Quiet Interaction: Designing an Accessible Home Environment for Deaf and Hard of Hearing (DHH) Individuals through AR, AI, and IoT Technologies

By Wentian Zhu

A thesis exhibition presented to OCAD University in partial fulfillment of

the requirements for the degree of Master of Arts in Digital Futures

OCAD U Waterfront Campus, April 4-6, 2024

Toronto, Ontario, Canada, 2024

Abstract

As technology rapidly evolves, voice-command-based smart assistants are becoming integral to our daily lives. However, this advancement overlooks the needs of the Deaf and Hard of Hearing (DHH) community, creating a technological gap in current systems. To address this technological oversight, this study develops a Mixed-Reality (MR) application that integrates Augmented Reality (AR), Artificial Intelligence (AI), and the Internet of Things (IoT) technologies to fill the gaps in safety, communication, and accessibility for DHH individuals at home. By employing the User-Centric design methodology, this study begins with a needs assessment through a literature review and online survey to understand the unique challenges and preferences of the DHH community. The key contribution of this study lies in its innovative integration of technologies within a Mixed-Reality (MR) framework, with the goal of creating a more inclusive and accessible home environment for the DHH community.

Keywords: Mixed Reality, Augmented Reality, Internet of Things, Artificial Intelligence, Human-Computer Interaction, Accessibility Technology, Deaf and Hard of Hearing Support, Smart Home

Acknowledgments

First, I want to express my deepest gratitude to my primary advisor, Dr. Alexis Morris, and my secondary advisor, Dr. Haru Ji, for their invaluable guidance, patience, and expertise throughout my research. Their mentorship was crucial in shaping this research and bringing it to fruition. I also extend my thanks to all members of the Digital Future faculty and cohort at OCAD University for their support during the development of this project and my final thesis exhibition. Additionally, I am grateful to all the participants who took the online survey; this project would have been impossible without their valuable responses.

I am deeply thankful to my family and my partner for their unwavering support and encouragement throughout this challenging and rewarding journey. Their belief in my abilities has sustained me through times of doubt.

Finally, I want to thank my cat, Xiao Mi, for his companionship. I look forward to sharing many more adventures and creating countless memories together.

Table of Contents

List of Appendices

List of Figures

List of Tables

1. Introduction

1.1 Motivation

My motivation stems from witnessing the challenges my grandparents faced with their hearing as they aged. This experience deeply impacted me, observing how something as natural as aging could create barriers in communication and interaction with the world around them. A particularly concerning moment was when they missed an appliance's alert which might raise safety concerns, such as the pressure cooker's whistle, which nearly resulted in an accident. This incident underscored the dangers and limitations imposed by their hearing impairment. Due to their difficulty in hearing, they began to rely on guessing during conversations, often responding inappropriately due to mishearing questions, further illustrating the extent of their challenges. This personal experience with my grandparents has motivated me to explore and develop solutions that enhance communication and safety for individuals experiencing hearing loss, aiming to mitigate the impacts of such barriers in their daily lives.

1.2 Problem Space

The concept of 'home' extends beyond its physical structure, encompassing aspects of security, personal identity, familial and social relationships, a hub for diverse activities, and a sanctuary from the outside world (Després, 1991; Intille, 2002). Understanding the multifaceted nature of home lays the foundation for understanding its pivotal role in individual lives. The trend towards dynamic home design emerges from integrating Internet of Things (IoT) technology into conventional settings, thus creating 'smart homes' that reflect an advanced approach to enhancing domestic environments (Ricquebourg et al., 2006).

IoT technology, central to the concept of smart homes, fosters an automated home environment by linking various devices and sensors for seamless interconnectivity within an internal network. The goal of integrating IoT technology into our homes is to simplify residents' lives and deliver solutions for comfort and security, thereby enhancing the overall living standards of residents (Cheng & Kunz, 2009; Ricquebourg et al., 2006).

While most home devices perform tasks through physical interaction, the advent and use of IoT technologies have diversified interaction methods people interact with their homes (Vandome, 2018). Smart assistant devices such as Amazon Alexa and Google Home epitomize this technological shift and are becoming widely adopted by the market (Bleakley et al., 2022; Kopytko et al., 2018). These smart assistant devices serve as a central hub, allowing users to link and control various home devices and sensors through voice commands, thereby transforming the way individuals interact with their homes (Dokhnyak & Vysotska, 2021).

However, this shift towards voice-operated technology creates a substantial barrier for people who are Deaf and Hard of Hearing (DHH). For instance, smart speakers and voice-activated home assistants, designed primarily for auditory interactions, are not accessible to those who cannot hear or use voice commands. Similarly, security systems and appliances that rely on sound alerts or verbal instructions further marginalize DHH individuals by not accommodating visual or tactile alternatives. As a result, the reliance on auditory commands in these devices fails to meet the needs of the DHH community, sidelining them from the benefits of such technological advancements (Wojtanowski et al., 2020). While efforts have been made to equip these devices with features to assist DHH users, the focus has primarily been on transcription services, which do not fully embrace the comprehensive vision of smart home living (Amazon.com, n.d.; Google.com, n.d.; Bleakley et al., 2022; Rodolitz et al., 2019). Life within the home encompasses a wide range of interactions and experiences, far beyond the scope of mere transcription services such as speech-to-text.

Navigating the complexities of accessibility within home environments requires an innovative approach, particularly for the DHH community. In this landscape, the convergence of technologies—each distinct in its capability—holds untapped potential. Augmented Reality (AR) introduces a dimension of visual engagement, converting everyday interactions into intuitive and informative experiences (Minaee et al., 2022). Artificial Intelligence (AI) brings its analytical strength, transforming sounds into actionable insights (Chaudhry & Kazim, 2021). Concurrently, the IoT integrates these distinct technologies into a unified network of devices, facilitating a seamless interaction within the smart home ecosystem.

1.3 How Can AR, AI, and IoT Technologies Benefit DHH Individuals in Home Settings?

Acknowledging the special needs of DHH individuals underscores the importance of integrating different technologies in shaping their smart home settings. AR stands out for its transformative potential in the way of conveyance of loT data, by fusing digital information with a user's perception of the physical world (Milman, 2018; Rauschnabel et al., 2019). AR facilitates the overlaying of essential virtual information directly into the physical environment, including visual alerts for doorbells, phone calls, or safety warnings. By integrating virtual elements seamlessly with the real world, AR enhances the accessibility and intuitiveness of user interfaces. It proves especially advantageous for DHH individuals, who depend on visual cues extensively (Almutairi & Al-Megren, 2017). Furthermore, AR's ability to visually represent complex information introduces an innovative possibility for smart home interactions and device control, offering users a new dimension of engaging with their environment (Mahroo et al., 2019; Žilak et al., 2022). By employing virtual control interfaces overlaid with physical appliances, AR simplifies and enhances user interaction, making it more intuitive to manage and understand home systems, and eliminating the need for voice commands (Inomata et al., 2020).

Thus, AR technology is demonstrated as a pivotal tool for delivering information and controlling devices in a direct and easily understandable manner for DHH individuals.

Having highlighted the advantages of AR in aiding DHH individuals, it is equally important to factor in their specific requirements for interpersonal interaction and environmental awareness (Findlater et al., 2019; Jain et al., 2020). The rapid advancement of AI, particularly in machine learning, opens up unparalleled opportunities in creating smart home environments that are intricately responsive, adaptable, and attuned to fit these individuals' special demands (Kopytko et al., 2018). For example, AI-driven models can be trained to recognize and interpret sign language gestures, enabling DHH individuals to control and interact with their home devices using their preferred method of interaction (Inomata et al., 2020). This approach to interpreting gestures into digital commands offers a more flexible and user-friendly smart home system operating environment.

While safety remains a paramount concern within the DHH community (Bragg et al., 2016; Findlater et al., 2019; Jain et al., 2020), AI demonstrates its potential to enhance auditory environment awareness for these individuals (Liu et al., 2017; Nakao & Sugano, 2020; Goodman et al., 2021). Sound recognition models, a key feature of AI, are trained to identify extensive audio events, including safety-related alerts like fire alarms and police sirens, as well as everyday household sounds like doorbells, and phone rings (TensorFlow, n.d.). These sounds are then converted into visual alerts or tactile feedback, ensuring that critical notifications are not missed and maintaining both the convenience and safety of the home environment for DHH individuals.

Given the substantial number of individuals who lose their hearing due to aging or accidents, and are not proficient in sign language (NIH, 2021; Findlater, 2019), AI models, particularly speech recognition models, have proven immensely beneficial (Kafle et al., 2020;

Dokania & Chattaraj, 2021). These models offer alternative means of real-time communication and information access, substantially supporting and addressing DHH individuals' psychological needs, playing a vital role in maintaining their sense of connection and ensuring they remain well-informed (Gao et al., 2020; Heine & Browning, 2009). Incorporating AR with this technique further enhances the experience by visually augmenting transcriptions in the user's field of vision (Peng et al., 2018). This integration offers an immersive and intuitive way for DHH individuals to engage with spoken language, ensuring they stay informed and connected.

IoT acts as the foundation, ensuring devices can collect and share data in real time, allowing for a seamless flow of information within the home ecosystem. For DHH individuals, this means that sensors and smart devices can detect and transmit information about environmental changes and critical information, as well as enable control of various devices, all on a single platform. For instance, a smart door sensor can notify the system of an entry, which then triggers an AR visual alert for the user or is processed by AI models for further action.

By integrating AR, AI, and IoT technologies, the home becomes not only more accessible but also more adaptive to the unique needs of DHH individuals. The continued data exchange facilitated by IoT, which allows centralized device control and monitors the home environment in real time, the capabilities of AR in providing visual cues and intuitive interaction within the environment, and the power of AI in interpreting information, form a holistic framework for developing an accessible and inclusive home environment, catering to both the practical and psychological needs of the DHH community.

1.4 Research Summary

1.4.1 Problem Statement

The vision of 'smart homes' promises to revolutionize living environments by integrating automation to streamline everyday household activities, enhancing convenience and efficiency for residents. However, there's a critical gap in this vision when it comes to inclusivity for DHH individuals. Central to this issue is the prevalence of dependence on auditory commands for device interaction—principally through voice-activated smart speakers and assistants—creating an environment that inherently excludes those with hearing impairments. While alternatives like application-based control systems on smartphones and tablets present a potential workaround, they frequently lack the seamless integration and intuitiveness afforded by voice command systems, thereby compounding accessibility issues rather than resolving them. Moreover, the accessibility solutions proposed by these smart assistant companies, often fail to capture the complex, spontaneous nature of interaction required for a genuinely inclusive smart home experience. This oversight in design philosophy not only marginalizes DHH users but also reflects a broader systemic failure to prioritize universal accessibility in technological innovation. The consequence is a disparity in the smart home experience, underscoring the critical necessity for a paradigm shift towards inclusivity in smart technology development.

1.4.2 Hypothesis

This thesis hypothesizes that the integration of AR, when combined with IoT and AI, through a prototype deployed on Head-Mounted Displays (HMDs) can significantly improve the home living experience of DHH individuals. By creating a network of interconnected devices that enhance control and security, alongside AR's ability to make IoT data visually accessible, and AI's role in analyzing surroundings and providing crucial information. Through this integration, I propose this prototype can help bridge the accessibility gap in smart homes for DHH individuals, offering them an inclusive and accessible living environment.

1.4.3 Research Question

How might the integration of AR, AI, and IoT technologies in HMD platforms potentially open up new avenues for improving the home living experience of DHH individuals?

