you encounter a button or link represented by an icon, how do you handle it?
- If two controls (links or buttons) have similar (or even worst) identical label/acc name, how do you manage it? What do you expect to happen?
- If you encounter the same functionality multiple times within the same view (e.g., a webpage), do you expect it to be presented consistently? If not, how does that affect you?
- If you encounter the same functionality multiple times within the same product (e.g., website), do you expect it to be presented consistently? If not, how does that affect you?
- Have duplicate names for operable controls ever caused issues, such as activating the wrong item? Do you remember any example? How did you handle this?
- Have you ever accidentally activated controls while dictating? If so, how did you address it?
- Have you ever had trouble remembering specific commands (cognitive)?
- Have you ever accidentally made a purchase and been unable to undo the action (this is an example, we want to know if after performing a legal/financial/etc. action the user was unable to undo the action)? What would you expect to say / do to undo an action?