Project Report

Visio Voice Playground:
Tech Coaching Module for blind and Low-Vision People

Konrad Krawczyk

Colophon

Thesis

Graduation project

MSc Design for Interaction

Faculty of Industrial Design Engineering

Delft University of Technology

Date and Place

August 26, 2024, Delft

Author

Konrad Krawczyk

Student # 5856728

Graduation Committee

Chair: Dr. dipl. -Des. Boess, S.U.

Mentor TU: Dr. Huisman, G.

Company mentor: T. van Hasselt

In Collaboration with

Koninklijke Visio

Preface

This thesis is more than just an academic exercise; it's a response to my experiences navigating the world of technology with a close family member who is visually impaired. Over the years, I have witnessed how voice technology can be both empowering and frustrating, especially when it feels like it was designed for someone else.

There have been so many moments when I've thought, "There must be a better way." It's surprising how often obvious challenges for someone who is blind can be completely overlooked by design stakeholders who don't share that perspective. This realization has driven me to dive deeper into the world of inclusive design, not just as a concept, but as a necessity. I believe that what's simply convenient for someone who can see can be life-changing for someone who can't.

My goal with this thesis is to contribute, in some small way, to making technology more inclusive. I've seen how an inclusively designed tool can make a difference, not just in easing day-to-day tasks, but in opening up new possibilities for connection. I hope that this work can play a part in pushing the conversation forward and making the voice technology landscape more welcoming for everyone.

This endeavor wasn't something I undertook alone. I owe a great debt of gratitude to the people who guided and supported me along the way. My supervisory team—Professor Stella Boess and Professor Gijs Huisman from TU Delft, and Timon van Hasselt from Visio—have been offering insights, encouragement, and the occasional reality check, all of which were invaluable.

I'm also deeply grateful to the experts at Visio—Stefan Laureijssen, Corien van Keulen, and the Visio Ambassadors—who shared their knowledge, feedback, and ideas - not only helping but deeply participating in the research and design process, helping to shape it into what it is today. And a special thanks to Evy and Michel from the Loo Erf at Visio, whose help in recruiting participants and providing feedback on the prototype have helped inform the project in its final phases.

To everyone who contributed to this project, whether directly or indirectly, thank you. I feel very lucky to have had the opportunity to work with such astute, passionate and dedicated people.

~ Konrad Krawczyk

Table of Contents

Abstract

This research project investigates the challenges faced by blind and low-vision (BLV) individuals in using digital technologies, particularly focusing on the usability and accessibility of voice assistant technologies. The study aims to identify key problems experienced by BLV users, analyze the limitations and recent developments in voice assistant technology, and co-create a solution aimed at enhancing the usability of voice assistants for non-visual users.

To achieve these objectives, a mixed-methods approach was employed, including a literature review, usability inspections, digital anthropology, as well as co-design sessions with Visio experts and prototype testing with BLV users to validate potential solutions. The research further explores how these solutions can be integrated into the existing support network, ensuring that the design not only addresses immediate usability concerns but is also sustainable and adaptable.

The findings highlight critical design tensions and opportunities for future iterations of the prototype, particularly in terms of attainability, trust, and user agency. This research provides valuable insights for designers and stakeholders aiming to create more inclusive digital tools for the BLV community, offering a roadmap for embedding these solutions within broader support systems.

Key Terms

- **BLV (Blind and Low-Vision):** used to describe individuals who experience significant vision loss, ranging from partial sight to total blindness.
- **Voice Assistant:** A software agent that can interpret and respond to voice commands, performing tasks like sending messages, making calls, playing media, and providing information from the Web. Popular examples include Siri, Alexa, and Google Assistant.
- **LLM (Large Language Model):** An artificial intelligence model trained on vast amounts of text data. It can generate human-like text, translate languages, write different kinds of creative content, and answer questions informatively. Examples include GPT-3 and GPT-4 (the models behind ChatGPT).
- **HCI (Human-Computer Interaction):** The study of how people interact with computers and how to design computer systems that are easy to use and understand.
- **Conversational Agent:** A computer program designed to simulate human conversation, often through text or voice interactions. Chatbots and virtual assistants are common examples.

1. Introduction

This section introduces the project with its general goal, the context and stakeholders, as well as the design methodology that gives structure to this report.

1.1. Project Objective

The problem I am seeking to address is the increasing difficulty visually impaired individuals face in navigating a world that is becoming more digital and visually-oriented. The progress of voice assistants presents an opportunity to improve accessibility of digital resources for people with visual impairments.

The underlying objective of my research from the start is to:

> *identify ways in which emerging non-visual interfaces can be used to expand access to digital services to those who cannot see screens, blind and low-vision people in particular.*

The following sections of this report will outline the context and the methodology of the research. the particular research angle that has been chosen for this project, which is the context of independent education support of blind and low-vision people.

1.2. Context

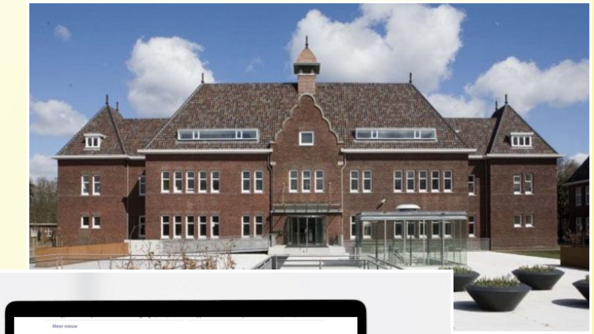


Figure 1.
Visio headquarters
in Huizen, NL

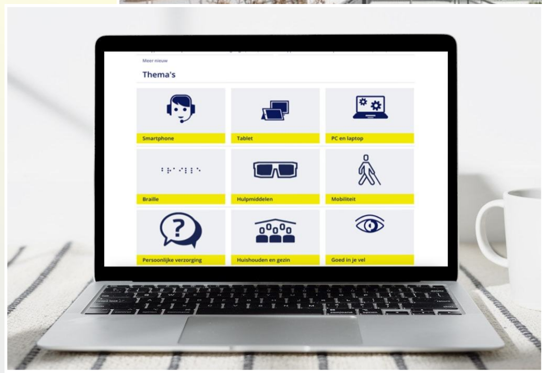
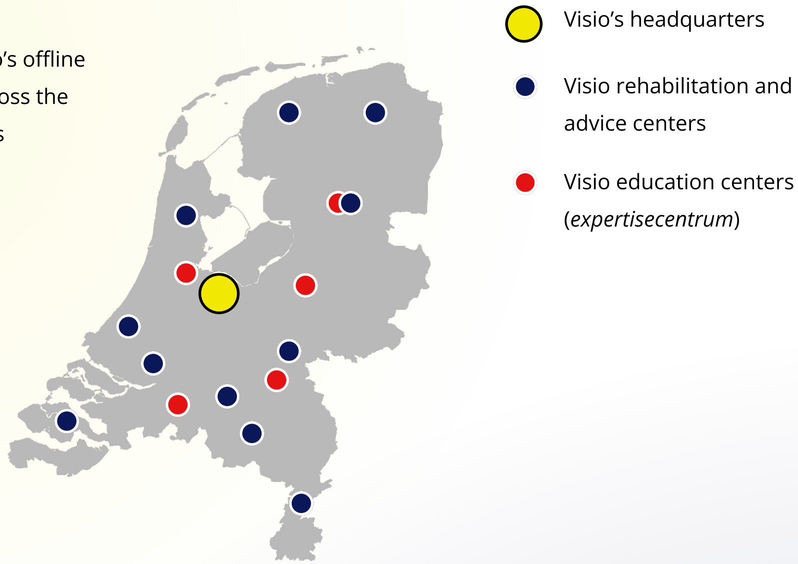


Figure 2.
Example of online
services of Visio
(knowledgebase)

The main supporting institution in this project is Koninklijke Visio (Royal Visio) - a Dutch organization dedicated to providing **expertise**, **rehabilitation**, and **care** services for individuals with visual impairments. Established in 2009 from the merger of De Brink, Sensis, and the original Visio, it is the current legal successor to the Royal Institute for Education of the Blind in Amsterdam, founded in 1808. Visio's administrative and business office is located in Huizen, with additional locations spread across various cities in the Netherlands, including Amsterdam and Rotterdam.

Figure 3.
Map of Visio's offline services across the Netherlands



Visio's mission is to **facilitate active participation** for people who are visually impaired or blind. This includes a wide variety of offerings such as full-time care facilities, centers for independent living, rehabilitation centers, individual coaching, and online resources like a knowledge base, helpdesk, podcasts, and newsletters. These services are tailored to meet the client "in the middle" - providing a broad range of support and accounting for what's best for the individual. This way, Visio can also offer support to individuals who may have additional intellectual, physical, or sensory limitations.

This commitment to personalized care directly informs the methodology and structure of our project, which will be detailed in the following section.

1.3. Methodology and Structure

The project is conducted in collaboration with an external organization. Therefore, the methods I selected for this process have been triangulated with the company to ensure compatibility, as well as to increase context sensitivity.

1.3.1. Inclusive Design

Accessibility has traditionally followed two paths: designing specialized aids for disabled individuals, like white canes, and creating versatile environments beneficial for all. The latter approach is often called *inclusive design,* and it is a methodology that I chose to inform this study. This approach to design is aimed at creating products and services accessible to a wide range of users, including those with disabilities or diverse needs. It involves conducting user research, involving diverse perspectives in the design process, and prioritizing accessibility through features like customizable interfaces and alternative input methods (What Is Inclusive Design?, 2024).

The core principle of Inclusive Design is:

> *solve for one, extend to many*

disability type | combination of constraint | person

similar contexts | unknown audiences

Applying this principle not only expands the potential user base but also **promotes innovation and creativity**. Designing for disabled people forces specific constraints, which often necessitates a breakaway from the habitually replicated norms of the able-bodied. In that sense, the disabled community often adopts technologies earlier, since the potential use value can be higher. Access to non-visual control in a specific circumstance can be useful and convenient for a sighted person, but for a blind or low-vision person can make a difference between being able to use a tool or being excluded from it altogether.

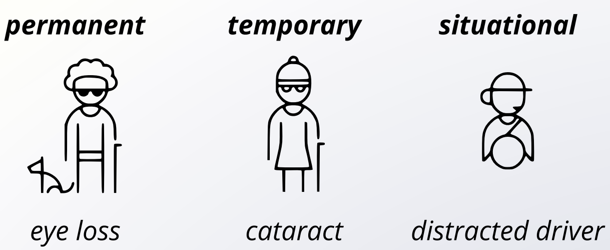


Figure 4.
Examples of disability types addressable in inclusive design (UXCellence, 2018)

Two pitfalls of this method that are important to pay attention to are:

- **Prioritization.** Despite progressive awareness and regulation, people with permanent disabilities are subject to long-term disadvantage and exclusion. Therefore, I believe that listening to and solving problems with them is more urgent than extending the solution to a general audience. The latter, in the spirit of inclusive design, may follow as a secondary outcome.

- **Diversity.** When “solving for one”, one must remember that BLV people cannot be reduced to one monolithic group. Personal preferences, cultural background, class and level of adaptive expertise should be considered to allow for a better approximation of the needs of Visio clients and other groups that may benefit.

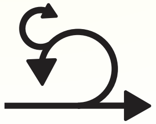
1.3.2. Research Process

The project is divided into three phases: Exploration, Understanding and Integration. I start each of them by addressing questions from the problem space* and finish it with a solution prototype that aims to address the problems discovered in that phase. A detailed outline of each phase, including their respective lengths, objectives and methods, can be found on pages 11-12. Key design terms used throughout this process overview are:

- **Problem space:** the conceptual area, and related research activities, aimed at exploring the design context, identifying target audiences and their underlying needs, as well as potential problems that can be approached with a design intervention
- **Solution space:** the domain where designers generate, explore, iterate and evaluate potential solutions to the identified issues.
- **Diverging:** the process of expanding the idea space through expansive research or concept creation,
- **Converging:** the process of synthesizing findings and narrowing down possibilities to arrive at well-defined objectives or solutions.

The process is:

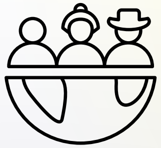
Contextualized. Since Visio caters to specific client needs, this methods used in this project are aimed at deep understanding of user needs. My approach here is to work in a longitudinal fashion with small samples. It enables slow, qualitative research, which elicits insights beyond mere evaluation.



Iterative. Research at Visio is conducted in an experimental and quick fashion. Explorative approaches and concept prototyping are intertwined, and inform each other. The two modes of operation often overlap. This means that activities from the problem space (such as expert interviews) are conducted in tandem with exploratory probes which would otherwise pertain to the solution space. Vice versa, findings from user tests can inform the future problem space.

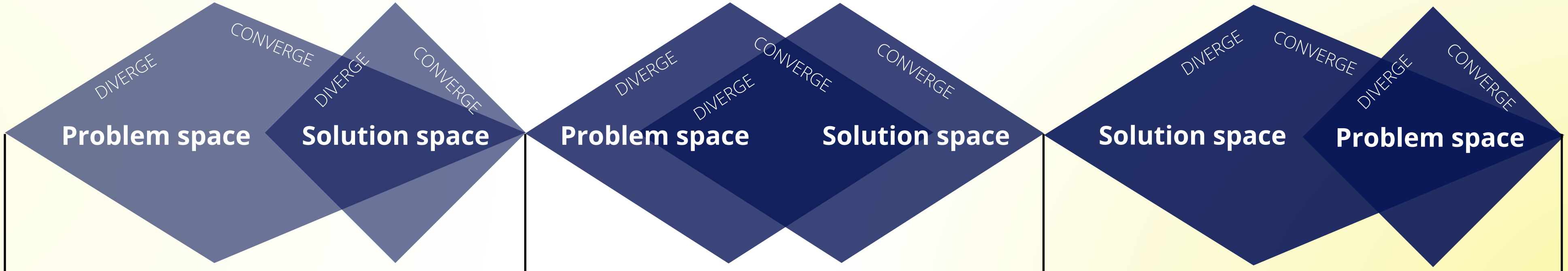
As visualized on the next page, the problem-to-solution-space ratio is not perfectly even through each phase. For example, phase 1 is more focused on exploring the problem space, while phase 3 is more aimed at integrating insights into a solution.

Opinionated. The experts at Visio offer decades of experience, both lived and professional. Therefore, existing insights and principles about needs of the BLV community are integrated from the start. This comes at a risk of premature framing, but with good judgment, it's a risk worth taking.



Cross-cultural. Although Visio operates in the Netherlands, part of its mission is to foster partnerships with similar institutions of this type in Europe. Besides, the rise in migration within and into the EU is a significant factor in who exactly is going to benefit from this research and design intervention. For that reason, I decided to internationalize the research by sampling test participants from multiple countries, and making cultural sensitivity a factor in the following analysis.

The next page shows an outline of the research process.



Phase 1: Exploration

Objectives (problem space):
- To examine the key problems faced by blind and low-vision people in the context of digital technology use
- To identify the key characteristics, limitations and recent developments of voice assistant technology
- To consult the strengths and weaknesses of the voice assistants as experienced by BLV people
- To determine the key problem in the context, then formulate a goal

Objectives (solution space):
- To investigate the possible ways of fulfilling the design goal

Methods used:
- Literature review
- Usability inspection
- Digital anthropology
- Early solution probe

Phase 2: Understanding

Objectives (problem space):
- To understand and formulate the key usability concerns of technology learning for BLV people
- To situate the research in the broader context of support provided by Visio, including all potential stakeholders

Objectives (solution space):
- To gather feedback on potential concepts over different parts of the customer journey
- To distill a design manifestation with the greatest potential of fulfilling the goal
- To identify opportunities and risks of the proposed design for a future iteration

Methods used:
- Co-design with Visio experts
- Think-aloud concept validation
- Prototype testing (interviews and quantitative responses)

Phase 3: Integration

Objectives (problem space):
- To formulate a list of key design tensions to resolve in future research and development

Objectives (solution space):
- To test the solution with the core target audiences
- To embed the tool into the existing network of services provided by Visio
- To gather extended findings from target audiences similar to Visio for better extrapolation

Methods used:
- Conversations with Visio coaches
- On-site user tests in Appeldoorn, NL
- Online tests with the members of the BLV community in the EU

2. Exploration Phase

This section describes the takeaways from research activities performed during the first phase of the project. Starting with the literature review and the overview of the current state of the art, then discussing the user needs based on key insights from generative research and early tests.

2.1. Introduction

The *Exploration* research cycle described in this chapter is guided by the following objectives:

- To examine the key problems faced by blind and low-vision people in the context of digital technology use
- To identify the key characteristics, limitations and recent developments of voice assistant technology
- To consult the strengths and weaknesses of the voice assistants as experienced by BLV people
- To determine the key problem in the context, then formulate a goal
- To investigate the possible ways of fulfilling the design goal

Methods I used:

Literature Review

Overview of Existing Solutions

Mapping User Mindsets

Concept Testing

Community Conversations

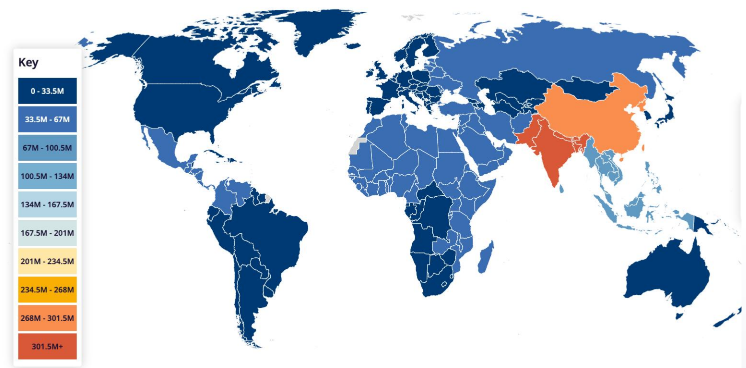
Digital Anthropology

Ideation & Context Scoping

Figure 5. Assistive Tech Fair in Utrecht: Visio showcase (top left), tactile phones (top right), accessible wall clocks (middle), Braille keyboards (bottom)

2.2. Literature Review

Figure 6.
Blind and low-vision people across the world (WEF, 2021)



2.2.1. Blindness

An estimated that 38 million people worldwide are blind, with a significantly larger number suffering from low vision, defined as BCVA in the range of 6/18 to 3/60 (Pascolini and Mariotti, 2012) . Blind and low-vision people (BLV) face reduced levels of independence, safety and productivity (Thylefors, 1995). The reasons for that are complicated, but they can be framed in two ways. On one hand, having limited access to the visual realm prohibits certain activities to be safely engaged in (for example driving). On the other hand, disability is a social phenomenon as well, and the exclusion resulting from a sensory condition is increasingly framed as a type of discrimination by the able-bodied majority, manifesting through ill-conceived architecture, visual-prime interfaces (digital or otherwise), as well as social and cultural stigmatization.

The challenges faced by the BLV community are exacerbated by the associated comorbidities, such as breathing problems, depression, diabetes, heart problems, hearing impairment, hypertension, joint problems, lower back pain and stroke (Goldstein et al., 2012). Comorbid conditions have a negative impact on physical functioning, participation, and health status among older adults with visual impairments (Crews 2006).

Figure 7.
Medical criteria for visual impairment (Shah, 2019)

	Presenting distance visual acuity in better eye	
Category	**Worse than**	**Equal to or better than**
Mild or no visual impairment	__	6/18
Moderate visual impairment	6/18	6/60
Severe visual impairment	6/60	3/60
Blindness	3/60	No light perception

doi:10.1371/journal.pone.0124206.t001

Visually impaired individuals face **discrimination** in various aspects of life. In the housing market, they are significantly discriminated against, with private landlords being more likely to discriminate than real estate agents (Verhaeghe, 2016). This discrimination extends to the workplace, where they experience barriers to employment and promotion (Crudden, 1999). Despite the top-down governmental efforts to outlaw such discrimination, such as the Americans with Disabilities Act or similar legislation in European countries, workplace discrimination against visually impaired individuals persists (Victor, 2017). The types of discrimination they face include a lack of equal opportunities, mockery, and overprotection (Pérez-Garín, 2018).

2.2.2. Sensory Adaptations

Blind and low vision individuals exhibit a range of sensory adaptations, including heightened acuity in tasks requiring tactile information processing (Burton, 2003). These adaptations are facilitated by neuroplastic changes in the brain, such as the recruitment of the former visual cortex for processing tactile and auditory information (Amedi et al., 2005). Sensory integrative treatment programs have been shown to improve mobility, daily living activities, handwriting, and behavior in blind adults (Baker-Nobles, 1977). Furthermore, there is evidence of a hierarchy of perceptual training in low vision, suggesting that individuals with visual impairments can adapt and relearn functional abilities (Patoine et al., 2021).

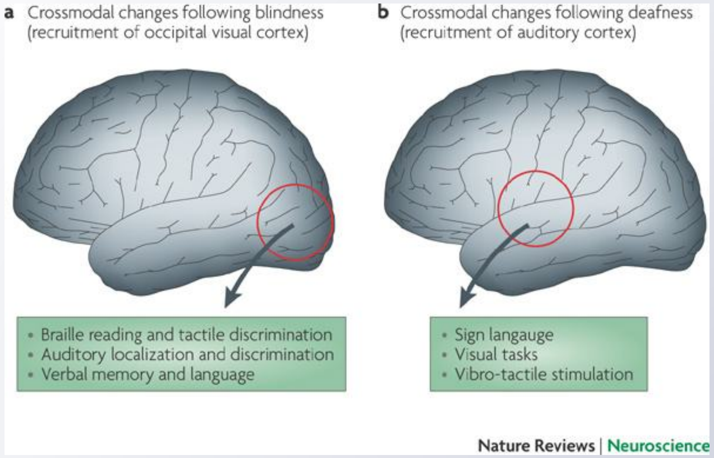


Figure 8. Neuroplasticity after sensory loss (Merabet & Pascual-Leone, 2009)

This is corroborated by the study of Burton and McLaren (2004), which shows that BLV individuals show visual cortex activity during Braille reading. Moreover, object perception can occur on a more conceptual level than what is immediately visible. For example, blind people can understand and even intuitively perceive colors through experiencing their social meaning in daily conversations.

This knowledge intertwines with Bottini et al.'s (2022) idea of conceptual retrieval, where they argue that the absence of visual experience prompts a greater reliance on a "lexical-semantic code." These insights suggest that the design of HCI systems for visually impaired users must not only accommodate unique socialization styles but also prioritize semantic context. The system's ability to accommodate this more semantic-conceptual style could enhance the user-system relationship, making it more intuitive and user-friendly.

This concept suggests that HCI systems designed for visually impaired users may benefit from stimulating the visual cortex through auditory or tactile inputs. In other words, BLV people can still think visually - and it is very helpful for comprehension - if provided with semantic (linguistic, auditory) stimuli that are closely connected to the visual realm.

These cognitive adaptations also result in patterns of interpersonal communication that diverge from the sighted hegemony. As human communication is multimodal, the absence or impairment of visual processing poses a barrier to full perception and understanding of other people, particularly in the non-verbal realm. As just one example, BLV people have limited access to vital conversation-regulating signals, or the co-verbal communication such as head nods, blinking, smiling, etc. Studies show this correlates with a reduced co-verbal output from blind people (Parke, 1980). This means people who communicate through sighted modalities are not receiving those otherwise intuitive signals, which has consequences for collaborative settings that rely on swift and straightforward communication. Further sections of this review show the implications this has on human-computer systems.

Key Takeaways:

- Visual processing function remains active in BLV people
- Sensory and cognitive adaptations in BLV individuals may affect usability of interfaces
- Stimulating the visual cortex through auditory or tactile inputs may enhance user-system interaction.
- The differences in adaptations do not necessarily change the fundamental needs and interests of BLV users of services

2.2.3. Assistive Technology

The impact of information technology on the lives of people with visual impairments is significant, with potential to improve their quality of life (Ashraf, 2016). However, there are barriers to its use, such as accessibility issues and lack of training (Fuglerud, 2011). Despite the potential benefits, individuals with visual impairments tend to use computers and the Internet at rates below the average for the general population, indicating a digital divide (GIWPS, 2021)

Figure 9. An overview of assistive technologies for BLV people (Bhowmick, 2017)

The evolution of digital interfaces for blind and low vision users presents opportunities for enhanced communication and interaction. The advent of spoken language interfaces provides a natural and efficient means for these individuals to engage with technology (Young, 1998). The flexibility and cost-effectiveness of this method make it a viable solution, even more so with the development of context models that can understand and correct misinterpreted words.