Sub-Questions

a. How can AR be utilized to visually represent auditory alerts, everyday auditory events like door knocking, phone ringing, or things dropping, and environmental sounds in a domestic setting for DHH individuals, thereby enhancing their interactive experience and situational awareness?
b. In the context of social interactions at home, how might AR visual aids, incorporated with AI interpretation of spoken language create an inclusive communication environment for both DHH individuals and their family members?
c. In what ways can IoT technologies be redesigned or interfaced with AR and AI to provide intuitive non-verbal interaction alternatives, bridging the accessibility gap for DHH individuals in voice-command-centric smart home environments?

1.4.4 Goal and Objective

The core aim of this research is to develop and deploy an AR application on HMDs, specifically designed to address the unique needs of DHH individuals within the home environments. This AR application seeks to:

- ***Enhance Interaction:*** Provide DHH users with an intuitive way to interact with their home environment, thereby mitigating the challenges posed by current technological solutions. This aspect focuses on streamlining access to device control through visual cues.

- ***Improve Social Integration:*** Integrate advanced AI-driven speech services. These functionalities are aimed at fostering inclusivity and psychological well-being by enabling DHH individuals to participate in social interactions and feel a greater sense of belonging more fully.
- ***Intuitive Alerts:*** Offer an intuitive approach to notifications through visual alerts, addressing the gap left by the absence of auditory cues. This is designed to ensure that DHH users receive timely and understandable warnings or reminders, enhancing safety and convenience.

1.4.5 Approach and Methodology

This study begins with a structured literature review divided into two main sections. The first section focuses on identifying the specific needs of DHH individuals through a careful examination of twelve key academic papers. The review then critically evaluates the current assistive technologies, with a particular focus on the intersection of AR, AI, and IoT technologies designed for DHH users. Mixed methodologies were applied to this study, specifically Research through Design (RtD) and User-Centric Design (UCD). An online survey targeting DHH participants yielded 52 valid responses. Insights from this online survey, along with the literature review, have guided the development and refinement of iterative prototypes throughout the research process. The final phase of this study involves a comprehensive reflection on the final prototype's design and functionality, followed by an evaluation to assess its effectiveness.

1.4.6 Contributions

This study advances the field of inclusive solutions for the DHH community in their everyday home life by crafting and applying an innovative combination of AR, AI, and IoT technologies, the key contributions of this research include:

- The development of a prototype that synergizes AR, IoT, and AI technologies to enhance home living experience of DHH individuals. This prototype offers solutions for communication needs, situational awareness, and more intuitive control over home appliances and devices, thereby enhancing their interaction and perception within home environments.
- An exploration of the potential of combining emerging technologies demonstrates how this synergy can elevate life quality for DHH individuals by providing tangible benefits in accessibility and interaction within their living spaces.

1.4.7 Scope & Limitation

Before delving into the details of this research, it's crucial to acknowledge the constraints and limitations that shape its scope. Here are several key limitations that should be considered:

- This study's prototypes were developed based on the findings from questionnaire results and selected literature reviews. The limited pool of participants in the survey and the scope of the literature review may introduce bias.
- Due to the limited resources, this study did not conduct user experiments with DHH participants. Instead, it relied on different approaches for evaluation. This research aims to provide direction for future researchers and studies.

- Privacy concerns, especially those arising from the use of sensitive information (i.e. environmental sound detection, camera data) in AI and IoT technologies, have not been extensively explored within the scope of this research.
- The prototypes central to this study comprise functions for communication, device interaction, and environmental sound recognition functions. The primary aim is to improve the overall home living experience for DHH individuals, rather than solving specific issues related to each function.
- This research operates under the assumption that future technological advancements will mitigate current limitations of HMD headsets, such as high cost and bulkiness.

1.4.8 Chapter Overview

This chapter introduces the disparity between the advancing smart home technology and the overlooked need of DHH users, posing the main research question and seeking to validate the efficacy of integrating AR, AI, and IoT technologies. This chapter also outlines the research goals, anticipated contributions, and the study's scope and limitations.

Chapter 2 presents a structured literature review divided into two main sections. The first section gathers insights into the specific needs of DHH individuals and reviews existing sound awareness systems in home settings. The subsequent section delves into assistive tools that leverage the synergies between AR, AI, and IoT technologies, examining how these technologies can be combined to assist the DHH community. A comparison table summarizes the related work discussed, guiding the design of the prototype.

Chapter 3 discusses the mixed methodologies employed in this study, specifically Research through Design and User-Centric Design, to guide the research process and flow. This section also includes an analysis of the online survey conducted among the DHH community, focusing on their home needs. Insights gained from this analysis further inform the prototype design.

Chapter 4 centers on the iterative development of prototypes, designed to bridge the identified gaps and special needs, and answer the main research question along with its sub-questions. The section details the prototypes' design processes, reflections on each iteration, and a subjective evaluation and comparison of their functionalities and performance. These critical assessments lead to the continuous refinement and finalization of the prototype's design.

Chapter 5 presents the final prototype design alongside its evaluation process. This chapter thoroughly analyzes and discusses outcomes to inform future design iterations, highlighting successes and identifying limitations of the current design.

Chapter 6 concludes this study by reflecting on the initially stated goals and the contributions. The discussion then delves into future work inspired by the evaluation findings and outlines potential directions for this thesis's continued development.

2. Literature Review

The literature review begins by exploring the unique needs and preferences of DHH individuals within the realm of sound awareness. It seeks to understand how the absence of traditional auditory cues in a home environment impacts DHH individuals. The review will then transition to a detailed examination of existing sound awareness systems, assessing their design, functionality, and effectiveness in addressing the specific challenges faced by the DHH community.

Subsequently, this literature review will delve into current solutions that address these special needs, with a particular focus on projects that integrate multiple technologies such as AR, AI, and IoT. By discussing and evaluating these projects, the review intends to pinpoint their shortcomings and contributions, thereby laying the groundwork for the development of my thesis project.

By synthesizing these insights, the literature review establishes the basis for the thesis prototypes, ensuring it is informed by the identified needs and the landscape of existing solutions.

2.1 Special Needs and Preferences of DHH Individuals in Home

Sounds possess a unique characteristic in enabling the transmission of environmental information without the need for direct visual observation (Jain et al., 2019). Therefore, by leveraging this special feature, the sound awareness system shows its potential to offer the DHH community access to the same information available to hearing individuals. The following sections will discuss the special needs and preferences of DHH individuals and investigate multiple sound awareness systems designed to assist them.

In 2019, Findlater et al. conducted a comprehensive survey with 201 DHH participants to understand their preferences regarding design variables in sound awareness technologies. This survey investigated sound types, characteristics, captioning preferences, and favored notification systems. The survey results on preferences of sounds align with the study results from Matthew et al. (2006), Mielke & Brück (2015), Bragg et al. (2016), Liu et al. (2017), Yeung et al. (2020), Nakao & Sugano (2020), and Goodman et al. (2021). These studies reveal that the DHH community has a strong interest in safety-related sounds and voices directed at them, such as someone calling their name. In addition, they are interested in social events, like the presence of others, and non-urgent auditory cues, such as notification sounds indicating the states of appliances.

Among these studies, Matthews et al., and Bragg et al., Yeung et al. further explore the specific sound interests that DHH individuals prefer to receive while they are at home. The results indicate that within the home setting, participants express a strong preference for receiving notifications related to safety such as emergency alarms, social events like door knocking and baby's crying, and appliance sounds.

Table 1 summarizes the key findings from these surveys and studies focused on the sound awareness needs and preferences of DHH individuals within home settings. It highlights the types of sounds they are interested in, the frequency at which these sounds are missed, and the interested sound characteristics.

Aspect	Findings	References
Most Interested Sound Types	1. Safety-related sounds 2. Voice directed to them (e.g., someone calling their name) 3. Social events 4. Non-urgent auditory cues	Bragg et al., 2016 Matthews et al., 2006 Liu et al., 2017 Findlater et al., 2019 Yeung et al., 2020
Most Interested Sounds	1. Appliances states (Normal Functioning, Accidental Malfunctioning, Alerts) 2. Emergency alerts 3. Doorbell and Knocking 4. Phone ringing 5. Intruders 6. People knocking things over 7. Their names	Matthews et al., 2006 Mielke & Brück, 2015 Bragg et al., 2016 Jain et al., 2019 Yeung et al., 2020 Nakao & Sugano, 2020
Interest in Sound Characteristics	1. Identity 2. Confidence 3. Urgency 4. Location 5. Volume 6. Duration	Matthews et al., 2006 Bragg et al., 2016 Findlater et al., 2019 Jain et al., 2019 Yeung et al., 2020

Table 1. Survey and study findings on home acoustic preferences of DHH individuals.

2.2 Assessing Current Sound Awareness Systems in Home Environments for DHH Individuals

Jain et al. (2020) designed a home sound awareness system for the DHH community, which proved its effectiveness by allowing users to review historical sound logs with detailed information like sound locations and classifications. However, their findings underscore the necessity of building personalized notification and filtering systems tailored to each user's specific needs and unique scenarios. This emphasis on customization and user-specific design, as also highlighted by the survey results from Findlater et al. 2019, plays a vital role in managing sound awareness systems effectively. Furthermore, Jain et al.'s research further reinforces the concept that such systems should not only be adaptable and user-centric but also carefully balanced to provide pertinent sound information without causing information overload.

Jain et al.'s and Findlater et al.'s focus on user-centric systems set the solid basis for exploring advanced personalization in sound awareness technologies. Nakao & Sugano's 2020 and Goodman et al.'s 2021 research takes this further by proposing a method where DHH individuals record their own audio samples and upload them to the system by utilizing machine learning (ML), thereby enhancing the system's ability to cater to the user's specific needs. This approach, while innovative, brings forth the challenge of how to accurately capture and interpret real-world sounds for DHH users. This concern, also highlighted in the study by Matthews et al. (2006), is identified as a key obstacle in the development of sound awareness technologies for DHH users.

Besides acknowledging the critical role of sound classification accuracy, Matthews et al.'s study expands on this by exploring the visual representation of sounds. Their study results indicate that DHH individuals prioritize designs that quickly and clearly convey essential sound information through glanceable and easy-to-interpret icons with their sources. Their study provides insights into the impact of different sound types on design choices for various scenarios. For example, in noisy environments, the map design excels in helping users identify and locate sounds, whereas the spectrograph is more effective in quiet settings for displaying sound amplitude and frequency, demonstrating an adapted approach to visual design based on environmental context. Asakura (2023) builds upon this research by combining real-time spectrogram display with ML-based icons on AR glasses, aiming to address occasions when the accuracy of ML models falls short. Although learning to interpret spectrograms requires extra effort from users, the results show that they can combine this outcome with ML detection to achieve the most accurate representation of environmental sounds.

Continuing the discussion on visual representation in sound awareness, Findlater et al.'s survey investigates the device preferences of DHH individuals. Their findings show a notable inclination towards multimodal feedback; a significant majority of participants favored both visual and haptic feedback, often on separate devices like smartwatches for haptic alerts and smartphones or Head-Mounted Displays (HMD) for visual cues. Additionally, the study underscores the need for balancing social acceptance and cultural context in device usability. It also highlights a demand for effective visual communication tools, particularly comprehensive captioning. This preference for comprehensive captioning aligns with the earlier findings of Maiorana-Basas and Pagliaro (2014), which emphasize the crucial role of captioning in providing equal access to information for DHH individuals.

While the preceding discussion in this section has primarily concentrated on the intersecting aspects of these studies, Table 2 provides a more detailed overview of each study. It outlines their main findings and the implications these have for designing sound awareness systems, thereby offering key insights in shaping future developments of sound awareness systems for the DHH community.

Study	Findings	Emphasized Implications for DHH Sound Awareness Systems
Findlater et al. (2019)	1. Priority should be given to urgent alerts and voices 2. Focus on the sound source and direction characteristics	1. Customizable sound filters and notification settings 2. Selection of devices 3. Balancing social acceptance and cultural context for device usability
Jain et al. (2020)	1. Historical sound logs proved effective 2. Enhancing Personal and Home Awareness	1. Importance of customization in notification and filtering systems 2. Adaptable UI Design 3. Problem of Sound misclassification 4. Privacy Considerations
Nakao & Sugano (2020)	1. Non-expert DHH individuals can be trained to understand ML in customizing models for their unique needs	1. Tailoring data collection to DHH users' capabilities in the ML training process 2. Addressing GUI design for easier DHH user interaction 3. Recognizing the diverse sound awareness needs in the DHH community.
Goodman et al. (2021)	1. Personal relevance influenced sound class choices 2. Uncertainty in sound perception affected class definition	1. Personalization of sound recognition for DHH users is key 2. ML tools must accommodate diverse DHH experiences 3. Support needed for nuanced sound classification

Study	Findings	Emphasized Implications for DHH Sound Awareness Systems
Matthews et al. (2006)	1. Preference for History Display 2. Single Icon display valued for event identification 3. Complex in the spectrograph, but recognized its potential	1. Sound identification and user customization are key components 2. Designs aiding sound's identity and user customization 3. Balancing sound details and ease in sound Visualization 4. Adapting Visual Designs to Context
Asakura (2023)	1. ML challenges with distinguishing background noise. 2. Training boosts users' environmental sound recognition (ESR) 3. The combination of Icons and Dynamic Spectrograms aids user recognition	1. Need to improve the AI model for accurate sound differentiation. 2. Importance of user training for effective sound recognition. 3. Combined visual aids improve sound identification
Maiorana-Basas and Pagliaro (2014)	The crucial role of captioning for equal access	Need for comprehensive captioning and visual communication tools

Table 2. Comparative overview of research studies on sound awareness systems for the DHH community.

In summary, the research into sound awareness technologies for DHH individuals underlines the importance of adaptable, user-centric systems while recognizing the diversity of needs within the DHH community. These range from preferences in notification mechanisms, and visual representations to the incorporation of multimodal feedback and comprehensive captioning. This trend towards personalized solutions indicates the importance of technology adept at boosting user autonomy and improving accessibility. Looking ahead, the development of sound awareness systems will be guided by a combination of user feedback, advanced sound processing techniques, and context-aware design. This direction aims to eliminate existing barriers, ultimately enriching the quality of life of DHH individuals and ensuring their equitable access to information.

2.3 Related Work: Current Solutions in Improving DHH Accessibility in AR, AI, and IoT Domains

Recent years have witnessed significant advancements in the domains of AI, IoT, and AR, each offering remarkable potential for developing inclusive accessibility solutions aimed at enhancing the quality of life of differently abled individuals, such as individuals who are DHH (Wald, 2021; Saifan et al., 2018; Iswahyudi et al., 2021). While AI, IoT, and AR have each made significant independent contributions to the field, the exploration of their combined potential is just beginning to emerge. Most existing studies, however, tend to focus on the interactions between only two of these domains at a time, revealing a gap in the research: the comprehensive integration of all three technologies—AI, IoT, and AR—has not yet been fully explored (Alahi et al., 2023; Suzuki et al., 2022; Wang et al., 2024). With this observation in mind, the following sections will discuss and evaluate existing works that focus on dual-domain integration, aiming to pinpoint beneficial elements for a more comprehensive merger of these three technologies, as well as to highlight aspects that require further development.

2.3.1 IoT and AI Integrated Solutions

The term Internet of Things (IoT) refers to an extensive network that includes physical devices, vehicles, appliances, and other tangible entities, which emerged in the late 20th century (Chin et al., 2019; IBM, n.d.). Enabled by the integration of sensors, software, and network connectivity, these objects form a network that is capable of autonomously collecting and sharing data, thereby eliminating the need for human intervention.

On the other hand, Artificial Intelligence (AI) is geared towards simulating human cognitive processes like problem-solving and learning (IBM, n.d.). According to Vaswani et al. (2017) and IBM Data and AI Team (2023), AI deploys predictive models, automation, and classification algorithms to carry out complex tasks typically performed by humans, such as facial and speech recognition, decision-making, etc.

Together, the data collection capabilities of IoT devices can feed AI algorithms, which in turn can make more intelligent decisions that can be executed through the IoT network, shaping the way we interact with the world today (Adi et al., 2020; Shi et al., 2022). The development and integration of IoT and AI have opened up new possibilities for enhancing human life, particularly in the realm of accessibility solutions for differently abled individuals, especially for individuals who are DHH.

Yağanoğlu and Köse (2018) designed a wearable device for individuals with hearing impairments that detects environmental sounds, converts them into vibrations, and allows users to understand their surroundings. The device comprises three key components: a Raspberry Pi, which is a compact single-board computer acting as the central processing unit; a microphone for capturing ambient sounds; and a responsive vibration motor. These components work in harmony, with the Raspberry Pi leveraging machine learning algorithms to process auditory signals from the microphone and translate them into corresponding vibrations through the motor. This system is particularly notable for its low production cost and high performance, as it boasts a general accuracy of 94%, factors that facilitate its easy integration into daily life.

While this work provides a valuable foundation, this prototype still comes with several limitations and constraints. One key constraint is the need for users to be trained to understand and remember the different levels of vibration and their corresponding meanings. Additionally, the device’s effectiveness could be highly impacted by the user’s personal sensitivity to the vibrations, which varies among individuals. External factors, such as ambient noise and simultaneous sounds could potentially interfere with the device's operation and lead to either false or missed detections.

Building upon this work, it is important to note that this model trained by authors only covered eight different sound types, which have a high potential to fall short of encompassing the complexity and diversity of real-world environments. These issues pose significant challenges for independent researchers or those with limited resources, as they may obstruct the development of specialized functions like sound classification across a wide and varied range of real-world scenarios. Given these challenges, this study plans to adapt existing models available on online platforms such as TensorFlow Hub or Hugging Face (TensorFlow, n.d; Hugging Face, n.d.). These platforms offer pre-trained models developed by various research groups, providing a valuable starting point and saving the time and resources required to train a model from scratch.

Nimmalapudi et al. (2022) created a device using machine learning and IoT, powered by Raspberry Pi, to interpret hand gestures and relay messages. The device's camera captures hand gestures, which are then interpreted by a machine learning model trained on five unique gestures. The decoded message is sent via Amazon Web Services (AWS)[1], a cloud platform offering various services such as databases and storage. However, this prototype is currently limited to transmitting emergency messages.

Although the authors demonstrated innovative collaboration of machine learning, IoT, and hardware, it still has a set of limitations. One primary concern is its limited functionality. As stated earlier, this system was solely programmed to send emergency messages, which narrows its range of applicability and limits its use in more diverse communicative scenarios. An

[1] Amazon Web Service (AWS) Introduction: https://aws.amazon.com/about-aws/?nc2=h_header

additional limitation lies in the dataset. The authors utilized their own photographs to train the model which could lead to a bias in the system's gesture recognition accuracy. This potentially reduces its effectiveness when interpreting gestures from different users. Finally, the system's reliance on AWS for message delivery introduces an additional constraint. This dependence introduces two distinct issues: one is the constraint on reliability and availability in areas where AWS services are not accessible; the other is the cost factor, as AWS is not free beyond a certain usage level.

Since the key aspect of this thesis study is the use of Language models for functions such as speech service, the need for accuracy and low latency in real-time conversation necessitates the success of the thesis's outcome. To this end, this study will leverage cloud services, which offer the advantage of high computational power, speed, and stability, ensuring efficient real-time processing. Platforms like AWS, Google Cloud Platform, Microsoft Azure, etc., provide sophisticated speech services, making them viable options for this purpose. Therefore, the study will also explore an optimal platform for the application under development, which is evaluated based on cost-effectiveness, stability, and efficiency, with the goal of improving communication for DHH users.

2.3.2 AR and AI Integrated Solutions

Augmented Reality (AR), a variant of Virtual Reality (VR), distinguishes itself by its relationship with the user's environment (Milgram & Kishino, 1994). While VR fully encapsulates the user in a simulated environment, AR overlays or blends the virtual elements onto the user's real-world view (Ifrim et al., 2021). Thus, in an ideal AR experience, users perceive the coexistence of both virtual and real objects within the same spatial context (Azuma, 1997). Given this unique capability, AR's fusion of virtual and real worlds invites a reimagining of human interactions and our perception of the surrounding environment, enriching visual cues for a more intuitive engagement (Kerdvibulvech, 2019; Wang et al., 2022). Specifically, AR has been beneficial for DHH individuals by enabling real-time, visually guided interactions and communication (Tigwell et al., 2023; Nguyen et al., 2021).

Peng et al. (2018) designed a system that uses the HoloLens head-mounted display (HMD) and utilizes Google Speech API, an AI-based speech recognition model for speech recognition. The system converts the transcribed text received from Google Speech into visual speech bubbles, which are positioned beside the speaker and overlay the real-world environment. This system aims to enhance the participation of individuals with hearing impairments in group discussions, addressing their fundamental needs for inclusion and engagement. The authors also highlight areas for future exploration, such as incorporating emotion detection into the captions and enhancing accessibility for individuals unfamiliar with spoken languages.

Guo et al. (2020) developed an AR-based system HoloSound which features three fundamental sound properties in real-time: *Speech Transcription, Sound Recognition,* and *Sound Localization*. This innovative system assists individuals with hearing impairments by not only transcribing speech but also identifying and locating environmental sounds. Additionally, this system alerts users to important sounds requiring immediate attention, such as phone rings and door knocks.

Although this study presents only an early-stage prototype, the authors propose its potential application in broader social contexts and more complex scenarios, such as group meetings and large lectures, where simultaneous representations of multiple sound sources are crucial. The authors also propose two key considerations for future design. First, they suggest integrating haptic feedback to leverage the heightened tactile sensitivity often found in individuals with hearing impairments (Pieniak et al., 2020). Second, they call for future research to focus on the design of the user interface, highlighting the importance of an information filter system to balance cognitive load with important sound information.

In summary, both projects offer significant contributions to assistive technology for DHH people. Peng et al. (2018) promote social inclusion by making group discussions more accessible. Their intuitive interface and real-world overlays effectively bridge the communication gap. Similarly, Guo et al. (2020) break new ground by not just transcribing speech, but also identifying and locating environmental sounds in real-time. Their systems showcase the potential to be especially useful in various scenarios, from daily interactions to emergency situations.

However, it is important to acknowledge that the research conducted by Peng et al. (2018) and Guo et al. (2020), as well as this study, are based on an important assumption that technological advancements will address the challenges HMD headsets face, such as the bulkiness and high cost, which currently affect the devices' accessibility and usability (Creed et al., 2023; Fang et al., 2023). This assumption is supported by multiple studies (Cipresso et al., 2018; Parekh et al., 2020; Liberatore & Wagner, 2021), which underscore the swift progress in AR technology, indicating a promising future for these devices.

Despite the assumed progress in HMD headsets, there remain areas for improvement. The primary limitation of the current work is its reliance on a text-heavy user interface, without fully exploring the potential of more intuitive visual cues within an augmented environment. Additionally, both works fall short in handling multi-speaker recognition. For instance, Peng et al. (2018) aimed to facilitate group discussions. However, their method of processing each speaker's data separately, centralizing all processed data on a single server, and finally distributing it to the user, can be inefficient in real-world scenarios. These limitations highlight the need for future research to develop more engaging and user-friendly interfaces that fully exploit the potential of AR, as well as more efficient approaches to data processing.