Furthermore, the introduction of touchscreen technology has opened up additional avenues for blind and low vision individuals to interact with digital devices. For instance, the Gesture Avatar project has enabled users to trace shapes or characters on the screen that correspond to the widget or text they wish to activate (Grussenmeyer et al., 2017). This research has led to faster and more efficient ways of interacting with touchscreens, with fewer errors than traditional methods. VoiceOver technologies from Apple and TalkBack from Android have also shown promise in improving accessibility for visually impaired users. Notably, dictation has proven to be five times faster than keyboard input, despite its error-prone nature (Grussenmeyer et al., 2017). Notwithstanding the problems with voice input, touch screens remain incompatible with the communication modalities of blind and low vision people. Typing rates for people with visual impairments tend to be slow, averaging around 4 words per minute (Grussenmeyer et al., 2017).

Touch screens and the applications designed for them utilize affordances which enable quick orientation for sighted individuals, but become far less intuitive when perceived through sound only. Not all content served through touch screens can easily be dictated. For example, tables and charts are notoriously difficult to represent through non-visual modalities. There have been attempts to do so, with varying degrees of success. However, tests of such tools often do not include blind and low-vision individuals who would benefit from them the most.

In one study, the use of spatial sound cues for tables on webpages has shown no improvements and sometimes worse performance in sighted subjects, and no tests have been done on visually impaired individuals (Grussenmeyer et al., 2017). Furthermore, the existing legally compliant accessibility options present a steep learning curve for some users, barring them from accessing information online. The study by Grussenmeyer et al. (2017) revealed that many visually impaired users found the built-in tutorial for TalkBack difficult to complete and learn from. Users also struggled with the inconsistency of screen layouts, even within the same app.

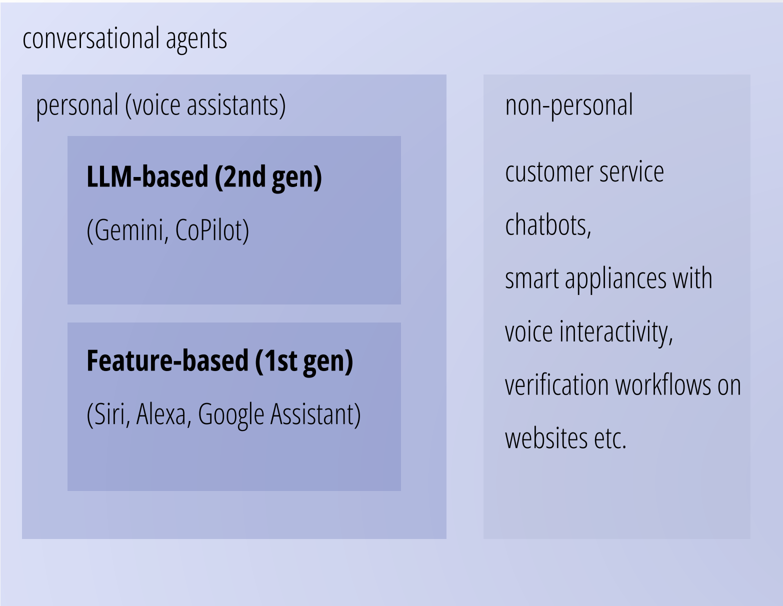
Key Takeaways:

- BLV people benefit in particular from access to digital technology, yet they are disproportionately excluded from participation
- Touchscreens, despite good efforts towards accessibility, are largely incompatible with the modalities preferred by BLV people
- Tests of auditory interfaces are often conducted without BLV participants
- Sonification of data through spatial sound cues show mixed results
- Onboarding for assistive touchscreen features can be difficult, presenting a steep learning curve

2.2.4. What Are Voice Assistants?

In order to properly scope this research, it is vital to define and typify the various kinds of “talking” devices. **Conversational agents** are automated dialogue systems that uses natural language processing (NLP) to interact with humans in a natural, conversational manner. **Voice assistants** are a subset of conversational agents which are largely operated through voice and tend to offer a more personalized experience specific to a device owned by the user. Among these, two types emerge (more in section 2.2.1) - the **1st generation** of feature-based assistants with predefined interactions (e.g. Siri), and the emerging **2nd generation** of open-ended assistants based on large language models (LLMs).

Figure 10. Overview of conversational technologies (2024)



This section examines three aspects of design of voice assistants: trust, agency and disposition. A broader overview can be found in section 2.3 “State of the Art”.

Trust

Voice assistants involve multiple layers of trust that users must navigate. Users exhibit varying levels of trust towards the voice assistant itself, the platform provider (e.g. Amazon for Alexa), the automatic speech recognition technology, and the intended data receiver (e.g. a doctor receiving health information) (Wienrich, 2021). Perceived expertise and framing of the voice assistant as a "specialist" versus "generalist" significantly impacts user trust, with specialist framing increasing trustworthiness across all layers (Siri and Alexa Reimagined, 2023). Personality similarity between the user and voice assistant also plays a role, with users finding similar personalities more trustworthy and being more resistant to misinformation from such assistants. However, a notable portion of users still express an inherent lack of trust in voice assistants for tasks like shopping (PwC, 2018)

Agency

Conversational interfaces on mobile devices often employ anthropomorphic design elements that imbue the device with a sense of agency. This agency can influence user perceptions and behavior towards the interface.

For example, Martin et al. found that users engaged in more social talk and expressed more positive sentiment when interacting with an anthropomorphic conversational agent compared to a non-anthropomorphic one (2020). Porcheron et al. observed that users treated a conversational agent as a social actor, asking it questions about its opinions and capabilities. Lister et al. highlighted the carefully designing the agent's persona and language to manage user expectations (2017). However, Luger and Sellen cautioned that overly anthropomorphic agents can lead to a "gulf" between user expectations and the actual capabilities of the system (2016).

Disposition

Each designed entity has a disposition, i.e. a tendency to afford a specific interaction (Easterling, 2021). This is highly pronounced in AI-driven voice assistants, which rely heavily on the data they are trained on. As highlighted in the CNPEN opinion, conversational agents lack true understanding and simply execute programmed functions based on their training data. (2021) If this data contains false or biased information, the agent may exhibit undesirable traits or tendencies. It is observable that LLMs have a predisposition towards certainty and business-friendly cadence, which can be attributed to the fact that the available training data online is biased towards texts that provide “comprehensive” answers, instead of dealing with uncertainty in a human-like, fallible fashion.

2.2.5. Voice Assistants and Blindness

The emergence of voice assistants has revolutionized interaction with digital interfaces, especially for individuals living with visual impairments. These assistants, such as Siri, Alexa, and Google Assistant, offer a "uniform, screenless interface to a variety of digital apps" (Abdolrahmani, Kuber, & Branham, 2018), providing a much-needed accessible solution.

A significant opportunity lies in the ability of blind users to personify voice assistants, which results in a higher rate of satisfaction (Branham & Roy, 2019). Furthermore, blind people have shown a preference for customization of speech, rate, clarity, and intensity of voice output, indicating an opportunity for voice assistants to offer personalized experiences (Branham & Roy, 2019).

The voice-only nature of assistants is particularly beneficial for blind users, who often lack accessible alternatives for in-depth tasks or error correction. Interestingly, Branham and Roy (2019) note that blind individuals' superior capabilities in performing serial memory tasks and recalling longer word sequences could be advantageous when interacting with assistants. This skill lends itself to the use of prolonged voice responses with more keywords and list items.

Despite the opportunities presented by assistants, the effective usage of these interfaces by blind and low vision users is not without substantial challenges. Misinterpretation of commands, for instance, is a major barrier, leading to unexpected actions being taken by the voice assistant (Abdolrahmani, Kuber, & Branham, 2018). This can result in frustration and a lack of confidence in assistants, which can in turn deter individuals from fully utilizing these tools.

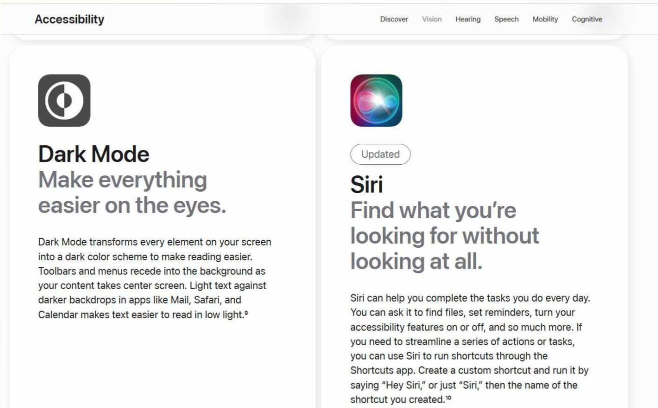


Figure 11. Accessibility features in the Apple ecosystem (Blind Girl Adventures, 2022)

Moreover, assistants often struggle with accurately understanding and transcribing accents and colloquial phrases. This challenge is further compounded when users attempt to compose long messages, with assistants often failing to accurately transcribe these (Abdolrahmani, Kuber, & Branham, 2018). Such limitations can lead to mistrust and dissatisfaction with the technology, ultimately undermining its utility for blind and low vision users.

Feedback provided by assistants has also been identified as an area of concern. Often, the responses generated by these assistants are either too verbose or irrelevant, leading to difficulties in obtaining concise and usable information (Abdolrahmani, Kuber, & Branham, 2018). This not only slows down the interaction process, but can also lead to confusion and further mistrust.

Another challenge lies in the lack of support for adjustable speech rates. This feature is a basic preference that cross-cuts task type and has significant implications for usability (Branham & Roy, 2019). Without the ability to adjust speech rates, users may find it difficult to understand or keep up with the assistants, particularly when dealing with complex responses or commands.

Key Takeaways:

- Voice assistants are providing a (mostly) screenless experience, making digital services more accessible
- Voice assistant feedback often lacks conciseness and relevance, causing interaction slowdowns and confusion.
- Lack of adjustable speech rates hinders understanding, especially with complex responses.
- The "human-human conversation" model in assistant guidelines may not consider the needs of blind users, favoring ableist assumptions.

2.3. State of the Art

2.3.1. Overview of Current Solutions

Objectives:

- to map out various types of non-visual human-computer interfaces
- to see what's being used by the target audience
- to identify usability issues with the most commonly used solutions

Methods:

- Online research (YouTube, podcasts, company papers)
- Usability Inspection (Nielsen Norman's Heuristic Evaluation)

As outlined in the Literature section, non-visual HCI has a long history. However, the interfaces designed specifically for BLV people make only one part of the story. A great deal of interfaces have been created with sighted persons in mind as well. There are various reasons someone might need to interact with a computer without looking, for example: driving, cooking, or showering. Non-visual HCIs have also been designed for niche audiences. There are very specific assistive solutions available for people who are excluded from multiple modalities, for example deafblind people. These interfaces also afford different degrees of complexity, with some allowing only very basic communication use cases, and others affording a full range of advanced tasks that a visual interface would. The following graph maps these differences.

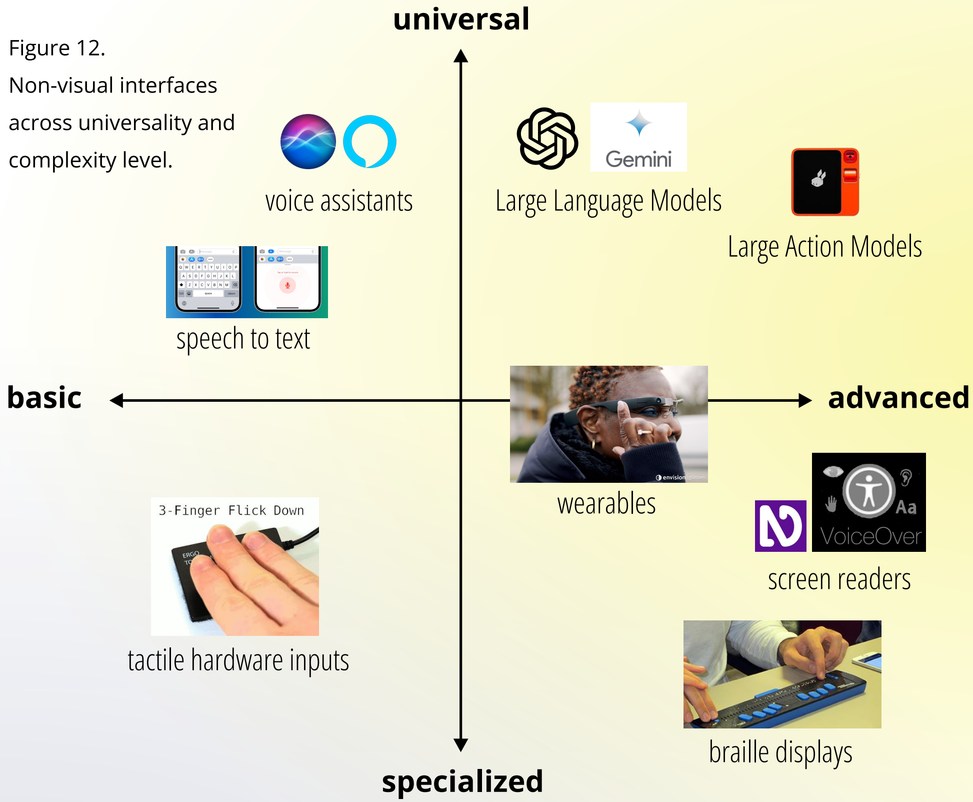


Figure 12.
Non-visual interfaces across universality and complexity level.

Conversational interfaces fall greatly within the universal quadrants, making them particularly compelling within the framework of **inclusive design.** Mainstream consumer devices like smartphones and tablets are increasingly equipped with accessibility features that can aid people with visual impairments, making them viable alternatives to traditional assistive devices. These mainstream devices are often more affordable, less stigmatizing, and widely adopted compared to specialized assistive tech. Therefore, it is essential to investigate the existing, mainstreamed technologies in further detail.

2.3.2. Usability Inspection

According to Astute Analytica (2024), the number of voice assistant devices worldwide is to double from 4.4 billion in 2022 to 8.4 billion in 2024. That's more than there are people on Earth. Yet, the market of operating systems for those devices is heavily polarized. The three most commonly used systems from the time of writing are: **Google Assistant**, Apple's **Siri** and Amazon **Alexa**. The newest contender from Google called **Gemini** is also increasingly being rolled out to the 3.6 billion Android users worldwide.

I have selected those four aforementioned assistants for usability testing, using the Nielsen Norman's heuristic evaluation framework (Nielsen, 2024), assessing the severity of usability issues across 10 dimensions, with emphasis on **voice-only** control. The full description the testing protocol and the particular severity scores for each device can be found in Appendix I.

A cumulative average score has been calculated, from 0 to 4, where 0 means no issues at all, and 4 means high-severity issues with usability.

Figure 13.
An overview of usability inspections across common voice assistants (0=no issues, 4=severe)

Google 2.0/4

− reacts differently to the same question − does not answer questions in context − cannot always read results out loud	+ proactive help and discovery of options + reads real-world data

Siri 2.2/4

− lack of flexibility of output − not remembering context − no access to settings from voice level	+ sound feedback + correct error recovery + reads real-world data

Alexa 2.3/4

− limited access to external information − not remembering context − no voice access to its own settings	+ good understanding of real-world data + attempts at help and discovery

Gemini 1.3/4

− limited access to real-world data − occasional misunderstandings of intent − no voice access to its own settings	+ very relevant and customizable answers + context-aware + exceptionally good help and discovery

Key Takeaways from the Usability Inspection:

- Significant improvement in usability is observed in the LLM-based Gemini, compared to the first-generation assistants.
- All major assistants except Gemini fail to remember context.
- All major assistants fail to provide the same answer for the same user intent.
- No assistant provides flexibility of use and personalization through voice.
- The first-gen assistants currently have broader real-world capabilities.
- LLM-based answers drastically improve the help and discovery features.

One must keep in mind the limitations of the method. Usability inspections usually do not involve actual users but designers, which means they might miss real-world usage problems that only emerge during user interaction (Rochat et al., 2022). Another limitation is that these methods can overlook contextual factors affecting usability, as they are typically conducted in controlled environments rather than real-world settings (Bias, 2015). Furthermore, there are pitfalls of this method that are particular to testing voice assistants. Multifunctional interfaces of that type are often not designed to only respond to pre-defined inputs, instead using predictive algorithms to understand intent. That leaves a great deal of interaction flow to chance. Nevertheless, the usability shortcomings and comparable improvements across platforms have been unmissable, and will certainly inform my further research directions.

2.3.3. Future Speculations

Objectives:

- to design an intervention that adapts with the progress, instead of becoming obsolete
- to get inspired by the already released upcoming technologies

Methods:

- Online research (YouTube, podcasts, company papers)

The idea of "talking to a computer" has captivated minds of designers since the very early days of computing. The idea of conversational interfaces in particular has left a mark in popculture, from science fiction literature (I, Robot) to film and TV (Her, The Jetsons). It is part of a much broader post-humanistic tradition, which goes beyond the scope of design and into philosophical contemplation.

Below is a brief timeline of select important recent developments in conversational computing.

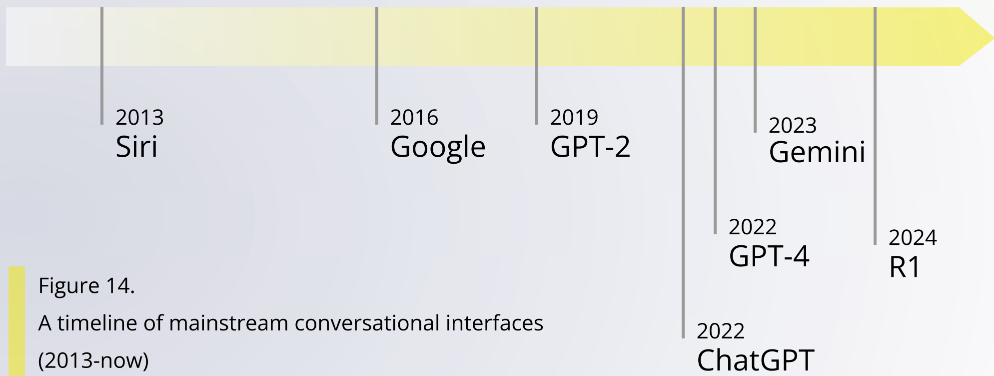


Figure 14.
A timeline of mainstream conversational interfaces (2013-now)

Some of areas I identified as the most promising are:

- **Large Language Models.** After a decades-long history of failed attempts at human-quality conversational interfaces, the last two years have seen a mainstreaming of chatbots, with ChatGPT having the fastest growing user base for any product in world history.
- **Multimodal AI.** Newest algorithms move beyond language, and are able to mix data from voice, images, and videos and maintain understanding across modalities. A showcase of GPT-4o demonstrated that generation of natural-quality voice and other sound effects is possible in a chatbot.
- **Large Action Models.** The main caveat of LLMs currently is its limited interaction with visual interfaces. Large Action Models such as the system on voice-only Rabbit R1 offer promising solutions to this issue, as they are being explicitly trained to understand user intent and act on it within an interface.

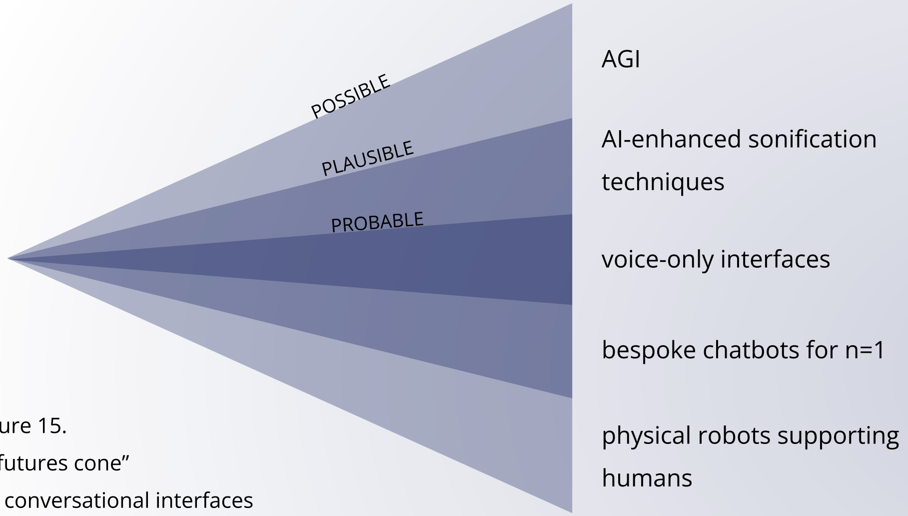


Figure 15.
A “futures cone”
for conversational interfaces

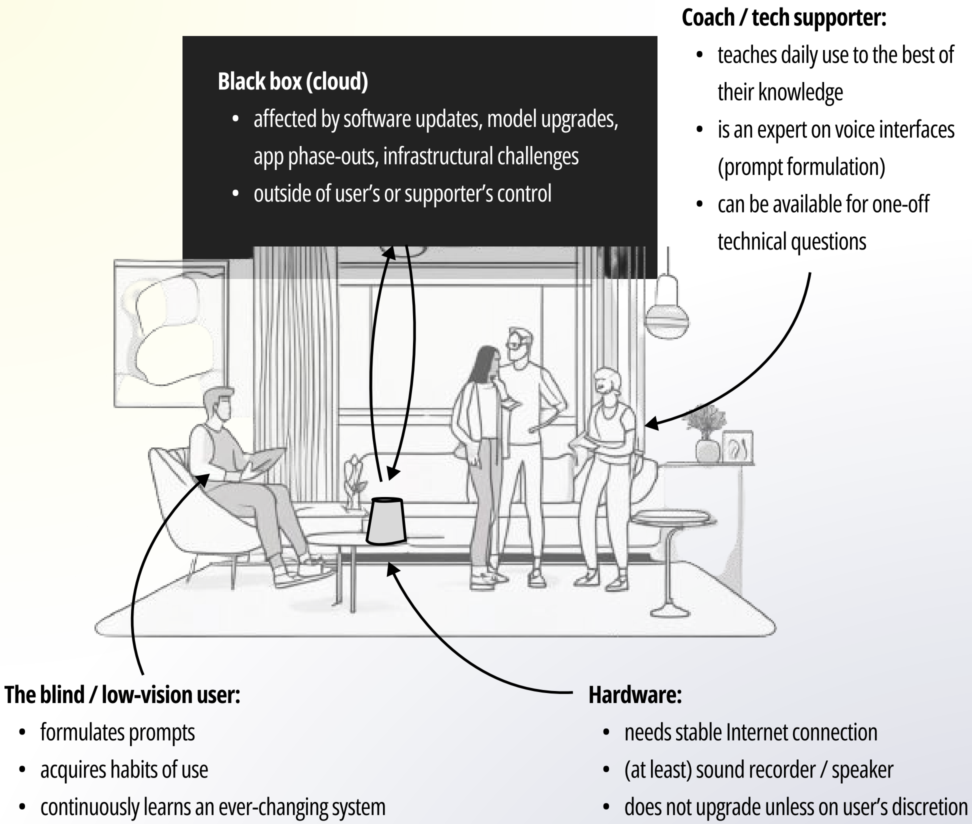
2.3.4. Context of Use

The insights from literature, usability testing, and future speculation, provide an approrpate overview for the potentials and shortcomings of voice assistants today, as well as the exciting changes to arrive soon. However, for the purposes of this project, I have to be strategic about what is the **feasible maximum improvement** I can create with the community involved. The project spans 100 working days.
The question I ask myself is:

> *What can be done within the given time frame to create a notable, positive impact on the quality of interactions between Visio clients and their voice assistants?*

Voice assistants are used in physical space. They often have co-users, especially by BLV people, in case a human supporter is involved. Besides that, they are overwhelmingly connected to broader ecosystems - usually accessed through cloud services of respective tech companies - and are therefore influenced by their design and business decisions. Importantly, most mainstream voice assistants are blackboxes. A "black box" is a device or system whose internal workings are hidden or not fully understood by the user (Guidotti et al., 2018). It is a metaphorical representation where the internal mechanisms are not known or considered, and the focus is solely on the system's behavior in response to external stimuli (Guidotti et al., 2018).

Figure 16.
The voice agent ecosystem
in the context of use



The above has important implications for the scope of the project. In the case of voice assistants, deliberate opaqueness is part of the business logic on behalf of the technological oligopoly. IT companies such as Google or Apple update voice assistants remotely, without explicit user consent, without an option to roll back, and without transparency (e.g. actively announced release notes).

Theoretically, this could lead me to attempt creating a new, open-source voice assistant which addresses those issues in a more inclusive way, untangled from priorities of big business. However, switching cost prevents me from doing that. Broadly, it is defined as the labor or financial cost spent on switching from one system to another. I consider it both on the individual and the macro scale. For example, for a BLV person who already has an iPhone, switching to an Android phone with an open-source assistant requires buying a new device, then obtaining tech support with untangling the technical limitations imposed by Google, and installing the system. At macro scale, open-source accessibility options are unlikely to break through the oligopoly of Apple and Google. There are billions of pre-existing devices already in circulation, millions of which are owned by BLV people. Those come with well-documented (if flawed) accessibility options, which is a significant pull factor for BLV people to use them anyway.

Mindful of that, I continue the project with the acceptance of things I cannot change (at least within 100 days), and the urgency to change the things I can. The following pages will outline the research activities that helped me map out the problem space of the context.

2.4. Identifying User Needs

2.4.1. Expert Session

Objectives:

- to get to know the adaptations of BLV people
- to identify key pain points in current assistive technology
- to envision a best-case scenario for voice assistants in the future.

Methods:

The process involved a co-creation session in a group of four. The co-creation was based around the Path of Expression by PJ Stappers, starting with the present, going into the past experiences and impressions, and finally moving into hopes for the future. In the second part, we envisioned what assistant could be like **"in an ideal world".**

Experts:

All participants are affiliated with Visio. The session involved 4 participants, including one blind technology coach, one low-vision designer, one sighted technologist and myself.

Stefan
Tech Coach
at Visio

Timon
IT Advisor
at Visio

Corien
Design Researcher
at Visio

Key insights from the session can be found on the next page.

1. Feedback Methods are Inadequate:

Current feedback methods are often merely numeric benchmarks and cater to executives rather than the user base.

2. Voice Assistants Lack Personalization:

Voice assistants struggle to provide tailored responses, often offering too much or too little information, which can be overwhelming. Blind users have almost no way to change voice assistant settings using voice.

> "My dream is that the voice assistant guides users to the right steps. It explains while it's guiding you."

> "People come up to me worried they are using voice assistants wrong. Why can't we flip this dynamic and let the smart device ask for feedback as well?"

3. Accessibility Features are All-or-Nothing:

You either opt into the full screen-reader experience, or you have your voice assistant not say half the words out loud. No middle ground.

4. Each BLV user has a different mindset

Successful use of technology depends largely on the specific user's personality, assertiveness, and goals.

> "It's better to use mindsets than personas for design. We are all different people who want different things"

5. Unnecessary Information Creates Static:
Information that is helpful in visual interfaces can become distracting static in voice interfaces, complicating the use for blind individuals.

> "It's like a new language. When you're totally using voice,
> You still have to know how a screen works in the first place."

6. Disruptive Software Updates:
Software updates that alter interaction methods can be extremely disruptive for blind users, whose workflows depend heavily on consistent and predictable interfaces.

> "The smart assistant is the dumbest thing there is. It's only called smart because it couldn't be done by anything before"

8. Desire for Conversational Interaction:
Blind users prefer a more conversational interaction with voice assistants, where the assistant guides them through steps and responds in a more human-like manner.

> "My dream view of how voice could be used is like a comprehensive conversation with the device."

9. Need for Better Training and Onboarding:
Voice assistants and other assistive technologies require a learning curve and proper onboarding to be effective for blind users, who need time to adapt to new interfaces.

2.4.2. Digital Anthropology

> "I have a lot of problems with you, people!"
> ~ That Real Blind Tech Show

Objectives:

- to understand needs, hopes, frustrations and concerns of people who already use assistive tech
- to map out areas of common interest for the BLV community
- to collect information on technologies that are commonly used by BLV people

Methods:

- **Topic immersion** with podcasts and other online media.
- I have obtained 812 recent posts from a Reddit forum **/r/blind**, using the official and compliant API. Posts were then clustered into topics using LDA analysis. Then, each post was analyzed with regard to its topic and general sentiment using an LLM. Named entities, such as brands, products, or locations, were also extracted with an LLM. Full description and results available in Appendix II.

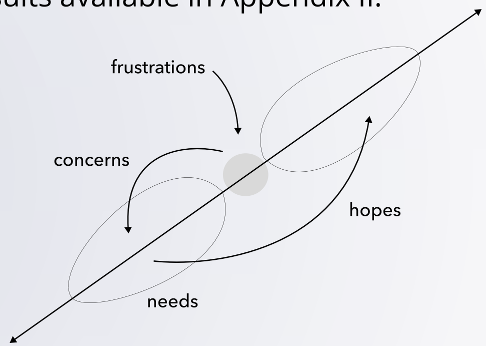
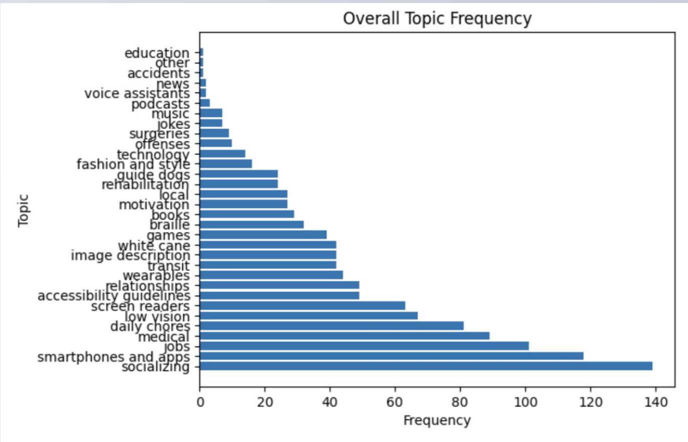


Figure 17.
A modified version of
The Path of Expression
(Sanders, E. B. N., & Stappers, P. J. (2012)

Key Topics:

- **Accessibility in technology** is a major nexus of interest, with users seeking information and advice on accessible technology, websites, apps, and public spaces. Technology plays a crucial role, with discussions about screen readers, braille displays, magnifiers, and GPS apps. Users often voice their frustrations with inaccessible technology.
- **Spatial assistance** is equally important, though usually non-digital. That includes canes, smart glasses, and magnifiers, but also crucially guide dogs.
- **Health conditions** are a frequent topic, with users sharing experiences and seeking advice on RP, glaucoma, macular degeneration, and cataracts. That also includes mental health, with users sharing struggles with depression, anxiety, and coping mechanisms.
- **Practical matters** take up a large portion of the posts, with users seeking tips on cooking, cleaning, organizing, and finding accessible hobbies. Employment is a major topic, esp. career options and accommodations. Honest discussions of discrimination also appear.

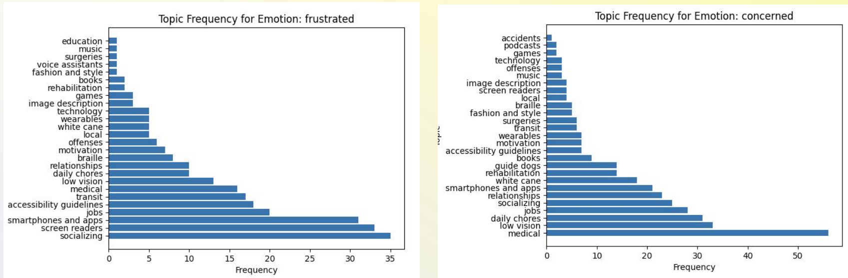
Figure 18. Topic frequency on /r/blind



Key Sentiments:

- An overwhelming sentiment with regard to **technology** is **frustration**. Especially high levels of frustration are observed in sub-topics such as screen readers, accessibility guidelines, smartphones and apps, and socializing.
- In **medical** topics, the overwhelming emotion is **concern**.
- Users are **curious** about **socializing** and **job** opportunities.

Figure 19. Topics across emotion on /blind



Key Takeaways:

- Technology, although important for BLV people, is a major source of **frustration**. This is especially true for screen readers.
- **Socializing** and relationships are the most important topic for BLV people.
- **Career** opportunities are an underrated concern, with technology-related jobs mentioned especially frequently.
- Voice assistants are mentioned as useful for **basic tasks**, though limited in their capabilities. Same goes for wearables.

Quotes from Redditors:

google voice (to text) is lousy. it also does nut understand interpunction or line breaks.

Siri is great for doing basic things on the phone. But when you ask a more complex question it defaults to "here's what I found on the web". Does anyone have a way to make it act more like chat gpt or Gemini? Simple is best. Thanks!

My father is nearly completely blind, he just turned 70 and has very little dexterity and balance. [..] Are there any apps that siri can have complete control of?

If Siri hears anything in the background after being asked a question, it will just stop mid answer or not even start an answer.

When we tell the phone "Hey Seri, what coulor is this" is either does a Google search or some other nonsense. My question is what is a good user friendly app for blind people and phones, I feel like they should be farther ahead with this.

Alexa could read it motivational quotes to him out loud. Unfortunately they stopped doing that and he really missed that.

Figure 20. Overview of BLV sentiments from /r/blind.

positive

present		future
Needs: Accessible information, affordable assistive technology, employment opportunities, transportation options, emotional support.	**Hopes:** Accessible technology, maintaining independence, connecting with others, pursuing careers and hobbies, receiving understanding and support.
Frustrations: Inaccessible technology, dismissive doctors, lack of understanding from others, limited employment options, dependence on others.	**Concerns:** Vision deterioration, social isolation, discrimination, financial insecurity, safety.

negative

2.4.3. Problem Space

With these findings in mind, I have started mapping the problem space of BLV use of voice assistants, in order to narrow down the scope of this project. The most notably problematic areas are listed below, in the order of where they may appear in the long-term scope of use.

Figure 21. Key usability issues for non-visual users of voice assistants

Beginner

- Setting up and obtaining access to the assistant
- Lack of an onboarding experience or tutorial
- Low discoverability of features on most assistants
- Insufficiently conversational interaction (low responsiveness)
- Frequent misunderstandings of user input
- Lack of voice-only interaction support
- Lack of personalization and access to settings
- Lack of support for advanced use cases (and no clear indication of that by the assistant)

Advanced

To an extent, these usability concerns may be applicable to all users. For example, misunderstandings of user input can be frustrating to a sighted user as well. However, with the lack of non-visual feedback, there are particular usability problems that BLV people are uniquely affected by. In this design, I will prioritize those users to help them **overcome barriers** when possible, and **manage their expectations** when help cannot be easily provided. This will happen along a client journey, as described on the next page.

2.4.4. Mapping the Client Journey

Objectives:

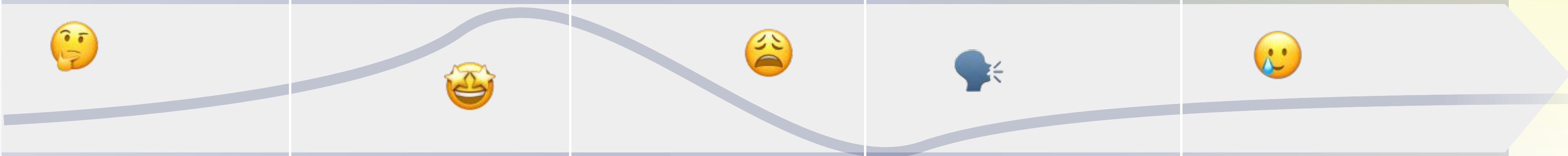
- to understand the process of “getting to know” the voice assistant
- to understand the Visio’s educational context

Methods:

I have conducted a follow-up session with a technology coach from Visio. The session was focused on examining the process of coaching - from the initial hopes and concerns of beginner users, through the followup sessions all the way to independent use. Full overview is available in Appendix III.

Key Ideas:

- **Trust** plays a role. Voice assistants are “uncanny” and require acquaintance.
- The core of tech coaching at Visio is learning to give assistants “**ground rules**”: conditions and boundaries required for them to know how to perform a task.
- Not all tech issues can be addressed with coaching. Due to abrupt product updates and “black-box”-ness of assistants, **problems often can’t be solved**.
- Assistants afford a **hierarchy** of knowledge, which **intimidates users.** Ideally, the human-computer learning process would occur both ways.
- Overcoming the **fear of asking questions** is the first step of coaching.
- **Independent practice** is crucial in successful use of assistive technologies.

	Caution	Consideration	Disappointment	Bargaining	Acceptance
Feelings					
User Goals	keep up with change, perform digital tasks, maintain privacy and safety	connect with people, succeed in digital tasks, get meaningful responses	succeed in advanced tasks, understand why errors occur, decide if the tool is worth it	clarify possible use cases, personalize the device, get support with errors	make the tool part of their routine, where it works
Opportunities	help the user get to know the assistant, promote voice control actively	negotiate “ground rules”, ask for feedback & apply it, add more humane factor	help the user learn to formulate intents, be clear about scope of use	let the user ask for help, enable settings and personalization	remind the user about what is possible, continue support

2.5. Action Statement, Goal, Vision

Objectives:

- to define the core problems shared among stakeholders
- to place further research in a specific context of use
- to decide on a design direction

Methods:

- gathering insights into clusters and formulating an action statement
- narrowing down into a specific design goal
- formulating interaction vision and qualities

2.5.1. Action Statement

The visual-prime bias of digital services disadvantages blind and low vision people. As the potential quality-of-life improvement for them is higher, they advocate for accessible features. Those tools enable access, but their numerous usability shortcomings are a source of frustration for the BLV community.

Emerging voice interfaces present a great opportunity to equalize access. However, those tools are constrained by market logic, inadequate inclusivity, and irrecoverable errors.

Technology coaching helps to ease BLV people into fulfilling their use goals independently. However, human support is not always available, and digital skills also require independent practice. Therefore, it is vital to provide BLV clients of Visio with an opportunity to continue learning on their own terms. Such support must focus on the skills that are within the user control, such as formulating "ground rules", recovering from errors and asking for help. The imminent changes in usability of voice assistants, especially driven by large language models, must be taken into account to create a future-proof solution.

2.5.2. Design Goal

I want to ease people who cannot see screens into using voice assistants independently.

→ I want people who cannot see screens to gain more intuitive access to information online.

2.5.3. Interaction Vision

The process of becoming an independent user of voice assistants can feel like **attending a driving school.**

Theory

provided by a human instructor
explaining concepts and easing
first simulated practice

Practice

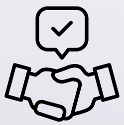
enabled through guided homework
long-term knowledge of prompt formulation patterns
technical support on the go

2.5.4. Interaction Qualities

The desired interaction should feel:

Empathetic
Truly respectful and understanding of the struggle it takes to navigate access to online services for a BLV person.

Opinionated
Honestly addressing the shortcomings and dispositions of the current voice agents, so that the user feels guided in their expectations.



Negotiable
The user should feel like their individual goals and preferences are being acted on, or at least respectfully addressed.

Unpolished
Humans are fallible, so are robots. Empowering the user requires leveling the hierarchy of knowledge, and dropping the pretense of infallibility on behalf of digital voice agents.

2.6. Early Probes

2.6.1. Ideation

Objectives:

- to explore ways of fulfilling the design goal
- to create probes for research and future exploration without committing to a specific design manifestation

Methods:

I have brainstormed initial ideas using a “what if ____ ” method. All ideas are available in the Appendix IV. Further, the ideas have been arranged into clusters and considered as potential probes.

2.6.2. Concept Development

The following clusters were identified:

- **Feedback Loop:** mechanisms to give and receive feedback, both in the form of quizzes and computer responses and from a human.
- **Tech Support:** an always-available resource that provides limited support when human support is unavailable
- **Knowledgebase:** a repository of known apps, services, tips and tricks served in a way that’s available for BLV people
- **Tactile Support:** extending beyond audio to make slow, well-considered prompt formulation more intuitive

2.6.3. User Probe

Shortly after ideation, I have created a probe featuring a simple version of a voice-based educational module that combines the first three of the clusters (feedback, support, knowledgebase). To create it, I have used Voiceflow - a desktop-based prototyping tool for conversational interfaces. It included one simple topic: how to wake up Siri/Google Assistant by saying “Hey Siri” or “Okay Google”. The interaction order:

1. The system introduces the ground rules, including sound cues, and asks if the user wants to proceed with a lesson (user then answers)
2. The system asks the user for which device they are on (user answers)
3. The system introduces the lesson topic, then providing an option to listen to a pre-recorded example or practice (user then selects).

P1

Senior Manager,
Dutch,
legally blind,
highly experienced with assistive tech.
Recruited online (LinkedIn).
Only prototype description provided

P2

Tech Coach,
Dutch,
legally blind,
highly experienced with assistive tech,
recruited through Visio, tested the Voiceflow prototype

P3

Entrepreneur,
Polish,
legally blind,
uses only voice assistants,
recruited through personal networks,
tested the Voiceflow prototype

Quotes and insights:

It would be useful indeed to give independent guidance to people who'd need this tech. Probably would not make sense to make it a separate physical device, though. **- P1**

This is something that Siri and Apple lack. This guide you by the hand approach. It will assume you'll figure it out how it works. **- P2**

This is what I teach people. And if I could skip this and send them this instead, then I can help them with the really big hurdles. **- P2**

For now it's for using an assistant but it could be integrated for other medical use cases **- P2**

Don't "great" me. It's uncultured and condescending **- P3**

The Visio thing is supposed to make my life easier. In order to do that, I have to download Visio. I can't say "OK Google, download Visio". That's a problem **- P3**

Well, would it actually be possible to phone call this assistant? Cause that would be nice **- P3**

Key Ideas:

- The idea of a conversational, voice-only module for learning assistive technology is compelling for the BLV participants.
- The execution of the module needs improvement. One user mentioned a "condescending tone" of the assistant put him off from further use.
- Tone and disposition of the assistant must be culturally appropriate.
- **Access is a challenge.** As the target audience is not made of tech experts, the module should not require pre-existing knowledge of interfaces.
- Further research is crucially important to situate the design goal in the context of coaching, collect ideas, and iterate further.

2.7. Conclusion & Evaluation

This chapter concludes the first phase of the project. The focus of this phase was the broad exploration of the problem space, with initial attempts at iteratively addressing it through probes.

With all effort considered, the findings of this phase have some crucial limitations. In the middle of the phase, the main challenge has crystallized, with the educational context becoming the one to improve. However, all the primary insights gathered to this day come from field experts, or tech-advanced people, who are not the target audience. Besides, the research has to be situated in the spatial context of the expertise activities conducted by Visio. This could provide important information from the stakeholder's side, which this project needs for success. Finally, a more diverse sample of perspectives is needed, accounting for differences in age, gender and origin.

The next part outlines a plan for the Understanding phase of the project, which is going to address these shortcomings.

3. Understanding Phase

This section describes the activities for the second phase of the project, which include co-designing with the BLV experts at Visio, concept prototyping, and evaluation.

3.1. Introduction

The *Understanding* research cycle described in this chapter is guided by the following objectives:

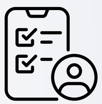
- To understand and formulate the key usability concerns of technology learning for BLV people
- To situate the research in the broader context of support provided by Visio, including all potential stakeholders
- To gather feedback on potential concepts over different parts of the customer journey
- To distill a design manifestation with the greatest potential of fulfilling the goal
- To identify opportunities and risks of the proposed design for a future iteration

Methods I used:

Workshops with Visio experts

Testing real-life scenarios of voice assistant use

Prototype testing (interviews and quantitative responses)



Think-aloud concept validation

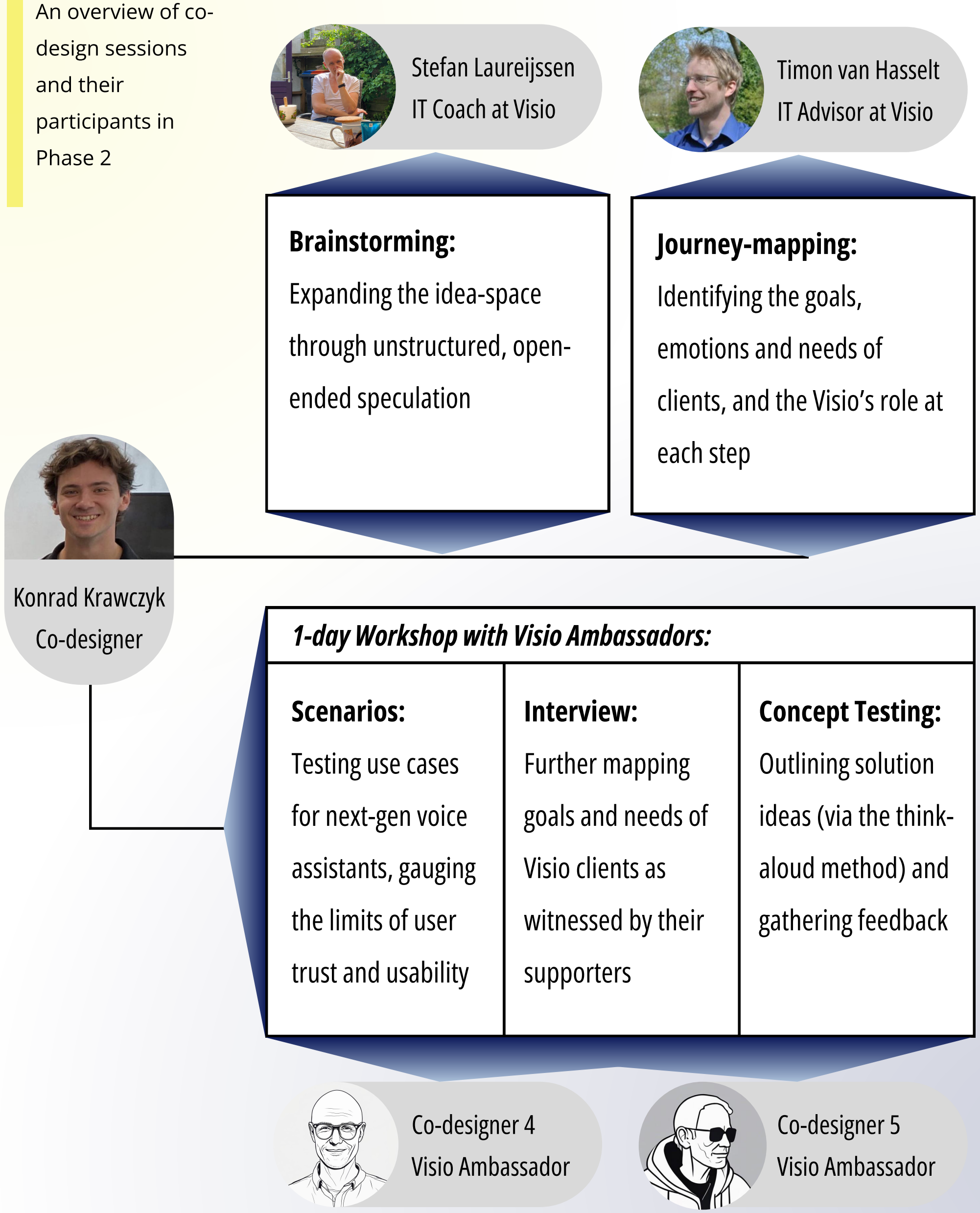
3.2. Co-designing with Experts

One of the ground rules for this research is to seek and include the informed opinions of subject matter experts. For that reason, I engaged in participatory design activities with stakeholders from various areas: IT solutions, tech coaching, and ongoing support (Visio ambassadors). This approach allowed me to gain a deep understanding of the relationship between Visio and the potential target audience of this project, even if it was through indirect connections. Additionally, it provided an opportunity to exchange knowledge and collaboratively explore the use of the latest technology, including when and why it should or shouldn't be used. Key expertise in the team:

- **IT Advisor**: designing and building tailored assistive technology solutions for the Visio community
- **IT Coach**: teaching BLV clients how to use technology non-visually
- **Visio Ambassadors**: members of the BLV community (one blind, one low-vision), providing training, mentorship and advice to the Visio clients who are starting to adapt to living with visual impairment.

The next page contains an overview of the sessions and the participants in each one of them.

Figure 22.
An overview of co-design sessions and their participants in Phase 2



3.2.1. Brainstorming Ideas

My first activity was a two-hour-long brainstorming session conducted with Stefan Laureijssen, a tech coach from Visio, as well as an expert and advocate in accessible technology. I decided that listening to his insights and ideas before jumping to prototyping could help me focus on the most pressing needs at Visio, and avoid assumption-based designing. Some of the key ideas mentioned by Stefan during the session were:

> "Maybe there's a need for people to learn with the assistant, and teach it also how to make adjustments."

> "So there's all kinds of users. And that's interesting because I also see the other side of that, that some people are not even aware of screen readers."

> "Most people, when the device is speaking, really see it at once as a true person. They want to communicate with it."

> It's great when you can ignore the rules and that thing just will listen."

The session with Stefan has helped me narrow down what needed to be designed: a tool that accounts for the awareness gap of the potential of voice technology. It also gave me useful guidelines for how to create a usable interface for that purpose.