2.3.3. AR and IoT Integrated Solutions

AR, as an emerging technology, has demonstrated its effectiveness in providing visual cues to environmental information (Asakura, 2023; Cosio et al., 2023). Complementing this, a study by Findlater et al. (2019) revealed a strong preference within the DHH community for both visual and haptic feedback. This aligns well with the capabilities of IoT technologies, which show great potential in meeting these specific needs. By integrating various devices into a cohesive network, IoT technologies can deliver a multifaceted user experience that includes both visual and haptic feedback.

Mirzaei et al. (2020) have drawn attention to the growing use of haptic devices in association with VR applications. However, they point out a lack of accessibility features tailored to the DHH community. To fill the void, they developed a system “EarVR”, a cost-effective and compact solution that largely enhances the immersive experiences for DHH individuals. The hardware setup consists of an Arduino board responsible for sound processing and two vibro-motors that provide haptic feedback. This coordinated effort enables EarVR to precisely locate the direction of sound sources within the VR environment and relay this information to the user through haptic feedback on their ears. The test results indicate that this system significantly narrows the experiential gap between DHH individuals and those without hearing impairments. Overall, EarVR stands out for its budget-conscious approach to solving a significant accessibility issue in today's VR technology, specifically for DHH users. Its streamlined architecture integrates seamlessly with multiple types of HMD headsets, making it a versatile solution for enhanced immersive experiences.

Yet, the system still shows room for growth. One notable limitation is the dependency on Arduino's processing capabilities, which might restrict real-time digital audio processing in more complex scenarios. Haptic feedback, a vital interaction medium for this system, could be enriched by leveraging the power of AI, thereby crafting a more context-aware experience tailored to individual preferences. Building on the proven success of haptic feedback in this study, the incorporation of visual feedback could further enhance the system's utility, offering a more immersive experience and broadening its appeal.

While the primary focus of this paper is on the DHH community, it's crucial to recognize the overlapping benefits of accessible design. Ghorbani et al. (2023) offer a comprehensive demonstration of how AR and IoT can improve the life quality of the elderly with cognitive impairment through the Ambient Assisted Living (AAL) system.

The AAL system integrates a variety of sensors and actuators within a home environment, all interconnected through a single network that operates using MQTT[2] — a lightweight network protocol designed specifically for device-to-device messaging. What sets this system apart is its object-oriented decision-making process, which allows for real-time adaptations based on user interactions. Additionally, an integrated AR interface extends this adaptability, serving as the channel for user notifications. As the system processes decisions, it sends messages to the user's AR device in response to specific detected events and interactions.

In addition to these features, Ghorbani et al. (2023) further identified the potential for scalability and customization within the presented model. The system can be expanded to accommodate more sensors, actuators, and scenarios. This adaptability lends itself well to enhancing the independence of individual users. Furthermore, the authors also emphasize the corresponding AR messages to ensure that the system can not only address a wider spectrum of needs but also deliver a personalized and resonant experience tailored to individual users' specific contexts and preferences.

[2] MQTT: https://mqtt.org/

In extending this framework to benefit the DHH community, several adaptations should be considered. First, revisiting the choice of sensors is crucial; the original system focuses on environmental aspects pertinent to the elderly, but adapting to the DHH community necessitates considering sensors that align with their specific requirements. This system's design also reveals its possibility of reshaping interactions with smart devices. While the majority of current smart devices can be activated via voice commands and through a control panel (i.e. smartphone applications), there is a need to explore more convenient and intuitive methods of interaction, especially for the DHH community. This would ensure that technology is inclusive and accessible to all, regardless of their hearing abilities.

2.3.4 Summary of Related Works

This section investigates systems and devices that utilize dual-domain technologies across AR, AI, and IoT, with a specific focus on enhancing DHH individuals' experiences while seeking further inspiration from a wider range of accessibility solutions. This section identifies key contributions and areas for improvement of these works, aiming to synthesize these findings to develop a more holistic accessibility solution for the DHH community within their home living experience.

The table below offers the key technologies used, the contributions, and the limitations of the related works. The contributions of these works are commendable, particularly in their innovative use of dual-domain technologies that advance the daily experiences of DHH individuals. Nevertheless, the limitations reveal a pattern of reliance on text-heavy interfaces and a need for more complex data processing abilities, as well as more comprehensive cognitive support. My research seeks to address these gaps by proposing an integrated approach that combines the strengths of AR, AI, and IoT, specifically tailored to meet the distinct sensory requirements of DHH users. Moreover, it is designed to provide a seamless user experience that reduces cognitive load and promotes intuitive interaction within their living spaces.

Related Work	Categories	Techniques used	Main Contributions	Limitations
Yağanoğlu and Köse (2018)	AI + IoT	Wearable device (Raspberry Pi + Microphone + Vibration Motor)	Compact, cost-effective, high-accuracy, easily integrated into daily life	1. User training required. 2. Unintentionally detects noise. 3. Limited to eight sound types. 4. No simultaneous sound processing
Nimmalapudi et al. (2022)	AI + IoT	Machine learning model on Raspberry Pi with webcam for gesture interpretation	Enables emergency messaging through accurate gesture interpretation for the safety of DHH individuals	1. Limited to emergency messages 2. Gesture recognition bias 3. User training required 4. Reliance on cloud service
Peng et al. (2018)	AR + AI	HoloLens integrates with cloud services to provide live captioning and speaker identification, displaying visual speech bubbles next to the speaker.	Enhance participation of DHH individuals in ground discussion through AR-guided visual communication	1. Reliance on text-heavy user interface 2. Requirement for separate data processing, which is inefficient in real-world applications
Guo et al. (2020)	AR + AI	AR application for speech transcription, sound recognition and localization	1. Cloud service for Speech transcription 2. Trained AI model for sound recognition 3. Respeaker module for sound localization	Text-based user interface, lacking comprehensive cognitive support.

Related Work	Categories	Techniques used	Main Contributions	Limitations
Mizaei et al. (2020)	AR + IoT	Arduino and Vibration-Motors for sound direction haptic feedback in VR environments	Facilitates sound localization with HMDs through haptic feedback, narrowing the gap with hearing individuals	1. Arduino limits complex data processing 2. Needs personalized user experience 3. Lacks more comprehensive cognitive support
Ghorbani et al. (2023)	AR + IoT	Sensors and Actuators, combined with AR for Ambient Assisted Living	Enhances elderly life quality and potentially offers DHH community benefits via scalable, customizable sensor integration	1. Adapt sensors to suit the needs of the DHH community. 2. Calls for more intuitive interaction methods and user interface

Table 3. Overview of related works.

3. Methodology

3.1 Research Methodologies and Thesis Workflow

This study employs a mixed research methodology that includes Research Through Design (RtD) and User-Centric Design (UCD), aiming to address the main research question and its associated sub-questions, and proposing an inclusive solution for the DHH community within domestic settings.

The RtD methodology is integral to an iterative process that yields various outcomes such as knowledge, theoretical validation, and solutions informing future design, representing a dynamic practice-based strategy aimed at solving specific problems (Gaver, 2012; Giaccardi, 2019). Zimmerman et al. (2010) articulate that RtD acknowledges using design practices and processes as a legitimate research methodology. This approach employs design-specific techniques, such as the iterative creation of prototypes and artifacts, to conduct inquiry and exploration (Zimmerman & Forlizzi, 2008).

UCD methodology can also be recognized as an iterative design process, such as utilizing surveys and interviews to understand user needs, with user feedback as its core driving factor (Anderson et al., 1988). This sets a foundation for a design approach that's responsive and grounded in real-world user experiences, ensuring that solutions are directly aligned with user requirements. In this study, the approach begins with a literature review to gain insights into the daily challenges that DHH individuals encounter at home and the advantages and limitations of the current solutions. This is followed by an online survey conducted with DHH individuals aimed at gaining a deeper and detailed understanding of their specific challenges and needs. After gathering insights from both the literature review and the online survey, this study proceeds to the prototype iteration process. These prototypes are then evaluated based on their capacity to improve the home living experience for DHH individuals, focusing on how well they address the specific challenges and needs identified earlier from both the literature review and survey findings.

Building on the context of RtD and UCD, this study also adopts the design thinking process. The design thinking approach emphasizes understanding user needs across several stages: Empathize, Define, Ideate, Prototype, and Test, as outlined by Jain (2015) and Brown (2008). In this study, the development of prototypes is both artifact-led and practice-based. Each prototype is designed to assess the conceptual and design effectiveness of specific parts of my thesis, thereby facilitating ongoing refinement for subsequent prototypes.

By integrating these methodologies, this study's workflow is divided into the following five stages:

- ***Stage 1: Empathize & Define:*** Starting from observing DHH individuals and through the literature review and an online survey to gain a deeper understanding of challenges and existing gaps in the current solutions that they will face at their homes.
- ***Stage 2: Ideate:*** Brainstorming solutions based on the identified challenges and needs of DHH individuals, with a focus on exploring MR's potential to address these problems and enhance their overall home living experience. The ideas from brainstorming lead directly to the design of the prototypes.
- ***Stage 3: Iterate & Prototypes:*** In this stage, prototypes are refined and improved through a self-reflective process, checking if each prototype adequately addresses the problems identified in the literature review and online survey. Comparisons are made to earlier prototypes using a radar chart to visually plot improvements and remaining challenges.
- ***Stage 4: Final Prototype:*** The final stage involves finalizing the AR application for DHH users. Interactive and visual elements are refined and polished to ensure intuitive and effective user interactions. The development process focuses on comprehensively

addressing the identified needs of DHH individuals based on insights gathered throughout the study.

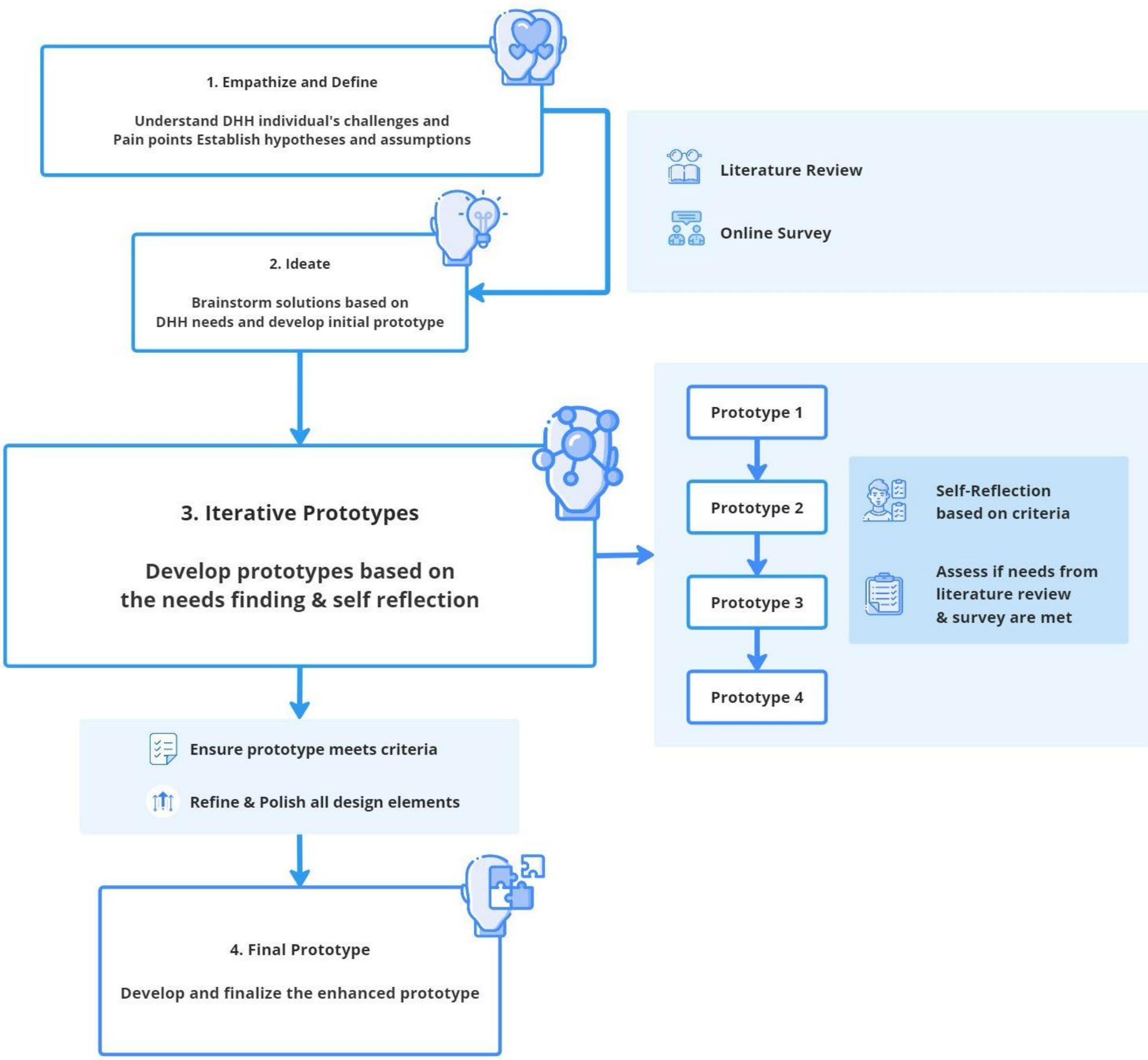


Figure 1. Diagram illustrating the research workflow.