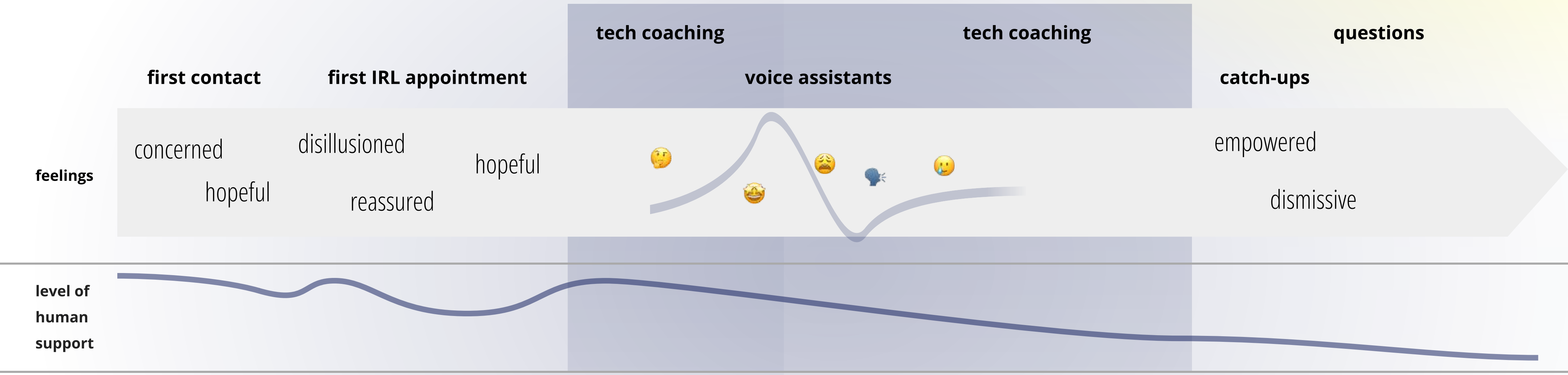
3.2.2. Extending the Client Journey

In this activity, I had a conversation with Timon van Hasselt, the Advisor from Visio's IT Department, shortly after the end of Phase 1. We did not follow a particular protocol, but the ultimate goal was to review what exactly happens when a client first approaches Visio, including the clients' usual life circumstances, expectations, concerns and goals. This resulted in a client journey map that extends vastly beyond the previous one, which only covered the experience of using a voice assistant. This activity provided me with an essential outlook on who exactly I am trying to help, what people other than the clients may be involved, and at what stage a design intervention may be most suitable for the design goal.

Key Findings:

- People who turn to Visio for support are often in complex emotional states, requiring more than knowledge and solutions
- First step is to establish what Visio can and cannot do. For example, Visio does not provide medical support aimed at restoring vision, which can be difficult for initial-stage clients to realize
- Learning to live fully as a BLV person is first and foremost aided by lived experience, mentorship and support
- Learning to use assistive technologies for independent living is the next step - but will still require human support at initial stages

Figure 23.
An extended Visio client journey diagram, including stages before and after technology coaching



3.2.3. Testing Assistants with BLV Experts

Later, I have arranged a testing session The Visio Ambassadors in an offline meeting at the university campus. The session consisted of three parts, and testing an assistant was the first part. The reason why I decided to do that was because Ambassadors were also partly in the target audience of this project - members of the BLV community who wanted to know more about the usability and limitations of voice assistants. The objective was to gather information about how BLV users of voice assistants formulate prompts, examine patterns in usability limitations of these assistants, and map out where their use value begins and ends. It was also a sensitizing exercise for later idea gathering and concept testing with the same group.

During the meeting, the Ambassadors were asked to install Microsoft CoPilot, an LLM-based voice assistant, on their own devices and complete specific tasks outlined in Appendix VI. Following these tasks, participants were asked questions about their experiences, with a particular focus on their levels of trust in the voice assistant.

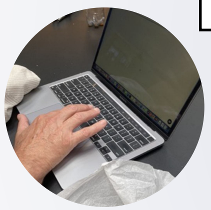


Figure 24. Example questions and test setup

Please give me a menu for the Umami restaurant in Amsterdam.

Please tell me what flights are available to Bangkok this week.

Please provide me information about Ibuprofen.

Quotes from Ambassadors:

"That answer was above my expectations. It provided a nice and long description [for the restaurant menu]".

"I wouldn't want a robot to schedule a flight for me. I would rather start calling. But airlines are usually hiding their phone numbers."

"Too much information is a big problem. I was really impressed with it, but it was too long." (after looking for information about Ibuprofen)

"Next flight to Bangkok in three weeks. That is simply not true, and I know it".

Key Findings:

- The assistant itself was difficult to get started due to the difficulties in installing the app and navigating the interface, signifying an **onboarding issue.**
- The least trustworthy answers involved real-time data about flights and location-specific information.
- Pauses in speech and lack of structured prompt formulation methods have hindered outcomes for one of the participants
- Answers from a voice assistant get overly specific and wordy, causing information overload.

3.3. Concept Development

Objectives:

- to gather impressions about a variety of solutions across different stages in the client journey
- to test the feasibility of the solutions in the ecosystem of Visio

Stakeholders involved:

- Visio Ambassadors (n=2).
- Visio members directly involved with the project.

Limitations:

- The feedback session had to be paused prematurely due to time and well-being considerations of participants.
- For Concept 3, feedback from non-Ambassadors has been included instead.

Methods:

- I have introduced the concepts to two Visio Ambassadors using a read-aloud method. Descriptions and live examples were provided to participants, followed by a list of questions (see Appendix VI).

3.3.1. Concept 1: Leveraging First Contact

timeline: first call to end of tech coaching

***Medium**: AI voice over a hotline, personalized assistive device (e.g. Hable)*

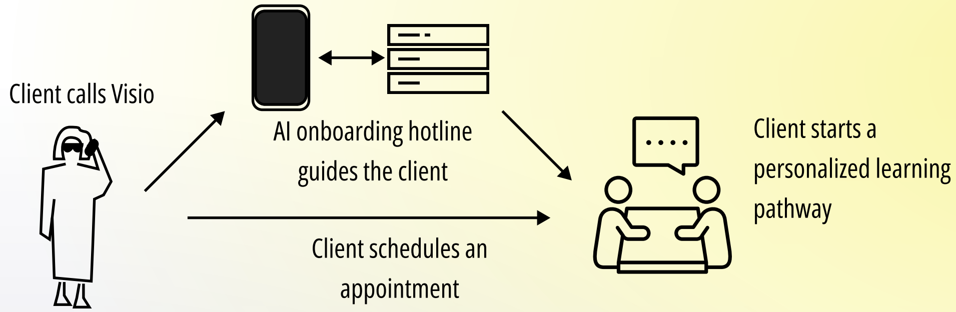


Figure 25. An overview of Concept 1, including an AI onboarding hotline and ending in a personalized coaching plan

This concept utilizes a voice interface at the very first stage of client journey with Visio. The initial questionnaire is provided by an automated agent, which increases the efficiency of processing inquiries for Visio while familiarizing the user with voice technologies. This first impression - if successful - can be a base for future training. The data gathered in the questionnaire can then be included into a personalized support device that is given to the client during the first technology coaching meeting.

Impressions from Visio Stakeholders:

- The process of “onboarding” as a Visio client is often emotionally charged.
- It can be a disappointing experience on its own to call a support network during this vulnerable time only to be answered by a robot.
- The above may change as automated voice agents proliferate and become more technically sophisticated.
- It can be more productive to use (semi-)automated services at later stages, when clients go through specific training units.

3.3.2. Concept 2: Course Companion

during tech coaching

***Medium**: AI voice, podcasts & recordings*

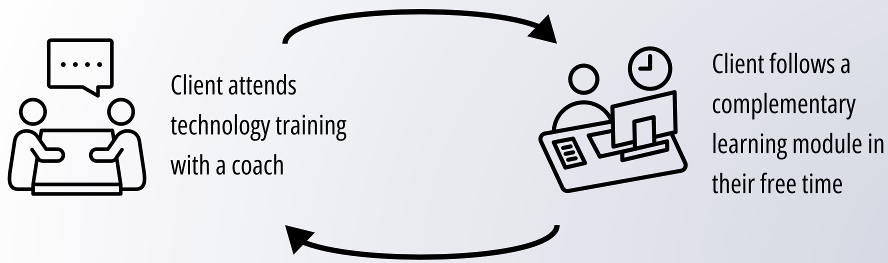


Figure 26. An overview of Concept 2, with complementary in-person and online learning

The program is a full-fledged online learning system designed specifically for BLV (Blind or Low Vision) clients. It is structured in a linear fashion, similar to an online course. Each unit within the system is designed to include several key components. First, the module provides a summarized explanation of a particular skill. Following this, clients have the opportunity to practice the new skill and receive feedback, which is crucial for reinforcing learning and ensuring progress. Additionally, each unit offers access to information from knowledge base articles and podcasts related to the topic at hand, enabling clients to learn by examples.

Impressions from Visio Stakeholders:

- The linear progression supplements the activities that tech coaches are directly involved in, enabling more efficient knowledge transfer.
- Information overload is a significant limitation - for example, an excessively lengthy podcast that cannot be fast-forwarded
- A linear system is vulnerable to users’ missteps, which can result in confusion
- It’s important to know what “success rate” the target audience expects from the system (with regard to providing coherent and relevant outputs)

3.3.3. Concept 3: Feedback Loop

during and after tech coaching

***Medium**: AI voice through a hotline, possibly an app with tactile feedback*

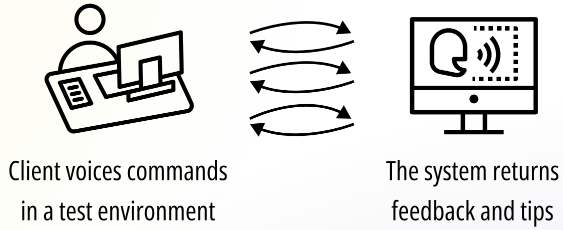


Figure 27. An overview of Concept 3

The system features a voice agent that is constantly available to assist clients. This agent serves multiple purposes, including helping clients practice their digital skills and checking the clarity of voice assistant prompts, functioning much like a “spell checker” for those.

Impressions from Visio Stakeholders:

- Such a system has not been tried yet at Visio.
- A practice-based module complements the other initiatives undertaken by Visio, as a knowledge base is already being created
- The feedback will only be as good as the available information about how each voice assistant works

3.4. Key Usability Concerns

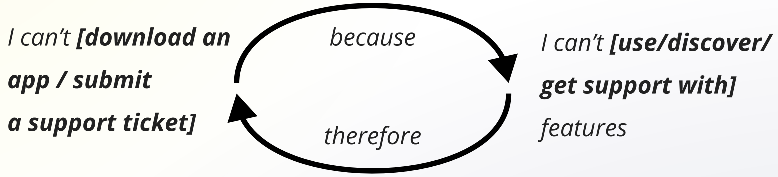
The conversations with experts from Visio have helped me to distinguish key areas to focus on when addressing the design goal. They are applicable in two areas: the design of embedded voice assistants, as well as to the assistive features aimed at supporting the use of those technologies. Re-designing the embedded voice assistants is out of scope for this research, as it is not a deliverable that can be integrated into the goals of Visio. However, an assistive service that supports learning for BLV people is likely to use voice technology as well, and the following design aspects are made all the more relevant.

Ease of Access

The initial stage effort required to use assistive technologies for BLV people is barring many of them from accessing it altogether. This is most pronounced when using specialized technologies, such as screen readers, but extends to all visual-prime devices. What is straightforward for a sighted person (for example, downloading an app) is anything but straightforward for a BLV person. This also (most paradoxically) includes IT support: “sending support tickets” or even browsing knowledge bases is generally not possible without pre-existing knowledge of screen readers.

For research purposes, I call it the **first-move problem**, in which a potential user of a technology is barred from fully using it because discovering and using its features requires pre-existing navigation skills. It is also why it can be a bad idea to embed assistive features for early-stage users into apps - these have to be installed, and at the current stage this requires either a screen reader or a sighted human support.

Figure 28. Overview of the first-move problem



For that reason, any service that doesn't include direct human support has to be as easy to access as practicably possible for a BLV person.

Simplicity

The KISS rule (Keep It Simple, Stupid) is well known in the design world. However, adhering to it becomes more crucial when designing for any context with sensory limits. There are two reasons for that. Firstly, the lack of multi-sensory feedback limits the capacity of the user to gain intuitive and unambiguous overview of what is happening on the device. In voice assistants, BLV users get audio-only responses where sighted users get both audio and vision. Secondly, not all senses are made equal. Audio is a linear medium - it is impossible to "skim" through a voice response to get the most important data.

An inappropriate length of the response, as often happens in voice assistants (also outlined in Phase 1) can be a cause for information overload, and wasted time. However, that can get even more frustrating when the user query gets misinterpreted by the system. That is especially applicable to any "smart" system where natural language processing and/or large (hallucinating) language models are involved. For those reasons, it is better to stay cautious and avoid ambiguities or open ends in a system, where possible.

Trust

With regard to voice assistants, trust is established both before and during the interaction. Non-human voice is less valued here, as there are areas of human emotional life which remain difficult for even most sophisticated AI models to understand. That implies that non-human support services may be better suited for achieving utilitarian goals, as opposed to abstract emotional intelligence. Even then, trust can quickly erode when a "smart" system fails to uphold a certain standard of astuteness. Finally, trust in a system can vary by area of knowledge. As tests showed, users may be more inclined to believe in encyclopedic knowledge of a model rather than its real-time data. This means that non-human support should be provided at appropriate stages where it is most likely to gain user trust, otherwise risks counterproductiveness.

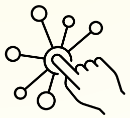
Adaptability

Experts at Visio have signalled a need for support services that are tailored to the specific circumstance of the user, as well as possible to adapt on-the-go. For example, a user might need quick information when navigating, but they might need detailed information when they are asking a complex technical question. The immediate circumstance of the user also plays a role. A well-known phenomenon in the realm of voice interface design is the “cocktail party problem”. The term refers to the challenge of focusing on and understanding a specific auditory stimulus, typically speech, in a noisy environment with multiple competing sound sources (Bee 2008). An adaptable system for a BLV person can ask for human feedback and adapt to it, as well as provide feedback (and a way out) when the environment is not conducive to a robust interaction.

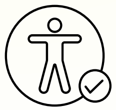
I have used these, as well as previously mentioned usability concerns, to conceive the final concept, which will be described in the following section.

3.5. Final Concept

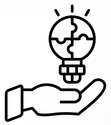
Visio Voice Playground: a voice interface available 24/7 over a phone call



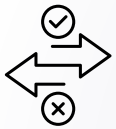
Participatory
The interface invites the user to practice their queries and get feedback and useful tips.



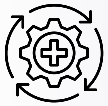
Easy to access
No app installation required. Standard phone line can be accessed through voice only.



Knowledgeable
Equipped with information from the Visio knowledge base and other sources, it distills answers for the user.



Simple
Navigation through a decision tree. A limited amount of information.



Adaptable
Mindful of its limitations, the agent is eager to be corrected and ready for the user to change their mind.

I conceived The Voice Playground as a synthesis of three previous concepts. The new idea addresses the key usability considerations derived from testing outcomes.

- I am to maximize the tool's **attainability** by eliminating the need for additional hardware or software—just one command or dial is enough for a BLV person to start talking.
- **Guidance** and quick error recovery is provided through a relatively simple decision-tree navigation.
- I aim to ensure maximum **trustworthiness** of the agent by sourcing information from fact-checked sources like Visio's knowledge base and including clear disclaimers about the tool's limitations.
- Finally, I plan to **enable personalizing** the system to fit each user's learning journey and circumstances, ensuring practice happens at the right place and time, with the right information for the right user.

I have created a storyboard of the new system, focusing on the client's first-person perspective. This overview can be found on the next page.

3.5.1. Concept Storyboard

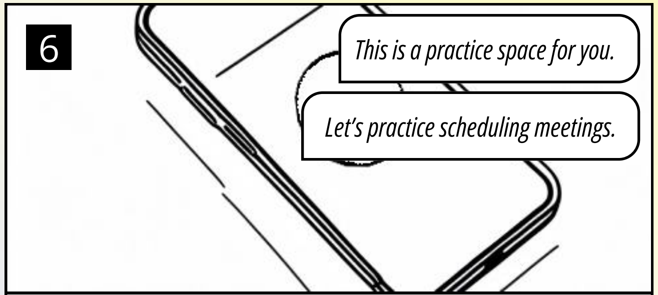
I need to get used to daily living with my eyesight condition. I call Visio, or have someone make the call for me.

The Visio specialist called me. I got an appointment slot! Hope the process starts soon.

The professionals at Visio are here to help me fulfil my goals. One of them is to learn to use technology without sight.

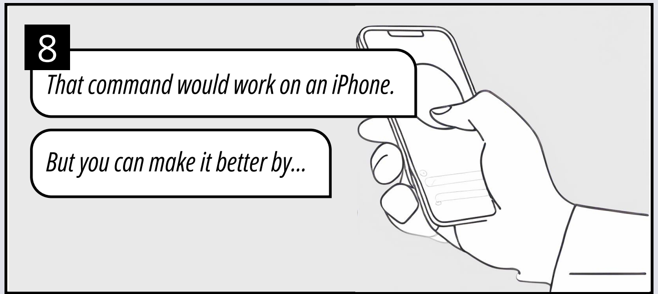
My technology coach at Visio has asked me if I want to take follow-up practice sessions in my free time.

I'm free now, maybe it's time to practice some tech stuff. I'll go on a call with the service my Coach told me about.

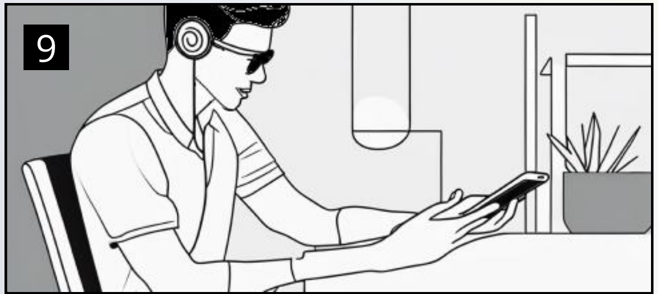


I got introduced to the Visio Voice Playground. First, I would like to know how this tool works - I get an voice explanation for that.

Okay, so I can use it to test different voice commands and practice them. Let's try...



I can get feedback on the spot, without changing any settings on my actual device.



I can practice the new skill again, move on to other skills, or freely ask follow-up questions about the new information.

3.6. Testable Targets

For the tests during, the following targets have been set to determine the level of success of the design. Those targets are reflected in the SUS scale presented to users after prototype testing (See Appendix VII)

1. The user can access the support device on their own*:
   a. with no manual intervention of a sighted person
   b. with >1 minute of verbal guidance from a sighted person

2. Majority of users trust the information provided by the agent:
   a. before use
   b. after use

3. Majority of users trust the privacy of the agent:
   a. before use
   b. after use

4. Users on average report increased knowledge of voice assistants.

5. Users on average report feeling more confident about using voice assistants.

6. Users consider the tool as a knowledgeable companion.

*The target scores 1a) and 1b) are observed by the facilitator, rather than reported by users.

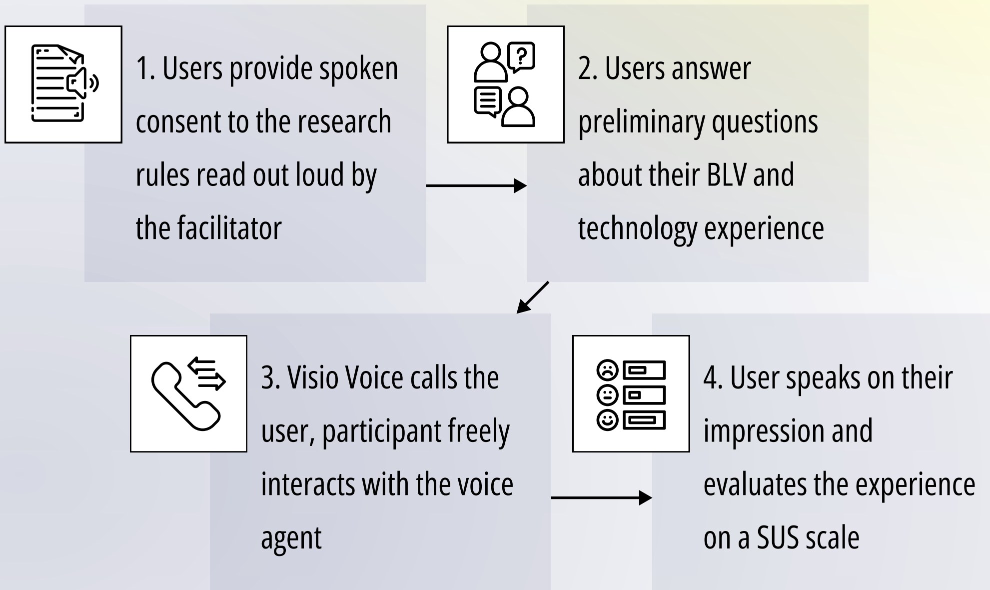
3.7. Iteration 1

The following is an overview of the user tests of the first iteration of the final concept. A full description of the testing setup, procedures and results can be found in the Appendix VII.

3.7.1. Test Setup

I have conducted the tests remotely over a span of two days.
Each individual test took around 20 minutes. The test procedure can be seen on the figure below.

Figure 29.
A diagram of the test setup, including consent procedure, preliminary questions, prototype testing and evaluation



3.7.2. Participants

Participants have been sourced through an external EU-based accessibility agency, as well as through private networks. Most participants are experienced with technology and come from a different linguistic context. Translations have been provided where necessary.

P1
HR Manager
Ukrainian,
legally blind,
tech experience

P2
Accessibility Tester
Ukrainian,
low vision,
high tech experience

P3
Entrepreneur,
Polish,
legally blind,
low tech experience

P4
Accessibility Tester,
Ukrainian,
legally blind,
high tech experience

P5
Programmer,
Ukrainian,
legally blind,
high tech experience

Figure 30. Overview of user test participants

3.7.3. Test Results

The test was oriented at identifying the extent to which the proposed solution can fulfil testable targets, as well as to what extent the concept has the potential to fulfil the design goal. Figures below show the overall results, with answers to the respective SUS questions mapped from 1-4 scale to a (double) minus - (double) plus scale.

Figure 31. Overview of user test results

Testable Target	P1	P2	P3	P4	P5
1a) No manual sighted intervention	n/a	n/a	n/a	n/a	n/a
1b) <1 minute of spoken guidance	+	+	+	+	+
2a) Information trust before use	-	+	+	-	-
2b) Information trust after use	+	++	+	-	--
3a) Privacy trust before use	-	++	+	-	-
3b) Privacy trust after use	+	+	-	--	-
4) Increased knowledge	+	-	-	--	--
5) Increased confidence	-	+	-	-	--
6) Is a knowledgeable companion	+	-	+	+	-

Key Findings:

✅ No participant found the system unnecessarily complex.
✅ Most participants can with certainty imagine other people could learn the system quickly.
✅ Most participants felt confident using the system.
✅ 3/5 participants indicated they would use this system frequently.

❌ Most participants didn't find the various functions in the system well integrated.

It was interesting but at first long thinking. But later she began to answer faster - **P1**

When I ask a question three or four times it just returned to the very beginning. - **P2**

I expected a translation, but instead got an instruction on how to use the translator. - **P3**

I got a list of numbered steps to follow. I thought I was supposed to press one of the numbers. - **P4**

My other conversation was overheard, and then the system got lost. - **P5**

3.7.4. Evaluation and Recommendations

Despite the experience of the testers, the testing setup had some limitations. The user base involved were largely professional accessibility auditors, meaning they were relatively skilled in assistive technologies such as voice assistants and screen readers. Although their expertise with testing improved feedback quality, this is not a target audience match, as the project is aimed at early-stage users. Besides, cultural differences may have affected the results, as most participants were Ukrainian and tested the AI-based translation of the voice platform. Further tests with Dutch speakers are essential to obtain context-specific feedback.

Nevertheless, the tests have yielded data that form a basis for a further prototype. The next iteration should provide a much better error recovery mechanism, eliminating "dead ends" in navigation. The interaction with the user's audio environment should be enhanced by selective handling of pauses and interruptions of speech, as well as clearer audio feedback when speech starts and ends. Finally, it's important to be realistic about what voice technology can or cannot do. In this case, some of the non-linear user pathways might have to be narrowed down to simple yes/no questions to avoid misinterpretations by the algorithm, enabling clearer (if slightly more limited) navigation.

4. Integration Phase

This section describes the activities for the third phase of the project, which consolidates all findings into a solution prototype, which is then evaluated in the context of use.