3.2 Online Survey: Daily Challenges and Needs of DHH Individuals

Central to the concept of the UCD methodology, I conducted an online survey to investigate DHH individuals' daily challenges and needs. The semi-structured survey[3] created by Microsoft Form was shared in a WeChat group[4] with 105 members (See Appendix A). To widen participation, posters containing the questionnaire link were also displayed across the OCAD University campus (See Appendix B).

Participation was structured to be both anonymous and voluntary, with the survey targeting participants who were:

- Adults aged 18 or older
- Members of the Deaf or Hard of Hearing (DHH) community
- Proficient in reading and writing for text-based communication
- Willing to share their experiences and challenges

By March 11th, 2024, the survey had collected a total of 54 responses. Among these, 52 responses were considered valid and will be used for the analysis in this study.

3.2.1 Survey Sections and Focus Areas

Considering that discussing daily struggles may provoke mild anxiety, the sole mandatory question in the semi-structured questionnaire simply confirms whether the participant is over 18 years old and inquiries about their hearing level. The questionnaire is divided into four sections:

- ***Inquiry of Hearing Level (Mandatory):*** This section aims to confirm if the participant is over 18 and ask about their hearing level, which is essential for categorizing responses to tailor the analysis more accurately to different needs within the DHH community.

[3] This study has received ethics clearance from the Research Ethics Board of OCAD University # 2024-01
[4] WeChat Group name: Southeast Hearing-impaired Group

- ***Navigating Communication: Challenges and Preferences***: This section aims to understand the participants' communication hurdles, explore their preferred methods and strategies, as well as investigate their psychological needs.
- ***Living with Technology: Accessibility at Home:*** This section investigates the physical barriers participants encounter at home and their use of assistive tools.
- ***Exploring the Potential: Mixed Reality as an Assistive Tool:*** This section seeks to understand participants' assumptions about how assistive functionalities could be implemented in a mixed reality environment.

3.2.2 Online Survey Summary

This section delves into the comprehensive findings from our online survey, segmented into distinct areas to gain insights into the experiences of individuals with DHH. Starting from an inquiry into hearing levels, navigating communication challenges and preferences, to the integration and accessibility of technology at home, and finally exploring the potential of mixed reality tools as assistive technologies. Each segment aims to uncover the nuances of the DHH community's daily experiences and their interaction with technology.

Section 1: Inquiry of Hearing Level

The initial section of the questionnaire collects data on participants' hearing levels, categorized into mild, moderate, severe, and profound impairment. Results indicate that over 75% of respondents fall into the moderate to severe categories (refer to Figure 2). Analysis of their responses reveals that those with more severe impairments tend to use simpler and shorter sentences. This finding underscores the need for the system to be designed intuitively and easily understandable, with a strong focus on visual elements to convey information and ensure a minimal learning curve for DHH users.

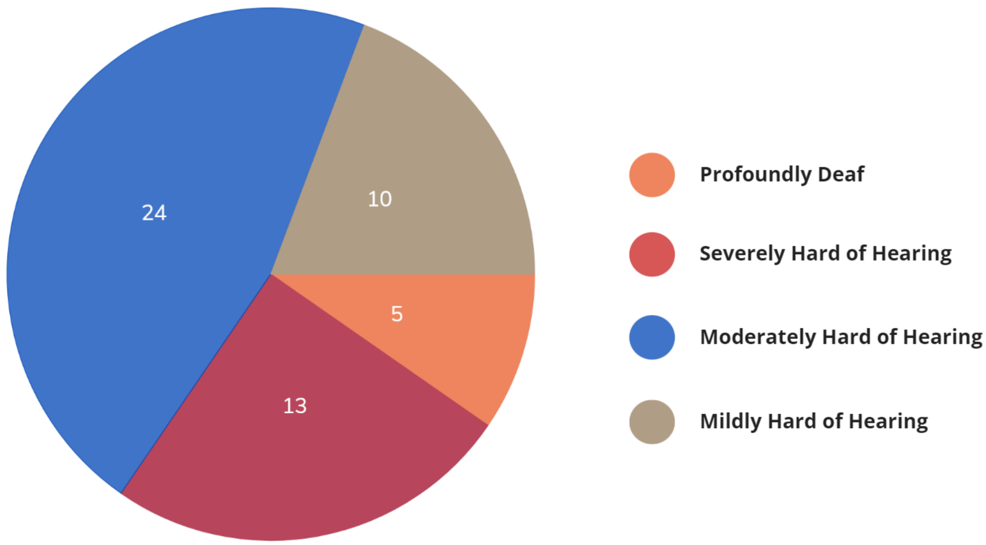


Figure 2. Survey results of participants' hearing level.

Section 2: Navigating Communication: Challenges and Preferences

The survey responses from this section reveal that DHH participants rely on a combination of technology and traditional non-verbal communication methods. Specifically, smartphones are frequently used to type out messages that are then shown to others in person. Sign language remains an essential communication tool, emphasizing the indispensable role of visual aids in facilitating daily interactions.

Miscommunication instances highlight the complexities DHH individuals face in daily interactions, particularly with verbal misunderstandings. For example, participant #17 shared an experience: *'There were many instances, but one time, while I was on the bus, an elderly lady was standing next to me. A man asked me if I could offer my seat to her. I didn't hear him and didn't respond. By the time I realized what was happening, everyone on the bus was staring at me, making me feel extremely embarrassed'.*

According to the data, participants frequently turn to clarifications, text-based communications, and gestures as solutions. Participant #24 described a specific incident: *'While* *walking on a path, I didn't hear a car honking from behind. Realizing this, I pointed to my ears to indicate to the driver that I am hard of hearing, waved my hand, and then wrote 'Sorry' in a notebook to show him.'*

These strategies, while often effective in preventing misunderstandings, occasionally lead to social discomfort and frustration, illustrating the complexity of navigating communication for DHH individuals.

The survey results highlight a significant reliance on digital messaging, such as typing messages on smartphones to show others, to facilitate conversations with those unfamiliar with sign language. Although this method provides a practical alternative, it also presents challenges such as delays and potential misunderstandings during exchanges. This underscores the necessity for developing more seamless communication methods.

Crucially, the emotional toll of hearing impairment on social isolation and psychological well-being is significant. Feelings of loneliness, frustration, and lowered self-esteem are common, reflecting the deep impact of communication barriers on mental health.

Section 3: Living with Technology: Accessibility at Home

In this section, participants describe how technology either mitigates or exacerbates challenges associated with hearing impairments in their living environments.

Many report difficulties in hearing notifications from home appliances or missing critical safety alarms from devices like pressure cookers, electric kettles, and rice cookers. Thirty-two participants shared experiences of consistently missing appliance notifications. For instance, thirty-two participants recount consistently missing appliance notifications. Participant #33 shares, *'Sometimes the fridge door is left open and it makes a sound, but I can't hear it. I only realize it the next time I open the fridge'.* Participant #32 adds, *'I often can't hear the pot boiling, so I need to stay close by to monitor it'.* Additionally, issues extend beyond the kitchen, as participant #37 notes, *'My phone will emit an alert during an earthquake, but sometimes I might not notice it'.* These experiences highlight the need for alternative alert mechanisms, such as visual signals or vibrations, to better meet the needs of the DHH community. In response, some participants have adapted by maintaining visual vigilance on device indicators or relying on family members for assistance.

Regarding smart home devices, like voice-controlled assistants, around 40% of participants indicated that smart home devices, particularly voice-controlled assistants, are less practical due to their dependence on sound for commands. For instance, *“I have a smart speaker, but I rarely use it”*, and another mentioned, *‘I don't have such a device; it's of no use to me’*.

A notable subset of participants who do not own smart speakers, instead rely on non-verbal smart devices like motion sensors to activate lighting, finding these alternatives both useful and accommodating, *‘The hallway in my home is equipped with motion-sensor lights. Although there's a slight delay, it's acceptable and very convenient’*.

Meanwhile, participants with mild hearing impairments describe challenges in using voice-controlled devices, often needing to repeat commands or increase the volume to understand responses. These actions can inadvertently impact family members and even neighbors. Participant #23 notes, *'I have a smart speaker but often can't hear its feedback. Increasing the volume might disturb my neighbors and family, leading to inconvenience'.*

As for the gaps and limitations in assistive technologies, the responses varied but generally pointed towards a desire for improvements in real-time communication aids and more intuitive interfaces for DHH users. Suggestions for enhancements included better speech transcription capabilities, more responsive and less error-prone smart assistants, and enhanced accessibility features that can be integrated seamlessly into everyday technologies.

Section 4: Exploring the Potential: Mixed Reality as an Assistive Tool

In this section, experiences with VR and AR among participants vary. Only a minority have tried VR or AR, praising its visual innovation yet noting a lack of features specifically tailored for the DHH community. They recognize the potential of VR and AR to simulate and enhance visually-based experiences that are otherwise inaccessible. They also suggest that these emerging technologies could revolutionize communication by incorporating features such as real-time sign language translation.

Concerning the role of VR and AR in addressing challenges faced by thoes with hearing impairments at home, participants envision functionalities like visual cues for alerts and subtitles for live conversations, enhancing accessibility and safety. Participant #1 emphasizes the importance of visuals in detecting and pinpointing sounds or activities within the home, *'Mainly visuals that would help notice and target noises or anything going on in the home'*. Participant #32 discusses the advantages of the HMD, noting, *'Wearing a headset might primarily boost my visual experience. While it may offer limited assistance for hearing, it essentially enhances visual utilization'*.

Despite the excitement around these technologies, there's a call for more targeted developments to fully meet DHH individuals' needs, emphasizing the importance of intuitive interfaces and immediate information access.

This section reveals a strong interest in leveraging mixed reality technologies to overcome daily obstacles, highlighting a hopeful future where AR significantly aids the DHH community.

3.3 Needs Finding

3.3.1 Needs Finding from Online Survey

Communication Challenges: DHH individuals rely on a mix of technology and traditional non-verbal communication methods, with a significant emphasis on visual aids. Miscommunications and the emotional toll of hearing impairments highlight the need for improved communication methods that are intuitive and seamless.

Technology and Home Accessibility: There's a necessity for alternative alert mechanisms, like visual signals or vibrations, for home appliances and safety alarms. The limited practicality of voice-controlled smart devices for DHH individuals points towards a need for non-verbal smart devices and interfaces designed with DHH needs in mind.

Interest in Mixed Reality Technologies: Participants acknowledge the potential of VR and AR technologies to simulate visual experiences and revolutionize communication through features. Visual cues for alerts and subtitles for live conversations are among the desired functionalities.

3.3.2 Needs Finding from Literature Review

Preferences for Sound Awareness: DHH individuals have a strong interest in safety-related sounds, voices directed at them (e.g., someone calling their name), social events, and non-urgent auditory cues like notification sounds indicating the states of appliances. This emphasizes the importance of prioritizing these sounds in sound awareness systems.

Customization and Personalization: The effectiveness of sound awareness systems can be significantly enhanced by incorporating user-centric designs that allow for personalization and customization. This could involve settings tailored to individual preferences for specific sounds and notifications.

Visual Representation of Sounds: There's a noted preference for visual cues that quickly and clearly convey essential sound information through glanceable and easy-to-interpret icons. This suggests a need for intuitive visual representation technologies, possibly augmented through AR glasses, that can adapt to different environmental contexts.

Multimodal Feedback: A considerable preference exists for multimodal feedback, indicating that DHH individuals favor receiving information through both visual and haptic feedback. Devices like smartwatches for haptic alerts and smartphones or Head-Mounted Displays for visual cues are particularly preferred.

Comprehensive Captioning: The importance of comprehensive captioning is highlighted, underscoring the need for effective visual communication tools that provide equal access to information for DHH individuals.

3.3.3 Summary of Needs Finding: Insights from Literature Review and Online Survey

Category	Source	Description	Possible Solutions
Preferences for Sound Awareness	Literature Review	Interest in safety sounds, voices, social events, and appliance notifications.	Sound awareness systems prioritize these listed sounds
Customization & Personalization	Literature Review	Preference for tailored sound and notification settings.	Personal sound settings and modes
Visual Representation of Sounds	Literature Review	Preference for visual cues that convey essential sound information through a glance.	Use of AR glasses for intuitive visual representation

Multimodal Feedback	Literature Review	Favor towards receiving information through visual and haptic feedback.	Smartwatches and HMDs
Comprehensive Captioning	Literature Review	Need for equal access to visual communication tools.	Effective captioning tools
Communication Challenges	Online Survey	Reliance on technology and non-verbal communication methods, emphasizing visual aids.	Improved communication methods that are intuitive and seamless.
Advancing Home Accessibility Through Technology	Online Survey	Necessity for alternative alert mechanisms for home appliances and safety alarms.	Visual signals or vibrations for alerts; non-verbal smart devices and interfaces.
Interest in Mixed Reality Technologies	Online Survey	Interest in VR/AR for simulating inaccessible experiences and revolutionizing communication.	Real-time sign language translation; visual cues for alerts and subtitles for live conversations.

Table 4. Summary of needs and preferences for DHH individuals. This table is summarized by the insights from the Literature Review and Online Survey.

3.4 Design Criteria for Evaluating Prototypes

For the comprehensive evaluation of the prototypes, a radar chart (refer to Figure 3) was employed, aiming to contrast the prototypes across several criteria to highlight their strengths and identify areas for enhancement. The criteria for this radar chart were developed by the author based on insights gathered from the needs-finding in both the literature review and online survey.

These criteria include *Visual Cue Efficiency, Interaction Quality, Accessibility Enhancement, Technology Effectiveness, User Engagement, and Usability*, with each criterion evaluated on a scale from 1 to 5. This tailored approach ensures a focused assessment aligned with the prototypes' design and functionality goals.

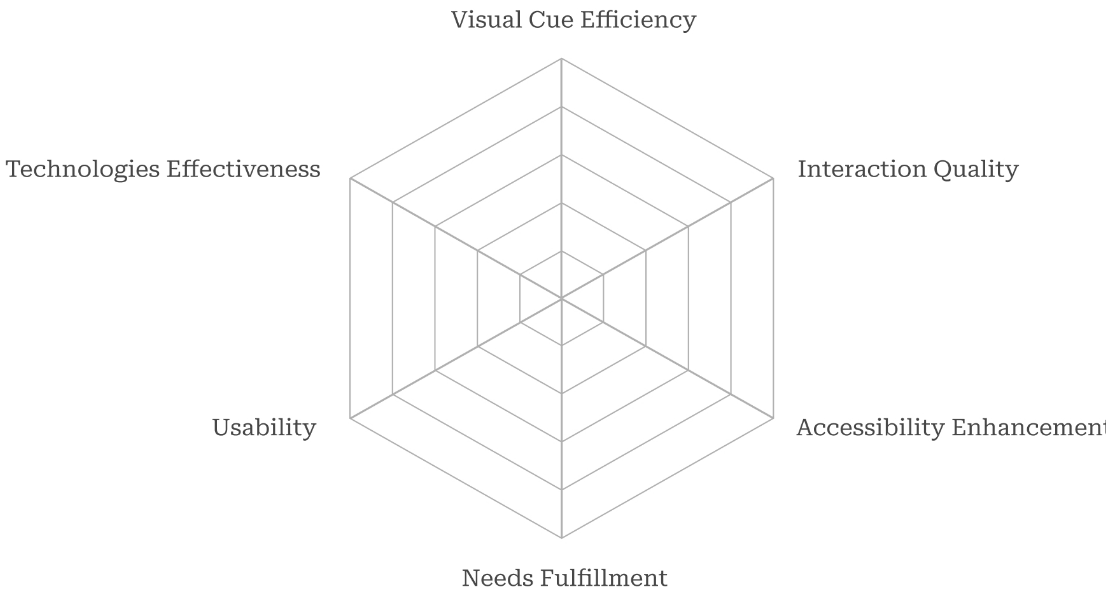


Figure 3. Prototype evaluation radar chart template.

Visual Cue Efficiency: This scale rates the prototype's ability to translate auditory information into visual formats effectively. A score of 1 signifies a minimal visual representation of sounds, while a 5 indicates a comprehensive visual system that intuitively conveys a wide range of auditory alerts and events.

Interaction Quality: This scale assesses the ease and intuitiveness of interacting with the prototype. A rating of 1 represents limited interaction capabilities, and a 5 reflects a highly interactive and user-friendly interface.

Accessibility Enhancement: This scale measures improvements in making environments more accessible to DHH individuals. A score of 1 indicates minor accessibility improvements, whereas a 5 signifies significant enhancements that substantially aid DHH users in understanding their environment.

Needs Fulfillment: This scale rates the prototype's success in addressing the needs of DHH individuals. A score of 1 reflects minimal impact, indicating the prototype meets only a few needs and preferences of DHH individuals. A score of 5 signifies a comprehensive satisfaction of DHH individuals' needs, indicating that the prototype fulfills both the essential and preferred requirements of users in home environments.

Technologies Effectiveness: This scale evaluates how well the prototype integrates AR, AI, and IoT technologies. A score of 1 suggests minimal enhancement of user experience through these technologies, whereas a 5 denotes a significant contribution of each technology, leading to seamless navigation and improved interaction with the surrounding environment.

Usability: This scale measures the overall usability of the prototype, including overall ease of use, learning curve, and user satisfaction. A rating of 1 indicates poor usability, whereas a 5 suggests excellent usability, with the prototype being easy to learn and use.

4. Iterative Prototypes

4.1 Early Stage - Dual-Domain Exploration

The foundational phase of this thesis project begins with the exploration of the combination of dual technology; AI with IoT, AR with IoT, and AR with AI. Each pairing is investigated with the goal of developing 'quiet interaction' solutions for DHH individuals in home settings. The rationale for this approach is to build a solid baseline for each technology pair, ensuring their seamless integration. Such a step-by-step method allows for a thorough understanding of how each integration can contribute to creating a more accessible and inclusive environment for DHH individuals, thereby laying the groundwork for the comprehensive integration of all three technologies.

Prototype 1: Smart Home Control through Sensory Data - AI and IoT Integration

Prototype 1 Overview

Prototype 1[5] explores the integration of AI technologies within the IoT framework, specifically in the context of smart homes that predominantly rely on voice commands. The design of this prototype directly responds to this identified limitation, as highlighted in the literature review and online survey, seeking to offer more inclusive interaction alternatives.

The following table illustrates the categories that Prototype 1 addresses and the methods it uses to meet these identified needs and limitations from previous works.

[5] Prototype 1 Video Link: https://www.youtube.com/watch?v=gj6nBTI1ecE

Category	Needs Finding from Literature Review and Online Survey	How Prototype 1 Addresses the Needs
Advancing Home Accessibility Through Technology	Challenges with **auditory-based** smart home devices due to hearing impairment	Utilizes **hand gestures** to reduce reliance on auditory cues, thereby **improving user autonomy** in device control

Table 5. Prototype 1 Overview. This table details how Prototype 1 addresses DHH users' challenges based on literature and survey insights.

Acknowledging the importance of user autonomy in the context of smart homes with the concept of 'Quiet Interaction' central to this thesis topic, my initial experiment focuses on enabling gesture-based controls for everyday devices, such as light bulbs, to facilitate non-verbal communication with the smart home system. For gesture detection, I utilized the AI framework called MediaPipe [6]by Google. Mediapipe is a machine learning framework, designed for rapid prototyping with sensory data processing, useful for extracting details such as hand landmarks (Lugaresi et al., 2019). To address privacy concerns within the system, I implemented face recognition to ensure that only authorized users can control the devices. For this purpose, OpenCV proved to be an excellent tool for handling this task because of its robust capabilities for facial recognition (OpenCV, n.d.). As an open-source computer vision library, OpenCV is adept at performing tasks such as face recognition and object detection across various platforms without requiring developers to write the code from scratch (Culjak et al.).

[6] Mediapipe: https://developers.google.com/mediapipe

User Flow of Prototype 1

The prototype hardware setup requires a device (e.g. a computer or a smartphone), a webcam, and Philips Hue bulbs. The operational process initiates when the system detects and recognizes the user's face by using OpenCV [7], an open-source library used for computer vision applications. Upon user recognition, the system activates and turns on the light by sending JSON requests to the Philips Hue API. Users can then adjust the bulb's brightness using a hand gesture recognized by MediaPipe. Specifically, the user changes the distance between their index finger and thumb. The system interprets this gesture using MediaPipe's hand landmarks tracking capabilities, translating the distance into a corresponding brightness level. This distance is mapped to a brightness scale of 0-100, where a larger distance corresponds to a higher brightness level and a smaller distance corresponds to a lower brightness level (see Figure 4).