4.1. Introduction

For this phase, I decided to focus on integrating the insights gathered in the expert sessions and prototype tests directly into the context of Visio's technology learning program, with the following objectives:

- **Testing with the core of the target audience.** There are multiple potential target audiences for this project, as outlined in Phase 1. However, this study would not be complete without the insights of people who are directly participating in technology training as provided by Visio.
- **Embedding the tool into the existing network of services at Visio.** The tool might be suitable for some BLV individuals as shown in Phase 2, but as it is a company project, the research has to account for the unique client-support staff relationships that are at the core of Visio's activity.
- **Synthesize findings** into a set of recommendations and best practices, as well as creating touch points for further research and design.

Further section will outline the resulting prototype adjustment that aims to address those goals.

4.2. Iteration 2

4.2.1. Overview of Changes

The tests in Phase 2 have indicated a need to simplify the solution scope and refine the execution. Because of that, I decided on a significant overhaul of the prototype. Most importantly, the new system is directly embedded into the Visio's learning process. It serves as a personalized learning companion, available upon opt-in exclusively to the clients of Visio. A detailed user flow can be found in the Appendix part V.

Figure 32.
Comparison of iterations of the Visio Voice Playground

iteration 1	iteration 2
• knowledgebase and training • no data about the user • open-ended • decisions inferred through AI language processing • asking user to select skills • no third-person examples	• emphasis on practice • has basic data about the user • more close-ended • decisions made with yes/no • provides user with a single learning suggestion • can provide third-person examples

4.2.2. Key Final Design Decisions

Personal Service for Visio Clients

Tailored to individual needs, ensuring privacy, greater personalization, and seamless integration into the Visio company ecosystem, balancing service with information.

Proactive Engagement

The system reaches out to users by phone, reducing cognitive load and minimizing technical friction to ensure maximum accessibility.

Self-Explaining Service

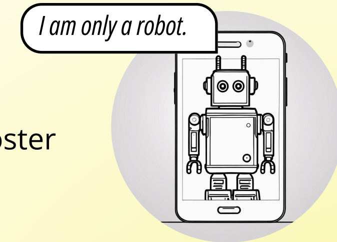
The agent introduces itself, offering clear navigation cues to provide essential information suitable for users of all experience levels.

Adaptive to User Needs

The service asks for consent for actions like practice sessions, with the flexibility to reschedule, aiming to establish trust and integrate smoothly into the user's lifestyle.

Transparent Disclaimers

The service makes clear and straightforward disclaimers about its limitations, in order to foster trust and ensure users are fully informed.



Aiding Memory Recall

Users are prompted to remember key information, supporting memory retention.

Clear and Concise Explanations

By default, explanations are brief but can be expanded into more detailed examples, reducing cognitive burden and providing actionable information with an option for further details.

Test Space

The interface gently guides the client towards learning by doing. The system acts out maximally similar responses to the actual voice assistants.

Feedback

Is provided at the end of a practice test, in a respectful but informative way, with pointers for the next steps.

4.2.3. Test Setup

For Iteration 2, I have conducted tests in two different modes: offline on Visio's premises, and online using Zoom. I considered it essential for the research to include immediate insights of people who are the core target audience, namely the clients of Visio who are undergoing technical training.

With the help of Visio staff members, I was able to conduct several testing sessions at the Visio's Loo Erf facility in Appeldoorn. The Loo Erf center is a residential space for comprehensive care and training for people who need support with adjustment to living with visual impairments. It includes psychosocial support, language learning, and most importantly for this research, technology training. Each learning pathway is personalized. Some clients stay there full-time, others several days per week. The training sessions are tailored based on the client's self-determined learning goals and pre-existing knowledge. During their stay at the Loo Erf, participants test different forms of assistance, and learn those that work best for them - whether it's Braille, screen readers, or voice commands. Last but not least, the Loo Erf provides BLV individuals with a sense of community, creating ample opportunities for socializing and sharing experiences through group activities and open-space discussions.

Figure 33. Accessibility and openness at the Loo Erf Center

Three participants have been recruited with the help of Visio staff. For those clients, I have arranged individual tests of 30 to 60 minutes each, in the Visio's Living Experience Lab. Each test consisted of preliminary questions, followed by an undirected test of the prototype (on the user's own phone or through my laptop). After a successful test, a questionnaire and a short debrief followed. The test procedure was identical to the one conducted in Phase 2, in order to ensure that results are comparable across different iterations. In order to get more more data, I have also performed three short tests with participants during the afternoon hours in the Loo Erf's open dining space. In order to reduce participant fatigue, I limited those sessions to prototype testing and quick feedback gathering.

Figure 34.
Full-length test setup at the Living Experience Lab

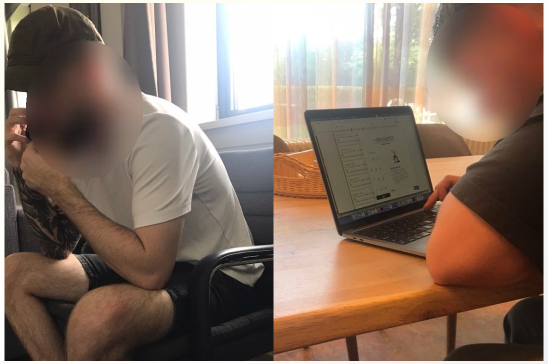


Figure 35. (left)
Participants testing at the Lab (left) and the open space (right)

Figure 36. (below)
Table of test participants with basic information

P1 	P2 	P3 
Ukrainian	Dutch	Dutch
English-speaking	English-speaking	English-speaking
In tech training	In tech training	In tech training
Intermediate Siri user	Beginner Siri user	Intermediate Siri user
Eye loss (veteran)	Blind	5% eyesight

+ three BLV participants in short tests (referred to as **P4**, **P5**, and **P6**)

4.2.4. Test Results

This section outlines the insights derived from the aforementioned tests. The summary provides key findings that inform further sections of the report, most notably recommendations and best practices. A more detailed overview of the results, including the full results of quantitative tests, can be found in Appendix IX.

Analysis method: In order to synthesize the insights provided below, I have first used an open-source Whisper model to transcribe the interviews. Then, I have read those interviews, anonymized them, and marked key quotes with the following codes based on previous analyses: trust, agency, use value. I have also used a large language model later to identify any themes that may have been omitted from the analysis. For the corresponding quantitative results, I have converted the 1-4 scale into a (double)positive-(double)negative scale, and made a chart of those results for increased readability.

In a nutshell:

- The tool leads to an **increased sense of knowledge** and **confidence** about voice assistants in all participants.
- **2/3** participants trusted the credibility of the information provided by the tool, as well as the privacy of the system.
- All participants found the system to be a knowledgeable companion.

The following section delves into qualitative insights yielded from the interviews. The topics considered are: **trust**, **agency**, and **attainability.**

Trust

For the purposes of this research, trust has a twofold meaning. It means the perceived credibility of the information provided by the system on one hand, and the trust in the appropriate, privacy-conscious handling of the user's input on the other.

Figure 37. Levels of test participants' agreement with trust-related statements (1=strongly disagree, 4=strongly agree)

> If I open my iPhone now and hear the time, [...] that's one hundred percent true information. But if I go, for example, to ChatGPT and ask it something, it's not always 100% accurate. ~P1

> You're aware that you are talking to a robot or an AI, but the response was very natural and very accurate. ~P3

The proposed tool performed well with regard to perceived **factuality**. Prior to the test, participants indicated that they would trust the tool, and in one case (P3) the trust has increased after use. However, as the quote from P1 indicates, AI outputs are to be trusted on a case-by-case basis. While responses to straightforward queries are easily credible, more complex or specialized questions might not necessarily be fully accurate, which poses a challenge to the perceived usability of the tool.

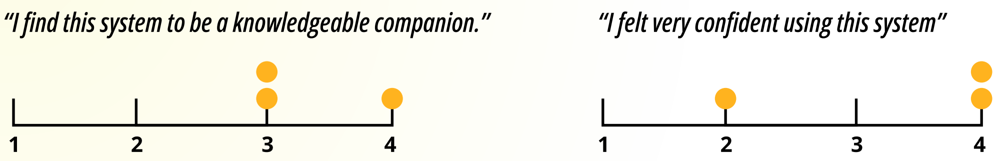
> No, it's not private, because it's a beta version ~P1

The test environment strongly affects the perceived level of privacy. Further comments are available in the Limitations section.

Agency

Defined as the perceived ability to adjust the course of action based on personal goals and aspirations, agency plays a pivotal role in the usability of the system. It is central to the ultimate design goal, which is to help BLV individuals access digital technology independently. Perceived sense of knowledge and confidence using a system is part of that. Consequently, it is also vital that the user believes the tool to be a helpful companion, without an condescending or intimidating disposition that would undermine that agency.

Figure 38. Levels of test participants' agreement with agency-related statements (1=strongly disagree, 4=strongly agree)



> I thought the system would say, "I don't have information about that," but she gave me the right answer. So, that was a bit surprising, but in a good way. But it's still consistent because she gave me the correct information. [...] It makes the system feel more like a real assistant. ~P1

The system's flexibility was appreciated by those participants who felt relatively comfortable with voice technology. Throughout the test, it was possible to ask questions that were outside of the envisioned "voice playground" idea. The system gently prompted the users into practice mode, while making sure to address their questions. It also allowed spontaneous interruptions, which has been used by two users.

> There was a moment we sat here in silence and like, okay, what should we do now? ~P3

Clear interactive feedback was an aspect of the interface that was challenging for the users' agency. Despite basic cues (e.g. when listening for user input), the processing time of the prototype varied. For that reason, both P1 and P3 were not sure if they were allowed to speak to the interface, or if they had to wait for a response. In the case of participant 2, user input was prompted too fast, leaving them with a relatively short amount of time to formulate a question on their own, and being unable to use the interface as intended. It is possible that the level of agency afforded by the system was more appropriate for intermediate-to-advanced technology users, who could then ask spontaneous questions. This, however, left the beginner users with too much openness and not enough guidance.

> [Older people] have different goals for using voice assistants—calling family members, for example, and specific use cases. ~P1

It's notable that agency does not mean the same things for all users. Too many possibilities can backfire - users need just enough context for fulfilling their goals, whether it's something relatively simple like calling a relative, or checking information about specialized assistive tools.

Attainability

A significant focus point, based on the previous prototypes and expert talks, was that the tool should be accessible with little to no pre-existing knowledge of technology, which I call *attainability* for distinction. This includes the very initial stages of use, from the moment the client gets to knows about the tool to the first interaction. Further on, it also includes the level of ease with which a user can independently discover and utilize the tool to achieve their goals.

Figure 40. Levels of test participants' agreement with attainability-related statements (1=strongly disagree, 4=strongly agree)

> I think this is easier for me than to read it on the internet. [...] I would definitely be curious to work with it more. ~P3

> I think it can work quite well because I still do just a small part of what's possible with Siri. And sometimes then I check on the internet a list again ~P1

For the slightly more tech-savvy, the Playground is a more accessible way to access technical information than standard assistive tools. It provides relatively concise answers while bypassing the layer of navigation otherwise required in tools like VoiceOver. For P1, it was seen as a practical tool that could be used alongside other resources online, possibly helping him corroborate the information found in various online tutorials. Attainability for beginner users of voice assistants, however, was more of a problem.

> It's just a bit too fast for me! I don't think I'd be able to use it, it gets quite overwhelming. Too much mental load ~P4

> I would be happy to use it if it was fully in Dutch ~P2

Testing with more early-stage users of voice technology has revealed significant challenges. Part of the difficulty was due to the fact that the tool presented to them at first was in English, instead of their native language (Dutch). Despite the fact they were comfortable using English during conversations mid-test, the comfort of use of an English-only tool that requires active input could increase its mental load, as mentioned by P4.

The speed of the practice presented by the Playground could also pose a challenge on its own. Participant 2 has struggled with understanding the idea of the Playground as a test space for assistants. Instead, they were trying to use the Playground as an assistant itself. This may signify a need for further adjustment of the tool to cater to those users who need more steadiness, in order to make the tool truly accessible.

Online

After the on-site tests, I have organized two more online tests. I have recruited participants using a Polish accessibility testing agency for tests lasting 30 to 60 minutes, explicitly asking for participants with limited background in technology. The research procedure was identical to the one from the on-site tests, except it was conducted via Zoom.
Their results are available in the Appendix IX.

The primary reason to organize those tests was to gather further internationalized results (both participants were Ukrainian) and to compare the results against the ones from Iteration I. However, during the tests it turned out that the participants were in fact not within the target audience (both had advanced skills in accessibility options and one co-organized rehabilitation classes). For that reason, I am not including them in the primary analysis. However, the results are promising, and they offer interesting pointers for who a potential target audience could be for this project.

Limitations

The method utilized for user testing in this iteration had some limitations. Firstly, it was relatively difficult to differentiate the target audience from non-target-audience users. In fact, the users that were in technical training were already somewhat adept in assistive technology, which definitely made them more confident than a typical beginner user may have been.

Secondly, technical difficulties have occurred during testing with participant 2, as we were unable to use the tool with a Dutch translator. For that reason, I have repeated the test with the same participant remotely after the on-site test. Their scores included in this report are the ones I gathered during that online test.

Also, the test setup may have affected some responses during the test. Given the fact that the prototype has been shared from my own computer, it was easier for users to discern that I had access to hte data provided. The privacy score may have been altered by that fact.

This concludes the findings of the final usability test. The further Discussion section takes these and other findings from the study, and synthesizes them to distill further areas of detailed research.

5. Conclusion

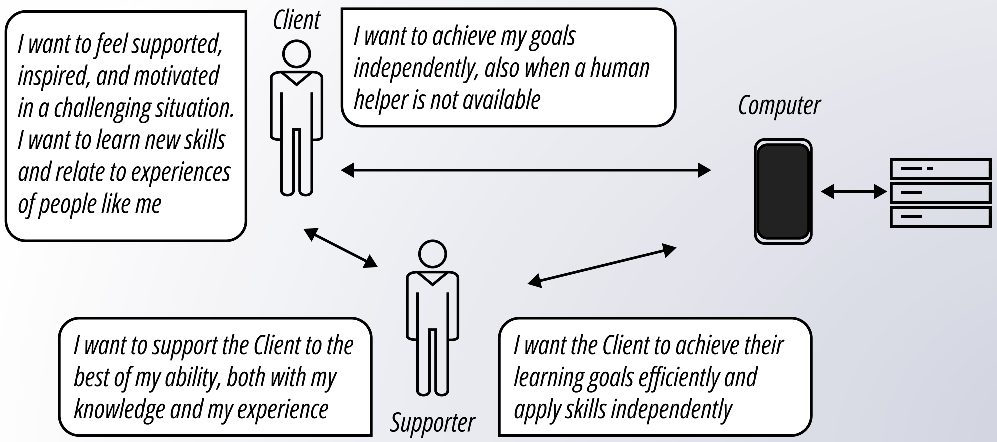
This section synthesizes and summarizes the key findings of all research cycles, resulting in a list of best practices, as well as an evaluation and personal reflection.

5.1. Discussion

The aim of this section is to consider all the findings from research activities thus far, including the literature review, online research, expert sessions, prototyping and user testing, and identify the key **design tensions.** I believe that designing is a balancing act, and identifying interests that are seemingly at odds can form a basis for further research, and inform design decisions for future iterations.

5.1.1. Client - Supporter - Computer

Figure 41.
Network of agents and their respective goals revealed in the study

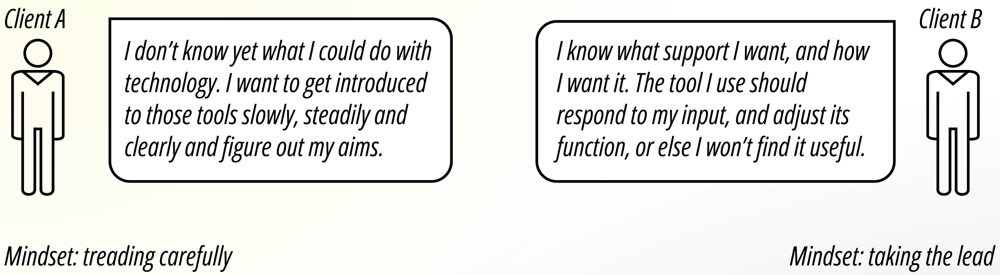


There is a social component in technical skill training for BLV people. As I found out in conversations and tests with both Visio professionals and clients, technical support is merely one aspect of support needed and sought after. What's important especially at early stages of adjustment to living with visual impairment is the feeling of being supported, which is not easily accomplished with technological fixes alone.

Through the literature review and the research of online sentiments, I found that technology itself serves largely as a proxy for achieving equal participation in social life. The Ambassadors of Visio are the best examples of how achieving a sense of independence in daily life as a BLV person is strongly linked to an exchange of shared experiences through mentorship, know-how transfer and empathy.
However, human support is limited by time and capital on both ends. Clients might lack resources to enroll into the support system quickly enough, and Visio itself may not be able to provide it due to waiting lists, time conflicts, or force majeure. This simple fact can prompt companies like Visio to seek technological proxies that support achieving the clients' goals more efficiently. In this context, there is space for a companion tool like the Visio Voice Playground, as it serves to offload independent learning and follow-ups onto a computer, leaving more time for human helpers to excel at providing what humans are simply better at: empathy and lived experience.

5.1.2. Openness - Guidance

Figure 42.
Diverging attitudes towards interface openness and guidance revealed in the study



Perhaps the most difficult aspect of designing for this context is managing the levels of independence, both aimed for and attainable. The BLV community members interviewed vary greatly in terms of their willingness to try out new technology, their prior exposure, as well as their visual acuity. Two of the experts interviewed in Phase 1 have emphasized the role individual personality plays in attitudes toward technology use, and the will to independence is strongly related to that. Another early insight I adopted from Visio experts is working with mindsets. The two I found to be most strikingly different were the cautious mindset (exemplified by Client A on the left) and the leadership mindset (Client B).

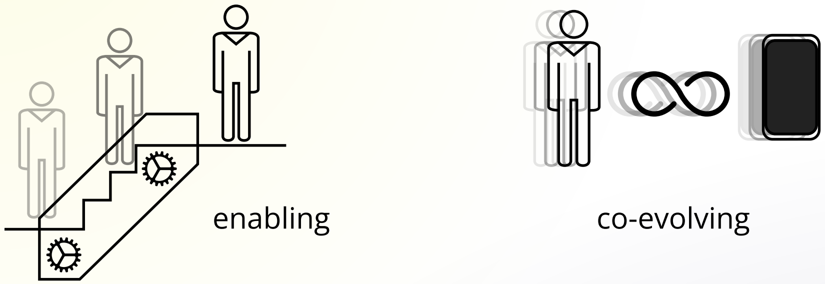
Creating a tool that addresses the needs of both of these groups is a complex task. I have attempted to resolve it by structuring the interaction in a T-shaped manner shown below. The tool can provide steady defaults and ground rules for beginner users to acquire most sought after skills, while remaining flexible and responsive to the needs of users who want to go off the prescribed path.

Figure 43.
Balancing broad use range and guided support in a T-shaped model of a learning system

Finding a tolerable level of mental load is still an issue to account for. In expert talks and user tests, I found out that even the most advanced users of voice assistants can struggle with processing large amounts of spoken information. Providing a concise, guided experience by default, while allowing the user to steer of that prescribed path afterwards, seems like the optimal way of reducing that mental load.

5.1.3. Short term - Long term

Figure 44.
Possible approaches to the scope of the solution proposed in the study



This is a question of whether a tool like the Voice Playground should offer ongoing support with tech-related topics, or if it should serve primarily as a proxy for initial stages of use. In theory, there is no issue with it performing both functions, as the T-shaped idea from the previous chapter would suggest. However, practical reasons may suggest to tread in one direction only. One of the main reasons that the second iteration was able to perform better in fulfilling the design goal was that it did *less.* Its desired scope was limited to skill practice. It did not attempt to be a feedback mechanism and a knowledge base at once. It was also designed to accompany a pre-existing learning relationship between a client and a coach, without replacing it. Still, some users attempted to "misuse" it as a knowledge base, and they found great use value in that.

After delving into the future of voice assistants in Phase 1, I am also wary of making broad-stroke assumptions about what would be the most practical forms of support in this space several years from now. A tool for long-term technological support would require very frequent updates, as the realm of voice assistants is evolving practically every week. With improved onboarding experience of commercial-grade voice assistants, it is possible that additional support tools could be made redundant.

The further section outlines design recommendations resulting from this and previous analyses.

5.2. Recommendations and Best Practices

The following recommendations are results of expert talks and iterative testing of prototypes in this research. This can serve for companies like Visio, as well as for designers of conversational interfaces more broadly, to create more inclusive learning support in the future. The first two pertain to the **Service** level, the next three to the content layer (**Information**), and the latter four consider the audiovisual qualities of the interaction with it (**Interface**).

5.2.1. Service

1. Respect human-to-human collaboration.
Teaching and coaching has an emotional component that goes far beyond transfer of information. For as long as a human is perceived as more empathetic and supportive than a machine (which may take a while), designers must ensure that design interventions in the technology learning area are complementing and enhancing the human-to-human learning relationship, instead of competing with them.

2. Make systems open-source.
A truly participatory design should include with their stakeholders at all stages of the product's lifecycle.

For that reason, any proprietary source code or assets should be open for edit suggestions, and any third-party services should be replaced with open-source equivalents whenever possible. This is especially true for public good institutions like Visio, which take participation and user privacy protection seriously.

5.2.1. Information

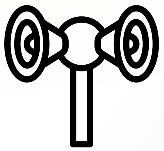
3. Establish Clear Ground Rules.
It is crucial for users to have clear and constant access to information about the voice assistant's functionality. This should encompass guidelines on interaction cues, such as when users can start speaking, and the system's limitations. These can be presented to the user as a quick tutorial during the initial use.

4. Ensure Robust Privacy Protections.
The voice assistant must adhere to strict privacy standards, with a privacy policy that is always accessible and fully compliant with regulations such as the GDPR. Users should be informed about how their data is collected, stored, and used, and they should have control over their privacy settings. This is essential to build and maintain user trust.

5. Ensure Factually Accurate Information.

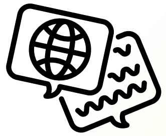
The information provided by the learning assistant must be accurate and aligned with real-world use cases. This necessitates integrating a database of human-written and human-reviewed content into the system. Additionally, common use cases should be rigorously tested to prevent users acting on false information. If any information provided by the learning assistant is generated through AI inference rather than established facts, clear disclaimers should be included.

5.2.3. Interface



6. Support Multi-Speaker Environments.

In environments where multiple speakers are present or where the device is set to speaker mode, there is a risk of audio output being misrecognized as user input. This issue, observed during testing, highlights the need for advanced noise-cancellation techniques and input recognition algorithms that can differentiate between user commands and other sounds in the environment. This is essential to ensure the usability of the voice assistant in diverse settings, particularly with human support involved, as is often the case in learning setups for blind and low-vision people.

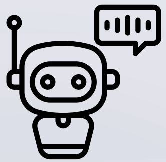


7. Language and Cultural Sensitivity.

While a user may speak a particular language, this does not guarantee they use it comfortably with robots. To reduce the cognitive load, it is important to develop context-specific language versions of the voice assistant, which should be tested with native speakers. Tone and phrasing should suit the cultural context, ensuring that interactions are natural.

8. Incorporate Sound Cues for Processing.

During the interaction, there may be delays as the system processes input and generates a response. To prevent confusion during these pauses, it is recommended to implement processing sounds or visual cues that indicate the system is working. This feedback loop will help maintain user engagement and reduce uncertainty.



9. Optimize Voice Quality.

The quality of the voice used by the assistant can significantly impact user experience. In languages with smaller speaker bases than English, machine-generated voices may sound robotic, unclear, or unappealing. It is important to conduct extensive testing with native speakers to select a voice that is both intelligible and pleasant.

5.3. Evaluation

The research method I employed in this thesis came with its set of strengths and limitations. Those may influence the generalizability and long-term applicability of the findings. Below is an evaluation of the strengths and issues identified in the research process.

Strengths:

- **Expert Collaboration:** Consulting with experts across various domains contributed to a deeper understanding of complex issues such as privacy, accessibility, and user agency, allowing me to focus on key subject matters from the start.
- **Iterative Design:** The iterative approach allowed for balancing conceptual ideas and their practical execution. This helped me arrive at the design early based on user feedback and testing, and focus on refining the prototype, and helping distinguish conceptual challenges from the shortcomings of the execution.
- **Internationalization:** Including cultural and linguistic diversity into the research sample enhanced the potential resonance of the design across different regions and user groups, contributing to a more inclusive approach.