[7] OpenCV: https://opencv.org/

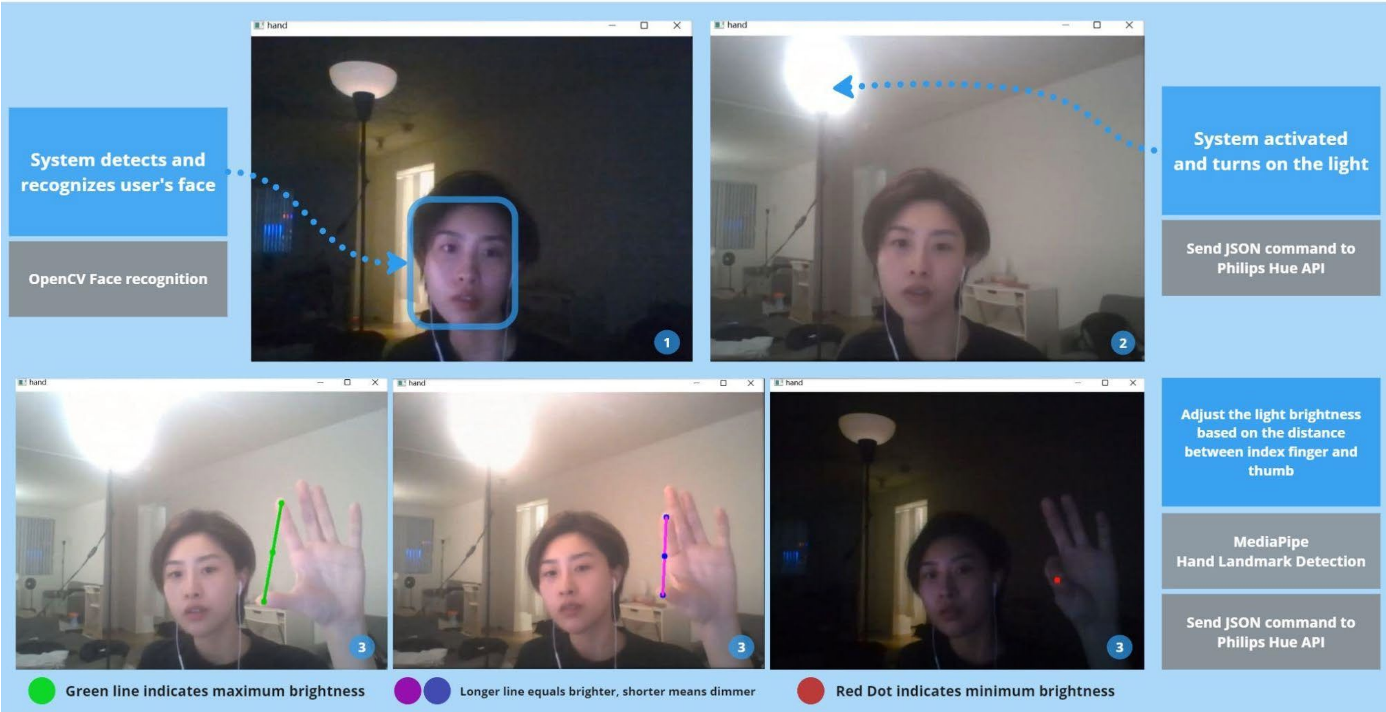


Figure 4. A sequence of four images demonstrating the Prototype 1 system design. The system uses facial and hand recognition to detect the user and adjust the light's brightness accordingly.

Reflection on Prototype 1

While this prototype showcases the success of the integration of AI and IoT in enhancing user interaction through intuitive gesture control, it encounters significant performance challenges. First, the FPS (frames per second) of the streaming video may drop as low as two during the implementation, likely due to factors such as Internet connectivity and the hardware's capabilities. Also, the system's sensitivity to environmental conditions, such as inadequate lighting, may hinder gesture detection and affect its reliability and consistency.

In addition, the fixed position of the camera-based system limits user control of its camera's field of view, leaving the movements outside this range undetected. Furthermore, unintended gestures within the camera's view can accidentally trigger the system, potentially impacting device performance. The adoption of facial recognition technology also raises privacy issues.

Subjective Evaluation of Prototype 1

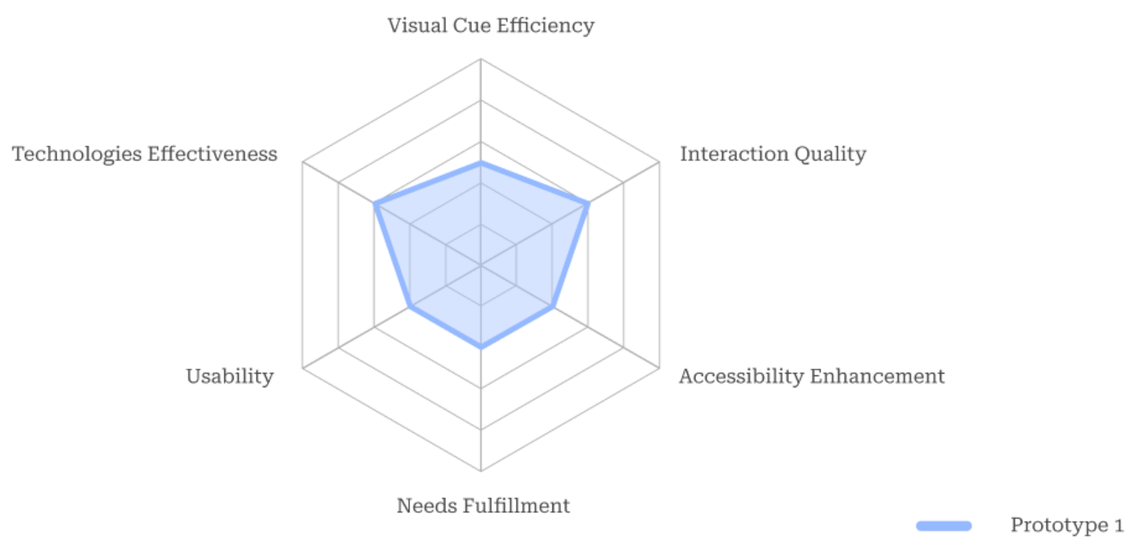


Figure 5. Radar chart for Prototype 1.

Visual Cue Efficiency (2.5/5): Prototype 1 utilizes visual feedback mechanisms by illustrating brightness levels with lines and colors drawn between the index finger and thumb, which demonstrates its capability to communicate information visually, yet within a limited scope. While there's substantial potential for broadening the variety and complexity of visual cues, this method provides a solid foundation for further development.

Interaction Quality (3/5): The gesture-based interaction introduces a user-friendly and intuitive approach, yet performance issues like FPS drops, fixed interaction position, and sensitivity to lighting conditions suggest the interaction quality is moderately effective.

Accessibility Enhancement (2/5): The gesture controls improve smart home accessibility for DHH individuals, with a primary focus on light control. Nevertheless, this focus represents just a fraction of potential interaction within the broader spectrum of home devices. Expanding the range of controllable devices such as doorbells, window blinds, and kitchen appliances would make it more applicable and beneficial for DHH individuals across various living scenarios.

Needs Fulfillment (2/5): This prototype introduces gesture-based control as a non-verbal communication tool for DHH individuals. However, its functionality is limited to interactions with a single type of device, falling short of addressing the broader spectrum of daily living needs.

Usability (2/5): Gesture control in the prototype offers an intuitive means for DHH individuals to adjust bulbs but it falls short in broader applications. The need for learning specific gestures for various functions makes the system less efficient and user-friendly. Simplifying gesture recognition and reducing the learning requirement are essential for enhancing usability.

Technologies Effectiveness (3/5): The integration of AI and IoT shows promise in enhancing smart home interactions. Nonetheless, technical challenges and sensitivity to environmental conditions underscore the necessity for additional optimization.

Prototype 2: Enhancing Interaction with HMD - Integrating AR and IoT

Prototype 2 Overview

Prototype 2[8] represents an advancement over Prototype 1, shifting from a webcam-based setup to the use of an HMD, specifically Meta Quest Headsets, examining the AR's capability in enriching the user experience by overcoming the restrictions of stationary devices, thus offering users greater mobility. The HMD serves as a medium that leads users into an interactive MR environment, which blends the real world with virtual objects, allowing users to interact seamlessly with both their physical surroundings and virtual objects.

[8] Prototype 2 Video Link: https://www.youtube.com/watch?v=EhVfvBKtWLw

The table below outlines how Prototype 2 addresses specific needs and limitations identified in previous studies and the online survey, along with a comparison to Prototype 1.

Category	Needs Finding from Literature Review and Online Survey	How Prototype 2 Addresses the Needs	Comparison with Prototype 1
Advancing Home Accessibility Through Technology	Challenges with auditory-based smart home devices due to hearing impairment	Uses **AR** for mobile, immersive control without auditory cues	Prototype 1 is limited to stationary interactions
Needs for Visual Cues	Dependency on **visual aids** for interaction due to auditory communication barriers	Leverage AR's capability for accessible visual cues and device control	Prototype 1 uses webcam gestures, reducing immersion

Table 6. Prototype 2 Overview.

In the development of the mixed reality experience, I prioritized user understanding and ease of use. To this end, I created detailed user interfaces that provide comprehensive setup instructions (refer to Figure 6). These instructions guide users through the setup process and usage of the system, ensuring they can make the most of the system's capabilities.

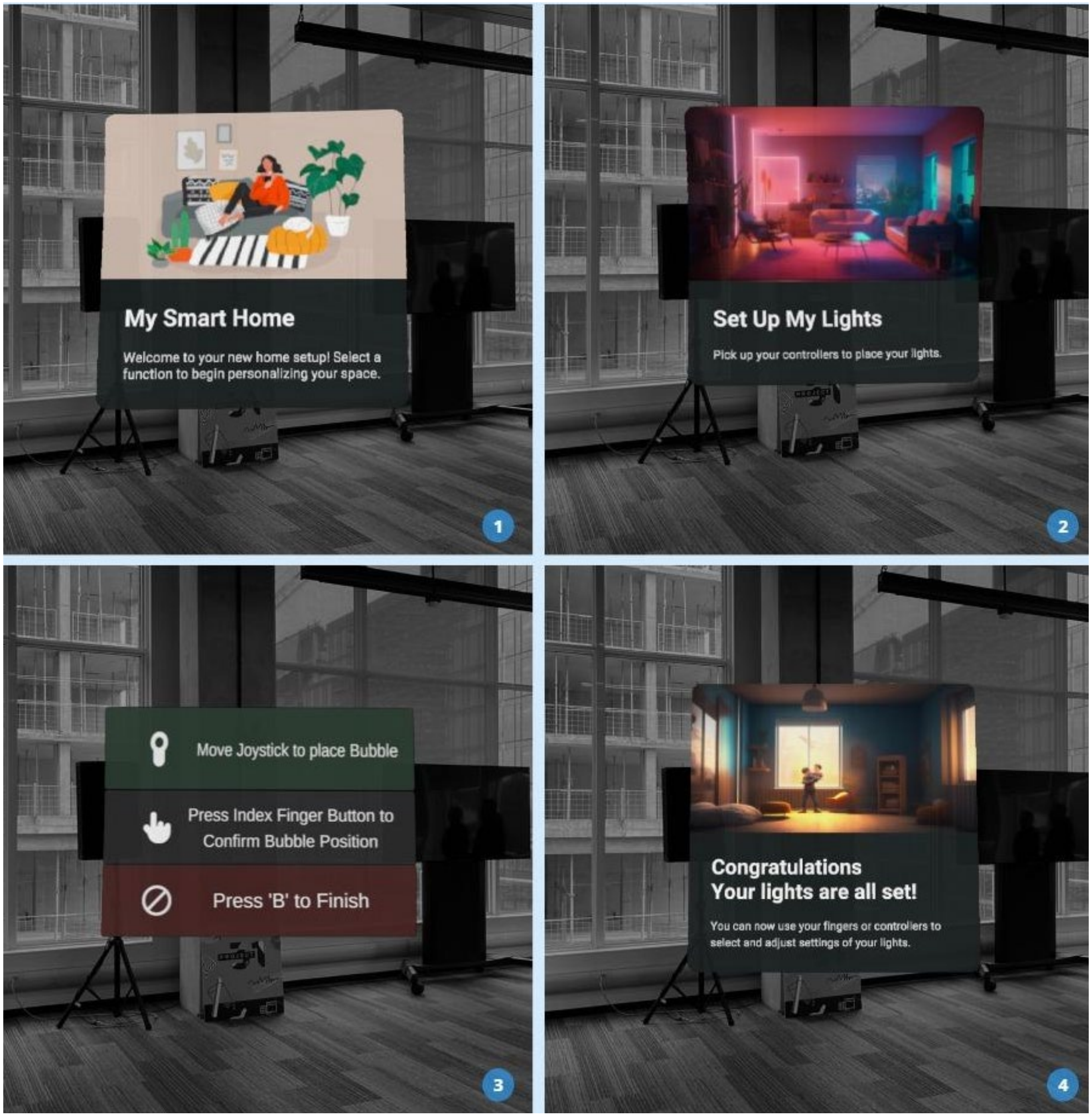


Figure 6. High-fidelity mockups for user interfaces of Prototype 2. These high-fidelity mockups serve as a guide for users during setup and will be implemented in the actual AR application.