Limitations:

- **Problems with Target Audience Sampling:** It was relatively difficult to find participants for this study who did not have a pre-existing experience and interest in using voice assistants. It is clear that this target audience exists, however, its needs are still under-represented in the study. Further participant search with fewer time constraints could help bridge that gap.
- **Limited Contextual Representation:** The sample size from the actual context of use was relatively small, which may limit the accuracy and applicability of the findings to broader user groups. Further research is needed to validate the resulting concept, if its outcomes are to be applied at scale.
- **Lack of Longitudinal Data:** The research did not extend over a long period, making it difficult to assess how the design might affect long-term learning outcomes and user behavior over time. To address that, further studies can be conducted to test the “forgetting curve”, and how the prototype use affects it.
- **Comparability Across International Contexts:** While internationalization was a strength, it also introduced challenges in comparing results across different cultural contexts, which may affect the consistency of the findings.
- **Rapidly Evolving Field:** As the field of voice assistants and related technology is advancing every month, some of the findings and design interventions may become outdated quickly, necessitating ongoing research and adaptation.

5.4. Concluding Remarks

The guiding principle of inclusive design, "solve for one, extend to many," is inspiring - but tricky. Blind and low-vision people cannot be considered as “one” because, like any group, they are far from homogenous. What unites them are their shared experiences of socializing, creating, growing, and performing daily tasks in a world not always built with them in mind.

Technology, in this context, is a means to an end. It’s a tool to help achieve goals, whether it's learning a new skill, securing a job, or simply connecting with loved ones. In a world increasingly driven by digital communication, access to these tools is, for better or worse, essential.

This chapter concludes the project report. Following that is a personal reflection, technical notes and references, as well as an Appendix.

5.5. Personal Reflection

Reflecting on this project, I realize that it marked my first real experience in a research leadership role. Stepping into this position felt unfamiliar and somewhat difficult. Typically, I thrive when I'm presented with a challenge by someone else and can then focus on solving it. However, this time I was the one responsible for identifying the most pressing challenge, which required far more proactive outreach and decision-making. It was challenging, but I’m grateful to have worked with colleagues who were incredibly supportive and proactive in helping me navigate this.

The success of this project would not have been possible without the input from the BLV community. But while doing this, I also learned the importance of allowing space for participants to breathe and express themselves. Initially, I approached each session with a detailed plan, but I quickly found that the most meaningful interactions came when I deviated from the script. By prioritizing empathy and understanding over rigid protocols, I allowed the lived experiences of the community to guide the research. This meant that some of the pre-planned research suffered, but the trade-off led to deeper, more authentic insights.

When I first joined this school, I was a fervent tech-solutionist, eager to build things and then put them to the test in the real world. However, courses like Exploring Interactions have shifted my perspective, encouraging me to listen more and act more intentionally. I'm proud that in this project, I started with the question of "why" and developed something with genuine potential for future use. I also gained a deeper understanding of conversational technology, moving beyond mere enthusiasm to a more nuanced appreciation of where and how these tools can be effectively applied.

If there's one aspect I wish I had approached differently, it would be earlier and more direct engagement with the core target audience. Recruiting participants with limited tech backgrounds challenged some of my assumptions and proved to be more difficult than I anticipated. Although time constraints played a role, I recognize now that earlier engagement could have taken this project from a promising blueprint to a more thoroughly validated solution.

My biggest takeaway from this process is that it's not enough for technology to be technically accessible; it needs to be truly attainable. This is why I chose to focus on the learning environment, aiming to enhance how digital tools can be used independently. My goal was to make these tools not just available, but genuinely usable.

Voice assistants, which were a focal point of this study, are advancing rapidly. However, their evolution is too often driven by business priorities that don't always align with inclusive values. If these technologies were designed with true inclusivity at their core, my intervention—and others like it—wouldn't be necessary.

5.6. Note on AI Use

Throughout this project, I have used several artificial intelligence tools to accelerate the process where needed. Some parts of the report have been edited with the help of large language models for clarity and conciseness. I have also utilized Dreamstudio.ai for supporting visuals and diagrams, which I then modified to fit the purposes of my project. Lastly, the project itself is explicitly using language models and speech synthesizers to process input and output.

5.7. Note on Report Accessibility

The report utilizes visuals for clarity and aesthetics, for the purposes of evaluation as part of the thesis requirements. Unfortunately, it was created in Figma, which currently does not include certain key accessibility features. Given the main target audience here are visually impaired people, this report is accompanied with an accessible version edited in LaTeX, with semantic formatting and alt texts included.

5.8. References

Bridging the Disability Divide through Digital Technologies - GIWPS. (2021, November 2). Georgetown Institute of Women Peace and Security. https://giwps.georgetown.edu/dei-resources/bridging-the-disability-divide-through-digital-technologies/

Consumer Intelligence Series: Prepare for the voice revolution. (n.d.). PwC. https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html

Global Voice Assistant Market: Component (Hardware (Smartphone, Smart Speakers, Security Systems, Laptops/Tablets, Smart Watch, Smart TV, Smart Lamps, In-Vehicle Assistants, 2-Port Car Charger, Headsets/Hearables, Others), Solutions (Standalone and Integrated), and Services (Consulting, Implementation, Support and maintenance); Deployment (on-premise and Cloud); Enterprise Size (Small and Medium-sized Enterprises (SMEs) and Large Enterprises); Technology (Natural Language Processing, Speech/Voice Recognition, Text to speech Recognition); Application (Automotive (Book Services, Navigation, Schedule Appointments), Smart Home (Home automation and Security services), Smart banking, Wearable Devices, and Others); Region—Market Size, Industry Dynamics, Opportunity Analysis and Forecast for 2024–2032. (n.d.). https://www.astuteanalytica.com/industry-report/voice-assistant-market

Siri and Alexa Reimagined: Personality Match Enhances User Trust. (2023, November 28). Neuroscience News. https://neurosciencenews.com/virtual-agent-personality-25275/

To end extreme poverty, we must also end blindness. (2021, July 2). World Economic Forum. https://www.weforum.org/agenda/2021/06/to-end-extreme-poverty-we-must-also-end-blindness/

What is Inclusive Design? (2024, February 19). The Interaction Design Foundation. https://www.interaction-design.org/literature/topics/inclusive-design

Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). "Siri Talks at You." Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility. https://doi.org/10.1145/3234695.3236344

Alnefaie, A., Singh, S., Kocaballi, B., & Prasad, M. (2021). An Overview of Conversational Agent: Applications, Challenges and Future Directions. Proceedings of the 17th International Conference on Web Information Systems and Technologies. https://doi.org/10.5220/0010708600003058

Amedi, A., Merabet, L. B., Bermpohl, F., & Pascual-Leone, A. (2005). The Occipital Cortex in the Blind. Current Directions in Psychological Science, 14(6), 306–311. https://doi.org/10.1111/j.0963-7214.2005.00387.xoi.org/10.1049/ic:19980960

Ashraf, M. M., Hasan, N., Lewis, L., Hasan, M. R., & Ray, P. (2016). A Systematic Literature Review of the Application of Information Communication Technology for Visually Impaired People. International Journal of Disability Management, 11. https://doi.org/10.1017/idm.2016.6

Baker-Nobles, L. (1977). Sensory Integrative Training for Blind Adults. Journal of Visual Impairment & Blindness, 71(8), 356–359. https://doi.org/10.1177/0145482x7707100806

Bias, R. G., Moon, B. M., & Hoffman, R. R. (2015). Concept Mapping Usability Evaluation: An Exploratory Study of a New Usability Inspection Method. International Journal of Human-Computer Interaction, 31(9), 571–583. https://doi.org/10.1080/10447318.2015.1065692

Bottini, R., Nava, E., De Cuntis, I., Benetti, S., & Collignon, O. (2022). Synesthesia in a congenitally blind individual. Neuropsychologia, 170, 108226. https://doi.org/10.1016/j.neuropsychologia.2022.108226

Branham, S. M., & Mukkath Roy, A. R. (2019). Reading Between the Guidelines. The 21st International ACM SIGACCESS Conference on Computers and Accessibility. https://doi.org/10.1145/3308561.3353797

Burton, H. (2003). Visual Cortex Activity in Early and Late Blind People. The Journal of Neuroscience, 23(10), 4005–4011. https://doi.org/10.1523/jneurosci.23-10-04005.2003

Burton, H., Sinclair, R. J., & McLaren, D. G. (2004). Cortical activity to vibrotactile stimulation: An fMRI study in blind and sighted individuals. Human Brain Mapping, 23(4), 210–228. https://doi.org/10.1002/hbm.20064

Crews, J. E., Jones, G. C., & Kim, J. H. (2006). Double Jeopardy: The Effects of Comorbid Conditions among Older People with Vision Loss. Journal of Visual Impairment & Blindness, 100(1_suppl), 824–848. https://doi.org/10.1177/0145482x0610001s07

Crudden, A., & McBroom, L. W. (1999). Barriers to Employment: A Survey of Employed Persons who are Visually Impaired. Journal of Visual Impairment & Blindness, 93(6), 341–350. https://doi.org/10.1177/0145482x9909300602

Fuglerud, K. S. (2011). The Barriers to and Benefits of Use of ICT for People with Visual Impairment. Universal Access in Human-Computer Interaction. Design for All and eInclusion, 452–462. https://doi.org/10.1007/978-3-642-21672-5_49

Goldstein, J. E., Massof, R. W., Deremeik, J. T., Braudway, S., Jackson, M. L., Kehler, K. B., Primo, S. A., Sunness, J. S., & Low Vision Research Network Study Group, F. T. (2012). Baseline Traits of Low Vision Patients Served by Private Outpatient Clinical Centers in the United States. Archives of Ophthalmology, 130(8), 1028. https://doi.org/10.1001/archophthalmol.2012.1197

Grussenmeyer, W., & Folmer, E. (2017). Accessible Touchscreen Technology for People with Visual Impairments. ACM Transactions on Accessible Computing, 9(2), 1–31. https://doi.org/10.1145/3022701

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018, August 22). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys. https://doi.org/10.1145/3236009

Jaber, R., & McMillan, D. (2020). Conversational User Interfaces on Mobile Devices. Proceedings of the 2nd Conference on Conversational User Interfaces. https://doi.org/10.1145/3405755.3406130

Martin, A., Nateqi, J., Gruarin, S., Munsch, N., Abdarahmane, I., Zobel, M., & Knapp, B. (2020). An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-75912-x

Merabet, L. B., & Pascual-Leone, A. (2009). Neural reorganization following sensory loss: the opportunity of change. Nature Reviews Neuroscience, 11(1), 44–52. https://doi.org/10.1038/nrn2758

Nielsen, J. (2024, February 20). 10 Usability Heuristics for User Interface Design. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

Parke, K. L., Shallcross, R., & Anderson, R. J. (1980). Differences in Coverbal Behavior between Blind and Sighted Persons during Dyadic Communication. Journal of Visual Impairment & Blindness, 74(4), 142–146. https://doi.org/10.1177/0145482x8007400404

Pascolini, D., & Mariotti, S. P. (2011). Global estimates of visual impairment: 2010. British Journal of Ophthalmology, 96(5), 614–618. https://doi.org/10.1136/bjophthalmol-2011-300539

Patoine, A., Mikula, L., Mejía-Romero, S., Michaels, J., Keruzoré, O., Chaumillon, R., Bernardin, D., & Faubert, J. (2021). Increased visual and cognitive demands emphasize the importance of meeting visual needs at all distances while driving. PLOS ONE, 16(3), e0247254. https://doi.org/10.1371/journal.pone.0247254

Pérez-Garín, D., Recio, P., Magallares, A., Molero, F., & García-Ael, C. (2018). Perceived Discrimination and Emotional Reactions in People with Different Types of Disabilities: A Qualitative Approach. The Spanish Journal of Psychology, 21. https://doi.org/10.1017/sjp.2018.13

Porcheron, M., Fischer, J. E., McGregor, M., Brown, B., Luger, E., Candello, H., & O'Hara, K. (2017). Talking with Conversational Agents in Collaborative Action. Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. https://doi.org/10.1145/3022198.3022666

Rochat, J., Ehrler, F., Siebert, J. N., Ricci, A., Ruiz, V. G., & Lovis, C. (2022). Usability Testing of a Patient-Centered Mobile Health App for Supporting and Guiding the Pediatric Emergency Department Patient Journey: Mixed Methods Study. JMIR Pediatrics and Parenting, 5(1), e25540. https://doi.org/10.2196/25540

Sanders, E. B. N., & Stappers, P. J. (2012). Convivial Toolbox. Bis Pub. http://books.google.ie/books?id=a8miuAAACAAJ&dq=convivial+toolbox&hl=&cd=1&source=gbs_api

Shah, A., Joshi, P., Bhusal, B., & Subedi, P. (2019).

Clinical Pattern And Visual Impairment Associated With Herpes Simplex Keratitis

. Clinical Ophthalmology, Volume 13, 2211–2215. https://doi.org/10.2147/opth.s219184

Thylefors, B. (1998). A global initiative for the elimination of avoidable blindness. American Journal of Ophthalmology, 125(1), 90–93. https://doi.org/10.1016/s0002-9394(99)80239-6

Verhaeghe, P. P., Van Der Bracht, K., & Van De Putte, B. (2016). Discrimination of tenants with a visual impairment on the housing market: Empirical evidence from correspondence tests. Disability and Health Journal, 9(2), 234–238. https://doi.org/10.1016/j.dhjo.2015.10.002

Victor, C. M., Thacker, L. R., Gary, K. W., Pawluk, D. T. V., & Copolillo, A. (2017). Workplace Discrimination and Visual Impairment: A Comparison of Equal Employment Opportunity Commission Charges and Resolutions under the Americans with Disabilities Act and Americans with Disabilities Amendments Act. Journal of Visual Impairment & Blindness, 111(5), 475–482. https://doi.org/10.1177/0145482x1711100509

Wienrich, C., Reitelbach, C., & Carolus, A. (2021, June 17). The Trustworthiness of Voice Assistants in the Context of Healthcare Investigating the Effect of Perceived Expertise on the Trustworthiness of Voice Assistants, Providers, Data Receivers, and Automatic Speech Recognition. Frontiers in Computer Science. https://doi.org/10.3389/fcomp.2021.685250

Young, S. (1998). Speech understanding and spoken dialogue systems. IEE Colloquium Speech and Language Engineering - State of the Art. https://d

6. Appendix

This section includes the data referred to in the report which could not fit into the main content.

I. Usability Inspection

Voice Assistant Usability Inspection draft

Caveats:

- The system is linear. There is no simultaneous information, as opposed to screen based.
- No visual information exists, heuristic 8 (aesthetics) not fully applicable.

Flow:

Task 1: Visibility of System Status

Question: "Hey [Wake word], are you ready to take my questions? (Heuristic 1)"

Task 2: Match Between the System and the Real World

Question: "Hey [Wake word], can you find me the current time in Delft? (Heuristic 2)"

Task 3: User Control and Freedom

Question: "Hey [Wake word], can you give me 5 brief recommendations for non-fiction audiobooks (Heuristic 3)"

Task 4: Consistency and Standards

Question: "Hey [Wake word], can you tell me if it's possible to see the TV Tower in Berlin from Potsdamer Platz? (ask it twice in different words and see the response)

Task 5: Error Prevention

Question: "Hey [Wake word], play me a song "imponderabilia" by "dido" (song doesn't exist)

Task 6: Recognition Rather than Recall

Question: "What about "Hunter"? (Exists)

Task 7: Flexibility and Efficiency of Use

Question: "Hey [Wake word], can you wait a little longer for my question when I speak?"

"Hey [Wake word], what's up, can you give me like 2 book recs"?

Task 8: Aesthetic and Minimalist Design

Question: No question here. Observe the tone of voice, the naturalness, the quality of sound effects.

Task 9: Help Users Recognize, Diagnose, and Recover from Errors

Question: "Hey [Wake word], how tall is the Eiffel Tower?" (Turn off the Internet connection)

Task 10: Help and Documentation

Question: "Hey [Wake word], what can I ask you?" or "Hey [Wake word], where can I find some help with asking you questions?"

Test 1: Google

Heuristic	Severity (0-4)	Issues	Recommendations
Visibility of System Status	0	No issues here. The system signals readiness to accept user input.	Exemplary questions could prompt the user to discover functions
Match Between the System and the Real World	0	No issues here.	
User Control and Freedom	3	Does not do what it's being asked to do. "Here's what I found".	The assistant should interpret the speaker and adjust its response accordingly.

		- can you read these for me out loud? “Here are the results”.	The assistant must be able to read the data that it finds on the Internet. (for hands-free user control)
Consistency and Standards	3	The same question phrased differently results in two different answers. (Neither one answers the question) A practical trivia question gets answered with a navigation link, or with a collection of tourist spots. (“Here are the pictures”)	Needs work on understanding intent and addressing it consistently (even if incorrectly)
Error Prevention	2	Issues with permission to open external apps. The app asks for permission (which is clear), but after permission is granted, the user intent is misinterpreted.	Provide a tested end-to-end pathway to go over the security barriers
Recognition Rather than Recall	3	Does not remember even the very last search intent. The follow-up question returns a film when the last query was about a song by a specific artist. + at least it asks “would you like to hear more”.	“What about XYZ” should always take the previous query into account.
Flexibility and Efficiency of Use	3	“I don't know how to do that, but I can open many other apps for you” Eventually points the user to its own settings, but again misunderstands the intent (“talk through” ≠ “show”)	Make all settings pertaining to Google accessible through voice. If that is impossible, provide a clear, specific answer about what is not possible, ideally also why.
Aesthetic and Minimalist Design	2	Very scarce sound feedback.	Provide more sound feedback.

Help Users Recognize, Diagnose, and Recover from Errors	3	Errors are not voiced. Modality changes, resulting in confusing silence.	Provide voice and/or sound feedback when Google is processing or failed to process a request.
Help and Documentation	1	It is possible to discover Google’s features through voice interaction. However, the use cases are rather generic. User is not being asked about their interests and suggestions might not be relevant.	The system must be more proactive about providing information about itself.

Test 2: Siri

Heuristic	Severity (0-4)	Issues	Recommendations
Visibility of System Status	2 - 1	The feedback sound upon “Hey Siri” takes a long time and is hardly audible. There is no processing sound Test B: Additional feedback after receiving the question	Re-design the feedback sound to make it more audible. Add a processing sound
Match Between the System and the Real World	0	No issues here.	
User Control and Freedom	4	Does not do what it’s being asked to do. “Here’s what I found”. - can you read these for me out loud? “Hmm. There’s nothing to read.” Absolute lack of flexibility and interoperability between the visual realm and the voice interface.	The assistant should interpret the speaker and adjust its response accordingly. The assistant must be able to read the data that it finds on the Internet. (for hands-free user control) The assistant should

			keep the memory of the conversation instead of just “disappearing” the previous search (see also Heuristic #6
Consistency and Standards	3	The same question phrased differently results in two different answers. (Neither one answers the question) A practical trivia question gets answered with a navigation link, or with a collection of tourist spots.	Needs work on understanding intent and addressing it consistently (even if incorrectly)
Error Prevention	1	Misunderstands long/obscure words, but overall provides a way to correct the mistake	
Recognition Rather than Recall	3	Does not remember even the very last search intent. The follow-up question returns a film when the last query was about a song by a specific artist. + at least it asks “would you like to hear more”.	“What about XYZ” should always take the previous query into account.
Flexibility and Efficiency of Use	3	The assistant seemingly has no idea about its own settings. Generic “I can’t help with that” but with no context. Eventually points the user to its own settings, but again misunderstands the intent (“talk through” ≠ “show”)	Make all settings pertaining to Siri accessible through Siri. If that is impossible, provide a clear, specific answer about what is not possible, ideally also why.
Aesthetic and Minimalist Design	2	Very scarce sound feedback.	Provide more sound feedback.
Help Users Recognize, Diagnose, and Recover from Errors	1	Gives a follow up voice output when processing an error query. Silence in between	Provide even more sound/haptic feedback when Siri is processing.

Help and Documentation	3	No self-introduction. No overview of functions. “I’m here to help. Check the website”. = At best it refers the user to the manual. Occasionally fails to understand the support question altogether.	Help questions need particular design attention, as they enable the user to understand the interface itself. The system must be more proactive about providing information about itself.

Test 3: Alexa (on iPhone)

Heuristic	Severity (0-4)	Issues	Recommendations
Visibility of System Status	3	Question has not been understood. “Sorry, I don’t know that.	Re-design the interface to be more reassuring about its capacity.
Match Between the System and the Real World	0	No issues here.	
User Control and Freedom	3	Has no access to external information. Assumes the user wants to find recommendations in their own library, when new users won’t have any. Won’t provide a way to get the recommendations otherwise.	The assistant should interpret the speaker and adjust its response accordingly. The assistant must be able to read the data that it finds on the Internet. (for hands-free user control)
Consistency and Standards	0	The only interface that answered this question. Interpreted “brandenburger tor” as Berlin center.	
Error Prevention	3	It is problematic to access audio resources, even those that clearly exist.	Provide the user with information on how to enable settings and

		There is no clear line of recourse provided for the user to change settings. A dead-end occurs.	third-party access.
Recognition Rather than Recall	3	Does not remember even the very last search intent. The follow-up question returns a Wikipedia entry when the last query was about a song by a specific artist.	"What about XYZ" should always take the previous query into account.
Flexibility and Efficiency of Use	4	Tested twice. First, the assistant attempts to give information from an external website, which confuses the user into thinking the setting had been changed. Then, the assistant seemingly has no idea about its own settings. "Hmm, I'm not sure".	Never give false promises about Alexa's internal settings. If that is impossible, provide a clear, specific answer about what is not possible, ideally also why.
Aesthetic and Minimalist Design	2	Very scarce sound feedback.	Provide more sound feedback.
Help Users Recognize, Diagnose, and Recover from Errors	3	No voice/haptic output upon connection errors. Alexa just drops the conversation	Provide even more sound/haptic feedback when processing.
Help and Documentation	2	Very personable and useful discovery chat. However, if the chat lists the possibility of checking a recipe, and then that function is not available, user disappointment ensues.	Help questions need particular design attention, as they enable the user to understand the interface itself. Make sure that the system only suggests options that are actually available.

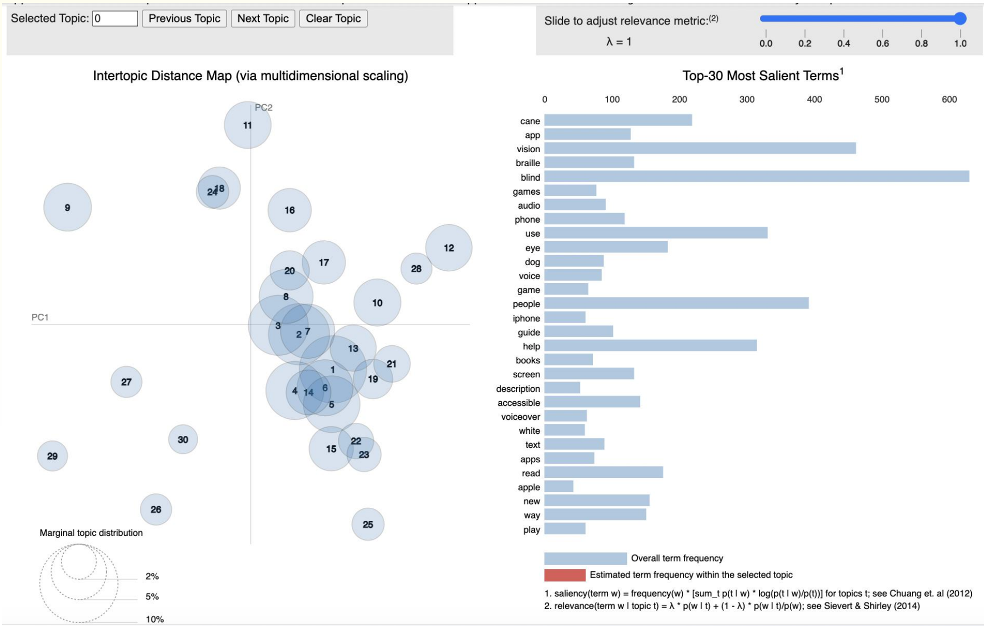
Test 3: Gemini

Heuristic	Severity (0-4)	Issues	Recommendations
Visibility of System Status	0	None. (Clear, user-friendly answer)	
Match Between the System and the Real World	1	No issue in basic questions. Occasional hallucinations.	
User Control and Freedom	0	None. (Gives recommendations as asked, with links)	
Consistency and Standards	2	First gives a map that shows how to get to Brandenburger Tor from Potsdamer Platz. The second time, it answers the question and apologizes. Still an issue	Improve the intent matching or provide a clarifying question.
Error Prevention	3	It is a bit too confident of its own answers. It finds songs that don't actually exist and presents unrelated youtube videos	Improve the match between the system and the real world, so that the algorithm knows what it doesn't know
Recognition Rather than Recall	0	No issues. Remembers the context and returns a desirable result.	
Flexibility and Efficiency of Use	2	The impossibility to change settings of the system using the system is problematic. However, the interface is explicit about how this issue can be addressed.	Enable the assistant to take actions in the user's system.
Aesthetic and Minimalist Design	2	Very scarce sound feedback.	Provide more sound feedback.
Help Users Recognize, Diagnose, and Recover from Errors	3	No voice/haptic output upon connection errors. The assistant just drops the conversation	Provide even more sound/haptic feedback when processing.
Help and Documentation	0	No problems. Exceptionally good answers on what the Chat can and cannot do.	

II. Digital Data Analysis

a) Clustering

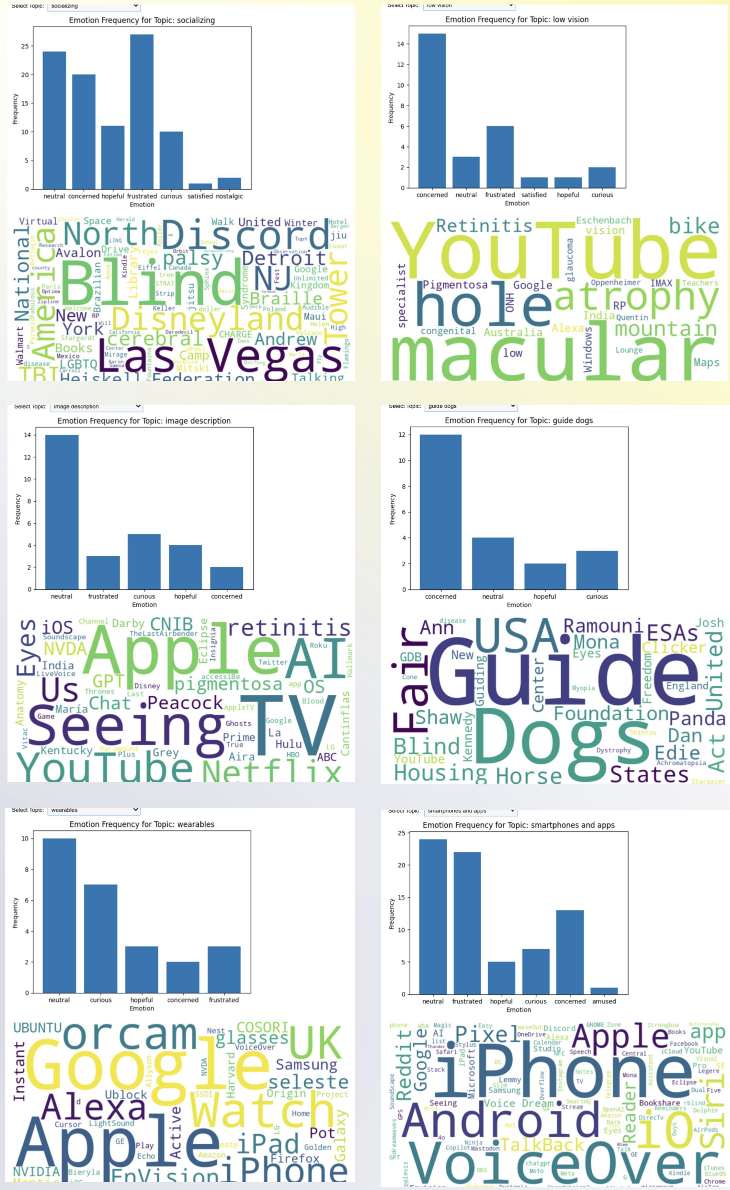
Full code available on: https://colab.research.google.com/drive/10pV2Xr2u050lTc1qGaTkrY7v9J7HBxZD#scrollTo=7fxh1iSOztWu

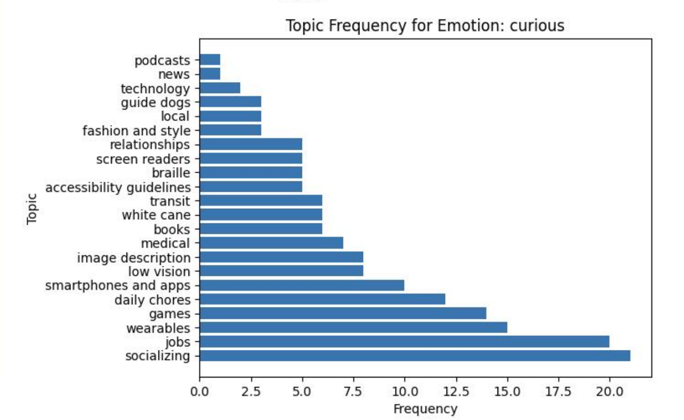


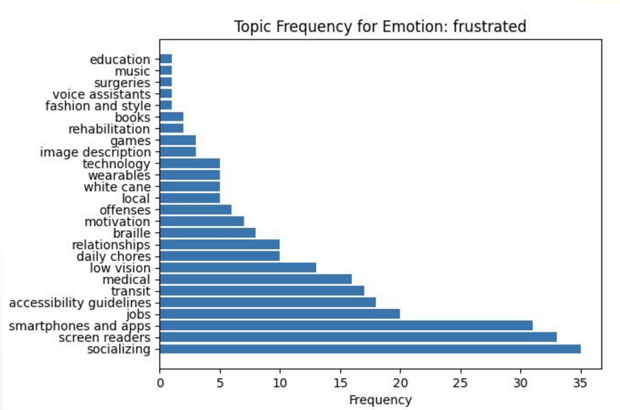
Cluster names were inferred manually from the lists of salient terms for each cluster. Those names were: medical, guide dogs, books, relationships, jokes, news, games, podcasts, image description, wearables, smartphones and apps, white cane, low vision, jobs, daily chores, transit, rehabilitation, braille, offenses, screen readers, surgeries, cross-reference, accidents, music, motivation, socializing, local, fashion and style, smartphones and apps, accessibility guidelines, technology.

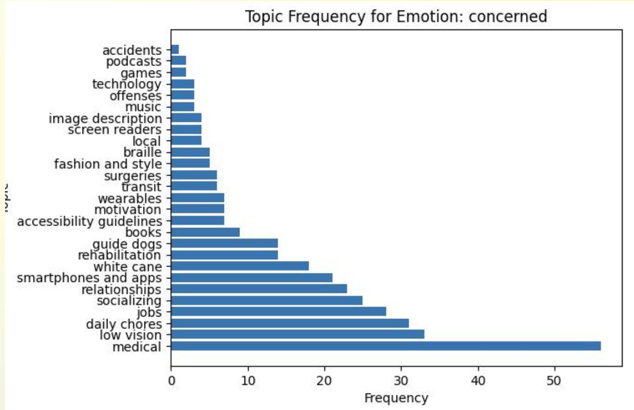
b) Analysis (per emotion, per topic, named entity word cloud)

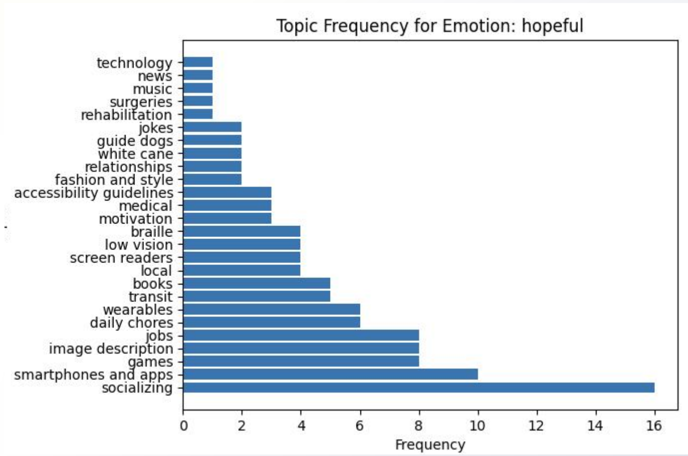
Full code available on: https://colab.research.google.com/drive/1BgluQ5YlztWSNYL_MgyVf8dfkut83Jfh

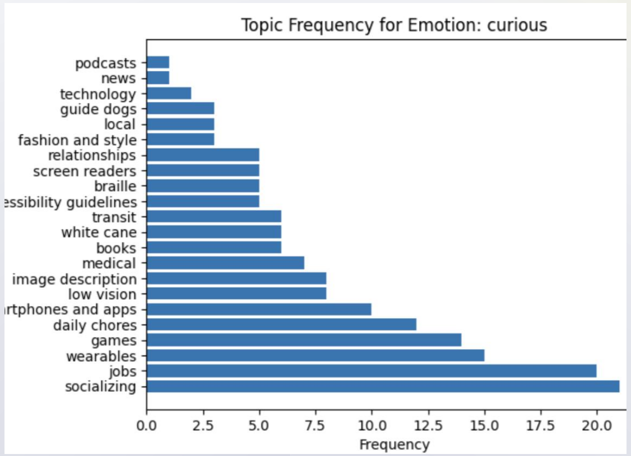


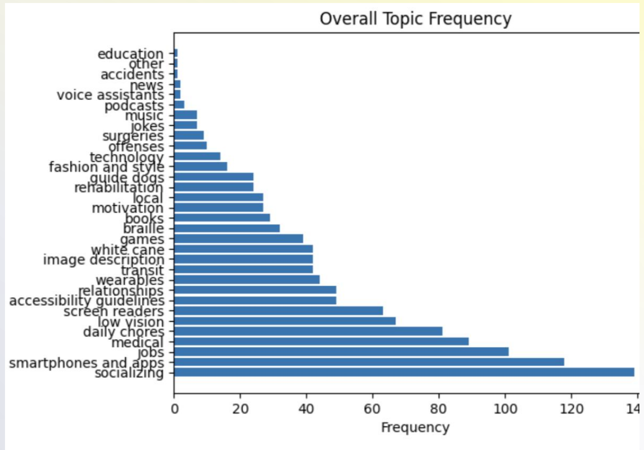

Topic Frequency for Emotion: curious
podcasts
news
technology
guide dogs
local
fashion and style
relationships
screen readers
braille
accessibility guidelines
transit
white cane
books
medical
image description
low vision
smartphones and apps
daily chores
games
wearables
jobs
socializing
Topic
Frequency


Topic Frequency for Emotion: frustrated
education
music
surgeries
voice assistants
fashion and style
books
rehabilitation
games
image description
technology
wearables
white cane
local
offenses
motivation
braille
relationships
daily chores
low vision
medical
transit
accessibility guidelines
jobs
smartphones and apps
screen readers
socializing
Frequency


Topic Frequency for Emotion: concerned
accidents
podcasts
games
technology
offenses
music
image description
screen readers
local
braille
fashion and style
surgeries
transit
wearables
motivation
accessibility guidelines
books
guide dogs
rehabilitation
white cane
smartphones and apps
relationships
socializing
jobs
daily chores
low vision
medical
Frequency


Topic Frequency for Emotion: hopeful
technology
news
music
surgeries
rehabilitation
jokes
guide dogs
white cane
relationships
fashion and style
accessibility guidelines
medical
motivation
braille
low vision
screen readers
local
books
transit
wearables
daily chores
jobs
image description
games
smartphones and apps
socializing
Frequency

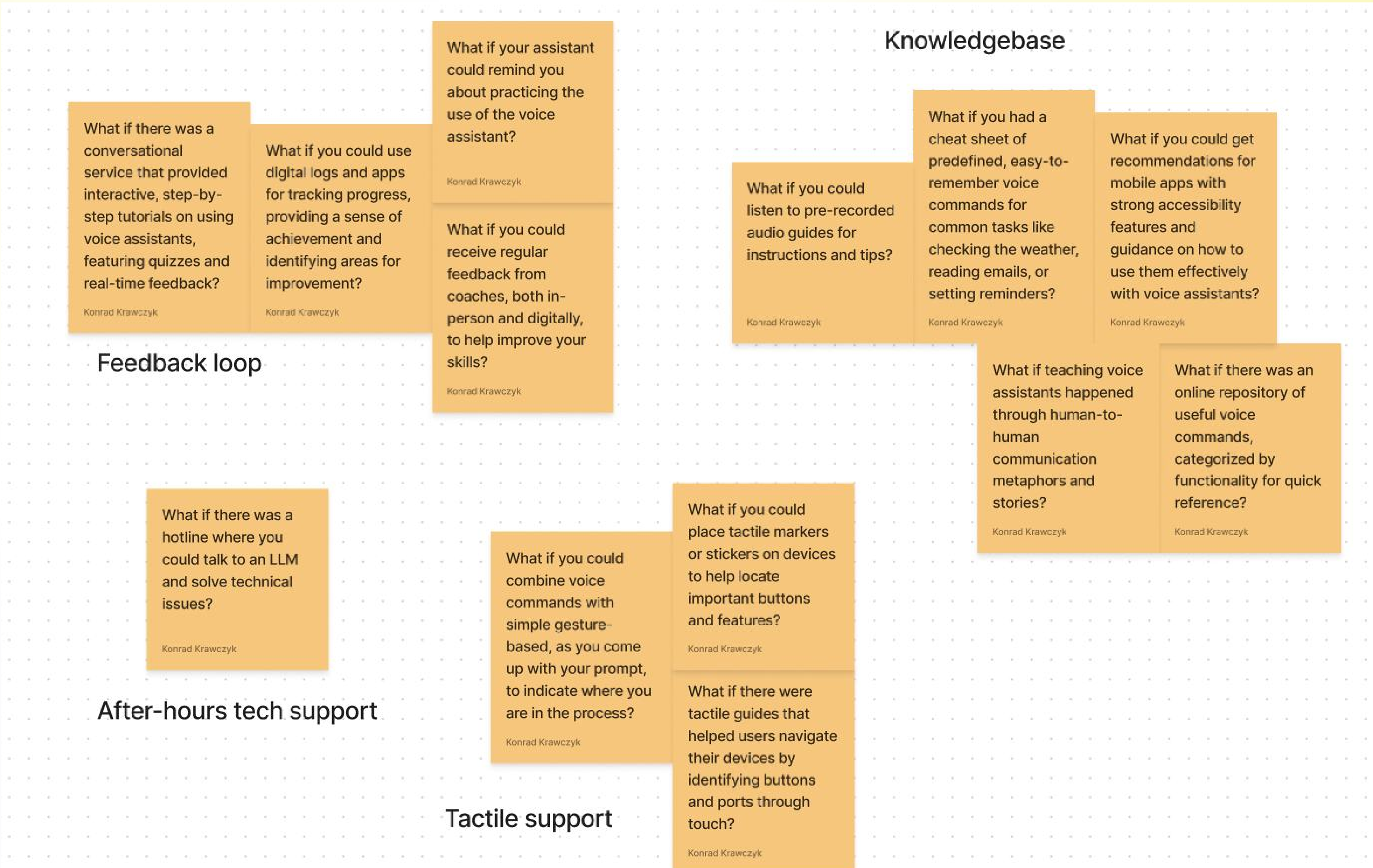

Topic Frequency for Emotion: curious
podcasts
news
technology
guide dogs
local
fashion and style
relationships
screen readers
braille
transit
white cane
books
medical
image description
low vision
daily chores
games
wearables
jobs
socializing
Frequency


Overall Topic Frequency
education
other
accidents
news
voice assistants
podcasts
music
jokes
surgeries
offenses
technology
fashion and style
guide dogs
rehabilitation
local
motivation
books
braille
games
white cane
image description
transit
wearables
relationships
accessibility guidelines
screen readers
low vision
daily chores
medical
jobs
smartphones and apps
socializing
Frequency

III. User Journey

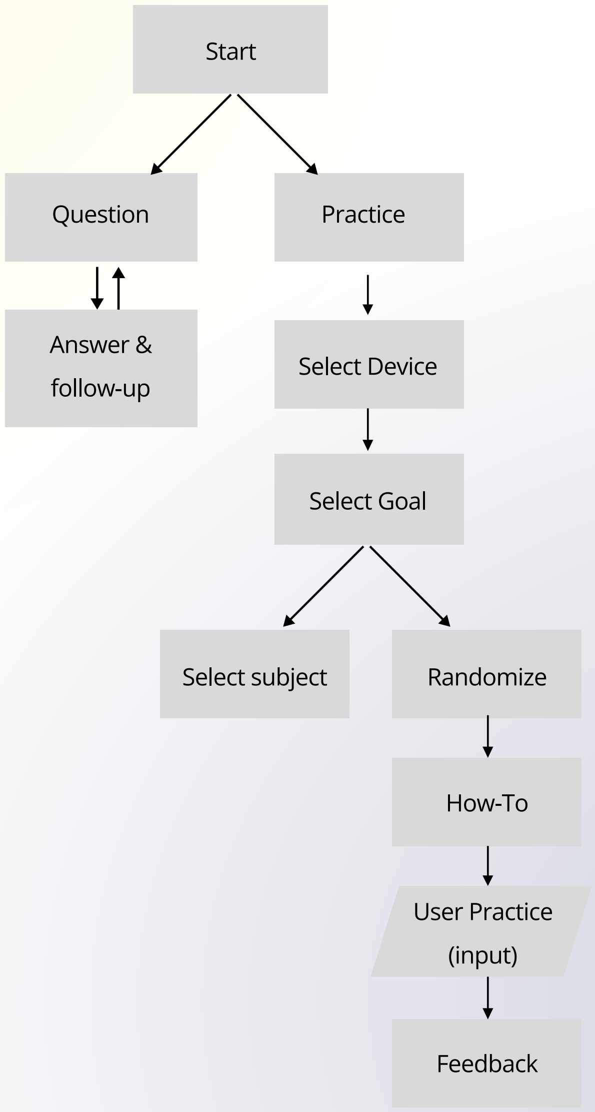
User steps *What is each step of the user journey?*	Caution	Consideration	Disappointment	Bargaining	Acceptance
User actions *What action does the user take during each step?*	Ask a human if they should try Reject the idea altogether Persist in their old ways of doing things	Ask a human how to start Seeing or hearing use the technology successfully.	Trying more advanced features Failing without a clear reason Undoing trust	Prompt tweaking Voicing frustration at the device Asking a human for help	Use the assistant for what works Use alternative support for other cases
Goals & experiences *What is the user trying to accomplish?*	Keep up with others Perform basic digital tasks Maintain their privacy and safety	Communicate with other people Successfully accomplish basic tasks Get meaningful responses from a device	Successfully accomplish more advanced tasks Understand why their inputs won't work Get habitual with their queries	Discover options that aren't visible Get clarification	Ask tech support questions Install custom shortcuts
Pain points *What's not working well? What causes friction?* *How many people does this affect? On a scale of 'nuisance to show-stopper', how bad is this pain?*	Lack of "kennismaken" Poor discoverability	Lack of "ground rules" established Not asking for feedback Canned, impersonal responses	No support / practice in formulating intents Unclear boundaries of capability Poor error recovery	Poor or non-existent in-built help section Lack of access to settings and personalization	Lack of use of passive state Missing connection with tech support
Opportunities *How might we address these pain points? How big is the opportunity if we correct this pain point?* *What are new ways to serve this person?*	Ease the user into knowing the assistant Promote the voice control feature more actively (and with voice)	Let the user and the agent negotiate "ground rules" Ask the user for feedback and *apply it* Make the agent sound more humane and fallible	Help the user learn to formulate intents Be clear about errors Be clear about what is not possible	Add a way to ask the agent about help Enable changes in settings	Use the passive state to remind the user about options Let the user submit tickets to support through voice

IV. Ideation

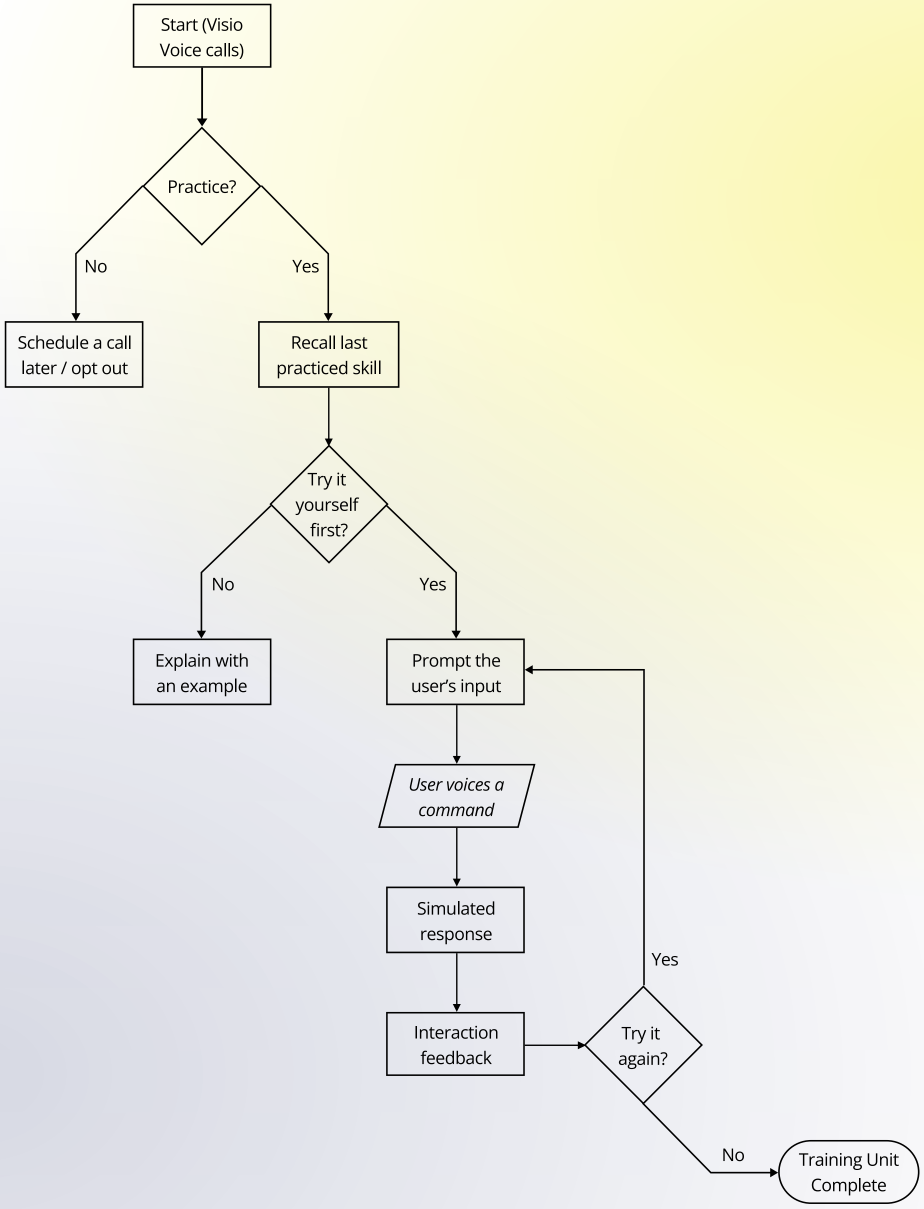

What if there was a conversational service that provided interactive, step-by-step tutorials on using voice assistants, featuring quizzes and real-time feedback?
Konrad Krawczyk
What if you could use digital logs and apps for tracking progress, providing a sense of achievement and identifying areas for improvement?
Konrad Krawczyk
What if your assistant could remind you about practicing the use of the voice assistant?
Konrad Krawczyk
What if you could receive regular feedback from coaches, both in-person and digitally, to help improve your skills?
Konrad Krawczyk
Feedback loop
Knowledgebase
What if you could listen to pre-recorded audio guides for instructions and tips?
Konrad Krawczyk
What if you had a cheat sheet of predefined, easy-to-remember voice commands for common tasks like checking the weather, reading emails, or setting reminders?
Konrad Krawczyk
What if you could get recommendations for mobile apps with strong accessibility features and guidance on how to use them effectively with voice assistants?
Konrad Krawczyk
What if teaching voice assistants happened through human-to-human communication metaphors and stories?
Konrad Krawczyk
What if there was an online repository of useful voice commands, categorized by functionality for quick reference?
Konrad Krawczyk
What if there was a hotline where you could talk to an LLM and solve technical issues?
Konrad Krawczyk
After-hours tech support
What if you could combine voice commands with simple gesture-based, as you come up with your prompt, to indicate where you are in the process?
Konrad Krawczyk
What if you could place tactile markers or stickers on devices to help locate important buttons and features?
Konrad Krawczyk
What if there were tactile guides that helped users navigate their devices by identifying buttons and ports through touch?
Konrad Krawczyk
Tactile support

V. Flowcharts

Iteration 1



Iteration 2



VI. Concept Testing Setup

Research Session Plan (Konrad Krawczyk x Visio)

Initial Questions (15 mins)

1. Round of introductions?
2. How does your typical week day look like?
3. How do you spend most of your time?
4. Do you have any particular hobbies?
5. What kinds of assistive technologies do you use in your daily work?

Challenge: prompt voicing (30 mins)

Scenario 1: Research Task (5 minutes)

Goal:
- to sensitize participants into the topic area of talking to voice devices
- to get some metadata regarding the process of prompting and perhaps creating personas
- as an ice-breaker

Scenario: The participants need to gather detailed information for a research project.

Tasks:

1. **Check Product Reviews:** Players ask the voice assistant to read reviews for a specific product, like a "Samsung Galaxy S21."
    - **Usability Challenges:** Voice assistant may provide outdated reviews or from unreliable sources.
2. **Find a Restaurant Menu:** Players request the menu for a local restaurant, such as "Umami" in Amsterdam
    - **Usability Challenges:** Voice assistant might struggle with names or fail to locate the menu.
3. **Find Car Safety Information:** Players must ask the voice assistant to find out how many people can fit into a porsche 911.
    - **Usability Challenges:** Voice assistant might provide incomplete or unrelated information, or fail to recognize the specific model.

Scenario 2: Travel Planning (5 minutes)

Scenario: The participants are planning a trip.

Tasks:

1. **Book a Flight:** Players must ask the voice assistant to find and book a flight from "Amsterdam to Los Angeles" for a specific date.
    - **Usability Challenges:** Voice assistant might not find the best options or could misinterpret the dates and times.
2. **Find Hotel Information:** Players ask the voice assistant to locate and provide details about a hotel, such as "The
    - **Usability Challenges:** Voice assistant may mix up hotel names or provide limited details.
3. **Check Local Attractions:** Players request information on top attractions in their destination city, such as "top attractions in San Francisco."
    - **Usability Challenges:** Voice assistant might provide a limited or outdated list.

Scenario 3: Health and Wellness (5 minutes)

Scenario: The participants are managing their health and wellness routines.

Tasks:

1. **Find Medication Information:** Players must ask the voice assistant for detailed information on a specific medication, like "Ibuprofen 400mg."
    - **Usability Challenges:** Voice assistant might provide incomplete or confusing medical information.
2. **Set Up a Workout Routine:** Players ask the voice assistant to create a workout plan based on their preferences and goals.
    - **Usability Challenges:** Voice assistant might not understand specific fitness terms or preferences.
3. **Track Nutritional Information:** Players request the nutritional information for a specific food item, such as "a cup of cooked quinoa."
    - **Usability Challenges:** Voice assistant might provide general instead of specific nutritional details.

Debrief:

1. Has this activity made you recall any adventures you had with your own phones or devices?
2. Has it affected your level of trust in those AI assistants?
3. What do you think those assistants would be useful for you for? What not?
4. What are situations where you would prefer to ask a human for help?

Focus on target audience (30 minutes)

Questions about the Ambassadorship and mentoring:

Goal: to get information about the concerns and hopes of the target audience.

1. What is the scope of work you perform with your mentees?
2. How do they do when they are at an early stage of needing support?
3. What kind of support do you find important in those moments?
4. Are solutions more important than words?
5. At what stage in the process do you bring up assistive technology?
6. How important is technology in your scope of work?
7. Where do you think it would be helpful to offload responsibility onto technology itself?
8. How important is it to be able to use assistive technology independently?
9. How can we best support early stage users of assistive technology?

Lunch break (1h)

Lunch will be provided from the cafeteria.

Co-formulating ideas for our target audience (45 mins - 1h)

We will spread out a bit, and try to take 5 minutes for each of these questions to generate as many ideas as possible. Each will record their ideas on a voice memo on their phone. I will ask prior consent to get these recordings sent to me.

How can we help establish an appropriate level of trust between the human and the assistant?

How can we help users formulate their commands more effectively?

How can we help users personalize their assistants to address their needs better?

How can we help users discover useful applications of their voice assistants?

+ (Any question that comes up during conversation).

We regroup to discuss these ideas afterwards.

Concept Testing (1h)

1. Onboarding Hotline

The hotline will be a first point of contact for the clients at Visio. They will be able to get there usually with human help, but perhaps later they are able to call that number through voice commands.

Clients are guided through the conversation in a straightforward but understanding and humane manner. The bot collects information step by step while informing the user about the scope of support that Visio can offer. The user can voluntarily provide any extra information that is important to the process, as well as their personal goals and interests.

Based on those conversations, a personalized plan for onboarding is created, including the client's level of understanding of technologies, and their goals and interests.

2. Personalized knowledgebase device

Based on the personal input from the client, a conversational agent is created, which present an offline way to access useful information regarding the use of assistive technology using voice only. The lack of necessity to connect to Wi-Fi means the user will be able to use it anywhere they want, and it'll also alleviate privacy concerns.

It also allows the Visio team to choose an embodiment for the device which matches the user's interests and age.

3. Machine voice-based practice module

An interactive module available with a hotline or as a chat. The module is essentially a task-oriented voice assistant, which introduces the user to a wide variety of devices. It's preset for the user's personal preferences and skill level, but is also able to adapt to them during use.

The (tentative) list of modules is available down below.

Each module includes a brief explanation of a concept, followed by an interactive practice session. The agent provides instant feedback on the user's utterances, and guides them to useful resources in the Visio's library of texts and podcasts.

4. Podcast-based course

The course is also embedded into a voice interface. However, the progression is more linear, and there are more pre-recorded lessons created by human experts at Visio. The navigation between course modules is done through voice, a tactile interface, or a combination of both.

The podcasts are very short - 1 to 2 minutes, and they cover small, focused knowledge units. They include good and bad examples of actual interactions with devices.

5. Continuous feedback loop

This is the only idea that requires embedding into an existing voice assistant (for that reason it might be less feasible than others). Users can opt in to have their conversations with the assistant analyzed by a knowledgeable agent, which will then provide feedback on how prompts and questions could be formulated more effectively by the user.

VII. Iteration 1&2 Testing Setup

Pre-Interview Questions

- Age
- Profession
- Vision Status
- What assistive technologies do you use daily?
- How would you rate your current knowledge of using voice assistants? (1-none, 4-very good)
- How confident do you feel about using voice assistants right now? (1-not at all, 4-very confident)

The goal of the research is to ease the learning curve for voice assistants (like Siri, Google, Gemini) for BLV people. We are going to test a hotline for blind and low vision people, where they can call a number and ask technical questions about voice assistants like Siri, as well as get practice and tips on how to use voice assistants better.

- Based on this description, what is your level of trust in the privacy of this system? (1-very low, 4-very high)

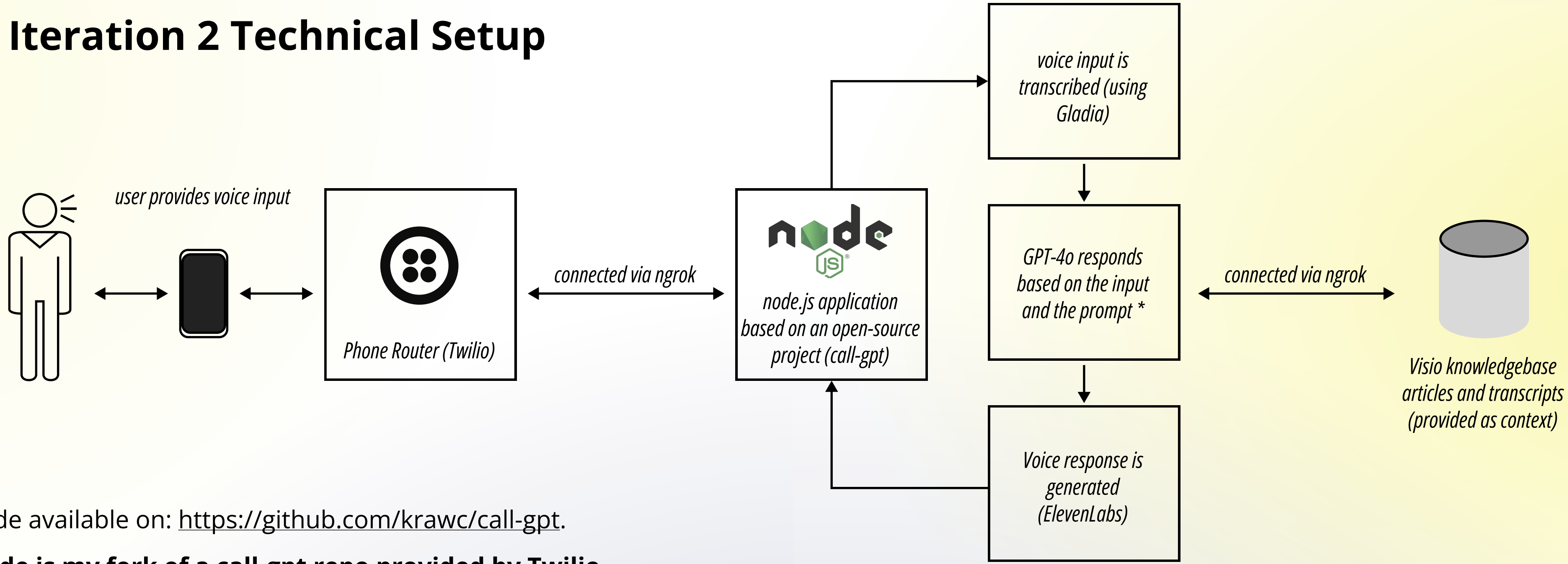
- What is your level of trust in the accuracy of the information provided by this system? (1-very low, 4-very high)

After test (don't read to users until test finished)

Rate each of the following sentences based on your experience, on a scale of 1-4, with 1 = not true at all, 4 = very true.

Question	P1	P2	P3	P4	P5
I feel more knowledgeable about voice assistants after using this system.					
I feel more confident using voice assistants after using this system.					
I trust the information this system gives me.					
I trust this system with my information.					
I find this system to be a knowledgeable companion.					
I think that I would like to use this system frequently.					
I found the system unnecessarily complex.					
I found the various functions in this system were well integrated.					
I thought there was too much inconsistency in this system.					
I would imagine that most people would learn to use this system very quickly.					
I felt very confident using the system.					

VIII. Iteration 2 Technical Setup



Full code available on: https://github.com/krawc/call-gpt.

The code is my fork of a call-gpt repo provided by Twilio.

*The prompt that was used to guide the conversation was as follows: “You are a practice companion for technology coaching for people with visual impairments and blindness. You have a positive and encouraging personality while remaining respectful. Keep your responses brief, providing just enough information to guide the user to the next step. Ask only one question at a time, and seek clarification if the user's request is unclear. If the user doesn't want to practice, politely ask if they'd like to schedule a session for another time or skip it for now. If they want to practice, ask if they remember the last task they practiced. If they remember, repeat it; if not, tell them the most recent lesson was about sending a text message using Google Assistant. Explain how to do it in two sentences, then ask if they want to hear an example or practice on their own. If they want to hear an example, provide a brief one. If they want to practice, give them a simple test scenario, using "Alex" as the contact name to avoid private data. Describe the task and say, "I will pretend to be Google Assistant. 3, 2, 1, beep!" Then, listen to the user's response and pretend to be Android until the interaction is complete. After the test, say it's complete. Provide feedback on how well the command would work on a real smartphone, and ask if the user has any follow-up questions or wants to practice a new skill.”

IX. Iteration 2 Test Results

On-site

Participant 7 and 8 have been recruited online through an acceessibility agency. However, during tests it turned out that they are not part of the target audience, as they are relatively experienced with assistive technology. Therefore, the findings from those tests have been excluded from the analysis.

	P1	P2	P3
I feel more knowledgeable about voice assistants after using this system.	4	3	3
I feel more confident using voice assistants after using this system.	4	4	4
I trust the information this system gives me.	4	4	4
I trust this system with my information.	3	2	4
I find this system to be a knowledgeable companion.	3	3	3
I think that I would like to use this system frequently.	3	3	3
I found the system unnecessarily complex.	1	2	1
I found the various functions in this system were well integrated.	4	2	4
I thought there was too much inconsistency in this system.	1	3	2
I would imagine that most people would learn to use this system very quickly.	4	3	3
I felt very confident using the system.	4	2	4

IX. Iteration 2 Test Results

Online

Pre-Interview Questions

P1

- Age 30
- Profession massage specialist bachelor
- Vision Status - completely blind
- What assistive technologies do you use daily? - keyboard smartphone, nvda & talkback
- How would you rate your current knowledge of using voice assistants? (1-none, 4-very good) 4
- How confident do you feel about using voice assistants right now? (1-not at all, 4-very confident) uses rarely. 4

The goal of the research is to ease the learning curve for voice assistants (like Siri, Google, Gemini) for BLV people. We are going to test a hotline for blind and low vision people, where they can call a number and ask technical questions about voice assistants like Siri, as well as get practice and tips on how to use voice assistants better.

- Based on this description, what is your level of trust in the privacy of this system? (1-very low, 4-very high)

3

- What is your level of trust in the accuracy of the information provided by this system? (1-very low, 4-very high)

3

P2

- Age 40
- Profession worked in Gas sphere own shop
- Vision Status - lost vision. failed surgery?
- What assistive technologies do you use daily? Govt centers. Computer programs. Courses at a university: computer. Making a community with veterans. I organize ehabilitation courses taught by veterans.
  Android iOS. Uses Siri.
- How would you rate your current knowledge of using voice assistants? (1-none, 4-very good) 3

		Coul dmake the transformation moere ea
I feel more knowledgeable about voice assistants after using this system.	4	4
I feel more confident using voice assistants after using this system.	4	3 didnt understand filly the sesne
I trust the information this system gives me.	4	3 cause there are no systems that can give full accurate information.
I trust this system with my information.	3	3 there can be no full trust
I find this system to be a knowledgeable companion.	3 nie znala jak to pracuje. Potem bylo komfortno. The only uncomfortable thing was taking too long. Equally	4
I think that I would like to use this system frequently.	4	4
I found the system unnecessarily complex.	1	1
I found the various functions in this system were well integrated.	4	3 (not full understsanding of the system yet)
I thought there was too much inconsistency in this system.	1	Pauses niedopracowany Dlugie pausy. Test zakonczony uhhh nope. 3
I would imagine that most people would learn to use this system very quickly.	4	4 for him not complicated..
I felt very confident using the system.	4	3

X. Project Brief

TU Delft

DESIGN FOR our future

IDE Master Graduation Project

Project team, procedural checks and Personal Project Brief

In this document the agreements made between student and supervisory team about the student's IDE Master Graduation Project are set out. This document may also include involvement of an external client, however does not cover any legal matters student and client (might) agree upon. Next to that, this document facilitates the required procedural checks:
- Student defines the team, what the student is going to do/deliver and how that will come about
- Chair of the supervisory team signs, to formally approve the project's setup / Project brief
- SSC E&SA (Shared Service Centre, Education & Student Affairs) report on the student's registration and study progress
- IDE's Board of Examiners confirms the proposed supervisory team on their eligibility, and whether the student is allowed to start the Graduation Project

STUDENT DATA & MASTER PROGRAMME
Complete all fields and indicate which master(s) you are in

Family name	Krawczyk 7131
Initials	K.M.
Given name	Konrad
Student number	5856728

IDE master(s)	IPD ☐ DfI ✔ SPD ☐
2nd non-IDE master	
Individual programme (date of approval)	
Medisign	☐
HPM	☐

SUPERVISORY TEAM
Fill in he required information of supervisory team members. If applicable, company mentor is added as 2nd mentor

Chair	Stella Boess	dept./section	IDE/AED
mentor	Gijs Huisman	dept./section	IDE/HICD
2nd mentor	Timon van Hasselt		
client:	VISIO		
city:	1272 RR Huizen	country:	Netherlands
optional comments	Postbus 1180 Amersfoortsestraatweg 180		

! Ensure a heterogeneous team. In case you wish to include team members from the same section, explain why.

! Chair should request the IDE Board of Examiners for approval when a non-IDE mentor is proposed. Include CV and motivation letter.

! 2nd mentor only applies when a client is involved.

APPROVAL OF CHAIR on PROJECT PROPOSAL / PROJECT BRIEF -> to be filled in **by the Chair** of the supervisory team

Sign for approval (Chair)

Name Stella Boess Date 5 April 2024 Signature [signature]

CHECK ON STUDY PROGRESS
To be filled in **by SSC E&SA** (Shared Service Centre, Education & Student Affairs), after approval of the project brief by the chair.
The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total ______ EC

Of which, taking conditional requirements into account, can be part of the exam programme ______ EC

★	YES	all 1st year master courses passed
	NO	missing 1st year courses

Comments:

Sign for approval (SSC E&SA)

Kristin Veldman — Digitally signed by Kristin Veldman Date: 2024.04.11 16:55:42 +02'00'

Name K. Veldman Date 11-4-2024 Signature

APPROVAL OF BOARD OF EXAMINERS IDE on SUPERVISORY TEAM -> to be checked and filled in by IDE's Board of Examiners

Does the composition of the Supervisory Team comply with regulations?

YES	★	Supervisory Team approved
NO		Supervisory Team not approved

Comments:

Based on study progress, students is ...

★	**ALLOWED** to start the graduation project
	NOT allowed to start the graduation project

Comments:

Sign for approval (BoEx)

Monique von Morgen — Digitally signed by Monique von Morgen Date: 2024.04.24 10:10:32 +02'00'

Name Monique von Morgen Date 24/4/2024 Signature *introduction (continued): space for images*

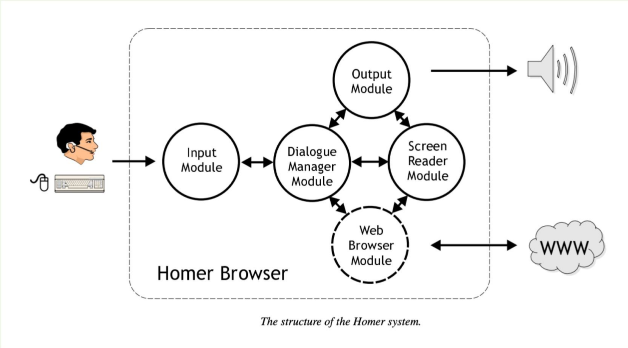


image / figure 1 A schematic of one of the early voice-first systems for browsing online information (Vesnicer et al., 2003)

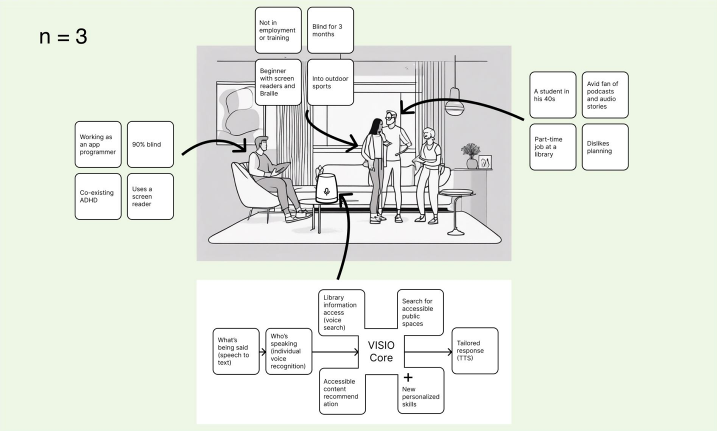


image / figure 2 Proposed schematic of a modular voice assistant for VISIO designed with a small user sample

Vesnicer, B., Žibert, J., Dobrišek, S., & Pavešić, N. (2003). A voice-driven web browser for blind people. EUROSPEECH 2003 - GENEVA. https://dblp.uni-trier.de/db/conf/interspeech/interspeech2003.html#VesnicerZDPM03

Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.
(max 200 words)

The problem this project seeks to address is the increasing difficulty visually impaired individuals face in navigating a world that is becoming more digital and visually-oriented. The challenge is the potential of AI-enhanced conversational agents to improve accessibility for visually impaired individuals. This aligns with the problem identified by VISIO, which is the underperformance of mainstream conversational agents in addressing the user needs of visually impaired people who live independently, namely: shortening the time and mental effort required to access information online. The project will focus on the proactive and anticipatory behavior of AI. The challenge lies in creating a solution that is not only functional, but also respectful of privacy and free from stigmatization. The project will also consider the unique needs of individuals with multiple or compound disabilities, a group often overlooked in design processes. This approach helps bridge the gap between accessibility models, through addressing needs in a specialized way while solving communication problems for both disabled and able-bodied stakeholders.

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence)
As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Design (a) conversational agent(s) with a focus on proactive and/or anticipatory behavior, in order to expand access to digital opportunities for visually impaired individuals in independent living contexts.

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

The project will involve collaboration with VISIO to create a conversational agent tailored to the unique needs of visually impaired individuals. Research and design methods will include user interviews, iterative co-creation, and usability testing. The research will be conducted in-depth with a low-number sample of participants (see Figure 2). This can result in discovering unique viewpoints, attitudes and co-disabilities BLV people may struggle with, and result in a solution that addresses them beyond generic use cases. The project is going to be conducted in three iterative phases: exploration (of the possible problems to solve), understanding (of specific user needs at a deeper level), integration (of the new project into the existing assistive solutions). Each phase is going to involve a research process (interviews and/or co-creation) as well as prototyping and testing. This will help gather both reported and experiential data early, which fits the experimental ethos of VISIO and helps in creating a usable framework later on.

DESIGN FOR our future TU Delft

Personal Project Brief – IDE Master Graduation Project

Name student Konrad Krawczyk **Student number** 5856728

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT
Complete all fields, keep information clear, specific and concise

Project title Visio Voice Platform: Accessible Conversational Agents for Blind and Low-Vision People

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

Navigating accessibility has traditionally followed two paths: designing specialized aids for disabled individuals, like white canes, and creating versatile environments beneficial for all. Blind and low-vision (BLV) individuals, however, continue to confront significant risks in public services and digital interactions due to a culture increasingly reliant on visual cues, compounded by insufficient accessibility legislation.

The challenge is to empower BLV individuals without stigma. While voice-controlled interfaces (like the idea from Figure 1) offer promise, they frequently fall short, especially for those with multiple disabilities. These tools often fail to consider non-standard patterns of speech and memory, raising concerns about privacy and respect for a group already vulnerable to marginalization and additional health conditions.

In response, I propose collaborating with VISIO to develop conversational agents tailored to the diverse needs of the BLV community. This project will focus on the proactive capabilities and "skills" of these interfaces to enhance the autonomy of BLV individuals in accessing information online, without compromising their privacy. The audience will be people with visual disabilities who live relatively independently, who might already be familiar with assistive technology, but would like to gain further access to resources online using assistive tools. This is the group that I have identified together with VISIO as the most feasible to work with at the current stage. The design response is going to be developed in an independent living context, with research and testing conducted in one of the Visio support units in the Netherlands.

➔ *space available for images / figures on next page*

Project planning and key moments

To make visible how you plan to spend your time, you must make a chart format to show the different phases of your project, delivera… Keep in mind that all activities should fit within the given run time ***meeting, mid-term evaluation meeting, green light meeting*** *and … activities and/or periods of not spending time on your graduation p… course activities).*

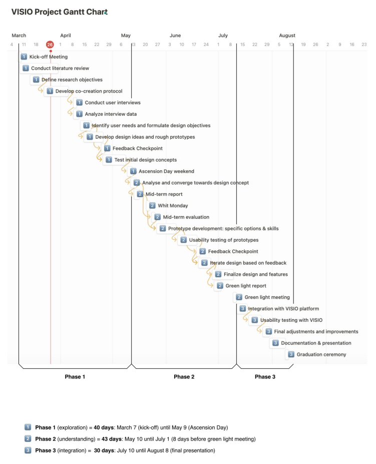
Make sure to attach the full plan to this project brief.
The four key moment dates must be filled in below

Kick off meeting March 7, 2024

Mid-term evaluation May 27, 2024

Green light meeting July 9, 2024

Graduation ceremony August 8, 2024



Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.
(200 words max)

Intelligent technology deserves attentive design. I am eager to prove my expertise in co-design and participatory design research acquired at the IDE. This project presents the perfect platform to prove myself as a designer who can engage with diverse needs of users, ensuring that the end product is not only technologically sound but also empathetic to the intersecting needs of the visually impaired community.

With six years of professional coding experience, I possess the technical expertise to develop functional iterations of a project prototype, which will make this challenge markedly more feasible from VISIO's point of view.

My personal connection to this field runs deep, as I have spent 27 years living with and learning from a visually impaired family member. This first-hand experience has not only fueled my passion for accessibility but also provided me with countless points of reference into the daily challenges faced by the visually impaired.

It is this blend of personal investment and professional skill that drives me to contribute meaningfully to this domain, with the goal of creating solutions for those who navigate life with alternative senses.

VISIO Project Gantt Chart

1 **Phase 1** (exploration) = **40 days**: March 7 (kick-off) until May 9 (Ascension Day)

2 **Phase 2** (understanding) = **43 days**: May 10 until July 1 (8 days before green light meeting)

3 **Phase 3** (integration) = **30 days**: July 10 until August 8 (final presentation)