To improve the sense of embodiment and enable tangible, intuitive interactions within the mixed reality experience, I developed a direct selection interface. This interface allows users to control a smart bulb, offering functionalities like toggling, adjusting brightness, and changing light colors (refer to Figure 7). Users can interact with this interface by simply "poking" at virtual buttons displayed on the menu. This action is performed by extending a finger and moving it towards a virtual object until it virtually "touches" it, simulating physical interaction. This design choice ensures that users can intuitively manipulate the system's features with minimal effort, enhancing the overall user experience by making virtual interactions feel more natural and instinctual.

Building on the foundational work from Prototype 1, I inherited the gesture control mechanism that leverages the spatial relationship between the index finger and thumb for value adjustment, an approach particularly suited for MR environments, where replicating the sensation of physical touch poses a challenge, as noted by Meta (n.d.). By expanding or pinching their fingers, users intuitively modify values, with the act of pinching to the closest point simulating a 'physical touch' sensation. This simulation enriches the virtual experience, offering a sense of tactile feedback where actual physical contact is absent. Furthermore, the incorporation of a circular slider bar for real-time updates and device value representation not only renders the interaction more intuitive but also offers immediate visual confirmation of changes, which significantly boosts user confidence and engagement with the system, as supported by research from Microsoft (2022) and Apple (2023).

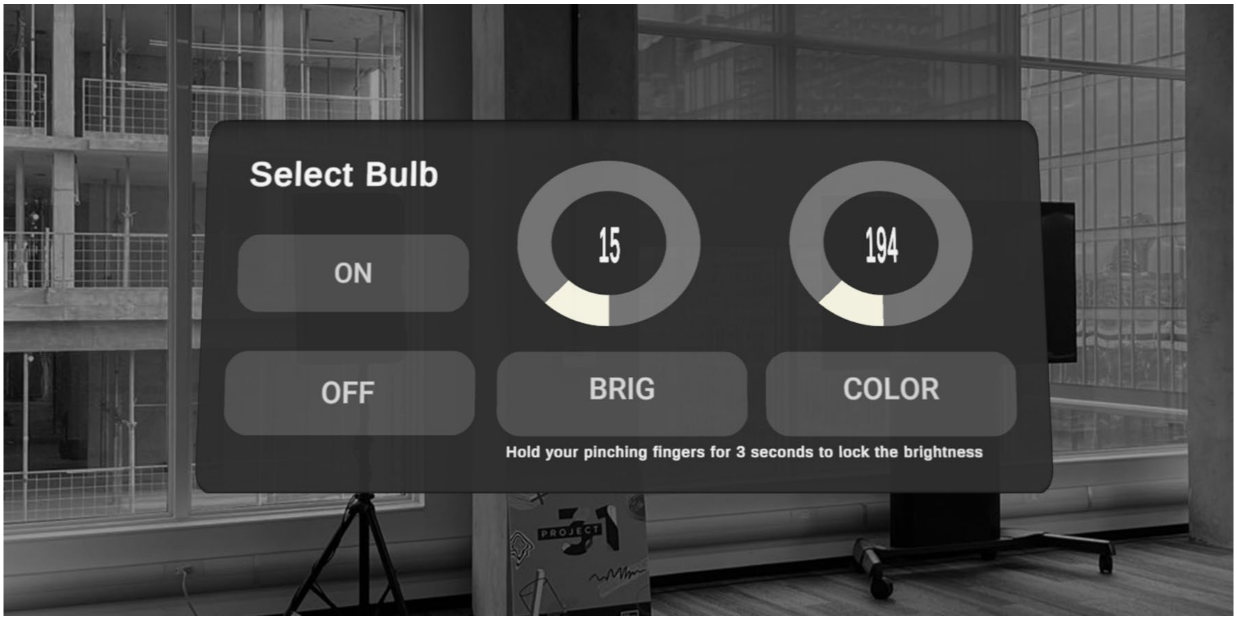


Figure 7. High-fidelity mockups for the device control panel of Prototype 2. These high-fidelity mockups serve as a guide for users during setup and will be implemented in the actual AR application.

Process of Prototype 2

In Prototype 2, I use the Meta XR Interaction SDK [9] to develop the AR application in Unity, specifically designed for development on Meta Quest headsets. Prototype 2 is deployed and assessed on both Quest 2 and Quest Pro headsets, where it functions impeccably without any bugs or conflicts, indicating that this application is compatible with the Meta Quest headset series. Consistent with Prototype 1, I retain the Philips Hue bulb as a representative element and as a preliminary step towards integrating a more extensive selection of smart devices.

A notable challenge with the Meta Quest headsets is their restricted access to raw camera data due to privacy concerns (Meta Quest Blog, 2022). Developers are not permitted to utilize the camera feed, which precludes the use of MediaPipe or any external AI tools requiring camera access. To resolve this limitation, I leverage the headset's built-in hand-tracking capability. By using the Meta SDK API for hand tracking in a Unity script, specifically 'OVRSkeleton', I

[9] Meta XR Interaction SDK: https://assetstore.unity.com/packages/tools/integration/meta-xr-interaction-sdk-264559

identify the bone IDs of the index and thumb fingertips and calculate their distance. This method allows me to convert the measured finger distance into JSON commands, which are then sent to the Philips Hue API to adjust the bulb’s brightness.

User Flow of Prototype 2

Upon the initial launch of the application, users are guided through a setup process to configure their devices by aligning virtual objects to their physical counterparts in the user's environment. While this system prioritizes controller input to ensure precise object placement, it is important to recognize the potential for battery depletion in controllers or the necessity to switch between tasks, such as pickup pens or cups. Thus, the system also offers hand gesture control as a versatile alternative. This ensures an uninterrupted experience, allowing users to seamlessly transition between interaction modes based on their immediate needs or preferences.

Once the setup is complete, the spatial data is stored, negating the need for users to undergo the setup process in future sessions. Users can interact with and control devices using either the controller, by moving the cursor to interact with the menu, or hand gestures, where clicking on the menu and performing gestures to adjust device settings.

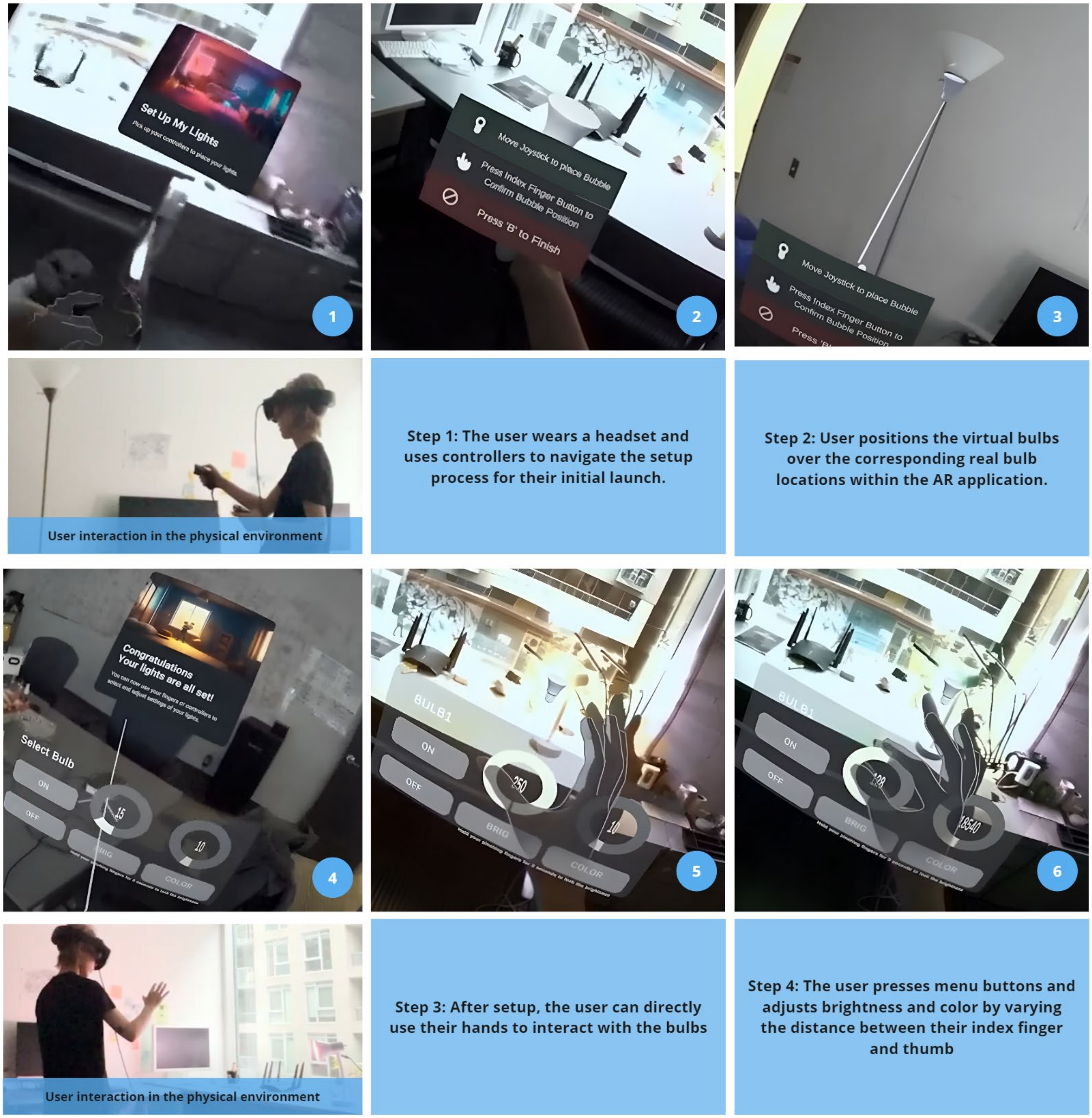


Figure 8. User Journey and Corresponding User Interactions for Prototype 2. Note: The titled screenshots occurred due to an error in the video capture process during application streaming. The actual interface displays correctly oriented, without any tilting.

Reflection on Prototype 2

While this prototype demonstrates the advantages of combining AR and IoT by transcending location limitations and enabling more flexible interaction, it also comes with several limitations.

First, aligning the models' positions with real objects for device control proves challenging when attempting to control devices across different rooms, underscoring the necessity of integrating a floor plane in future iterations. Second, delicate gesture detection's reliability is undermined by latency and accuracy challenges, since hand-tracking demands high levels of ambient light and relies heavily on both hardware performance and hand-tracking algorithms, making precise control a challenge. As a result, accurate finger distance detection becomes difficult, where minor hand movements can lead to unintended and significant adjustments in device settings. Therefore, exploring alternative interaction methods that ensure greater precision but ease of control becomes crucial. Last but not least, this prototype still focuses on bulb control, which, in real-world scenarios, represents just a fraction of the array of home devices. Consequently, the design of this prototype does not present a universal interface design solution, underscoring the necessity to extend experimentation to a broader range of devices to devise a more universally applicable solution for interface design.

Subjective Evaluation of Prototype 2

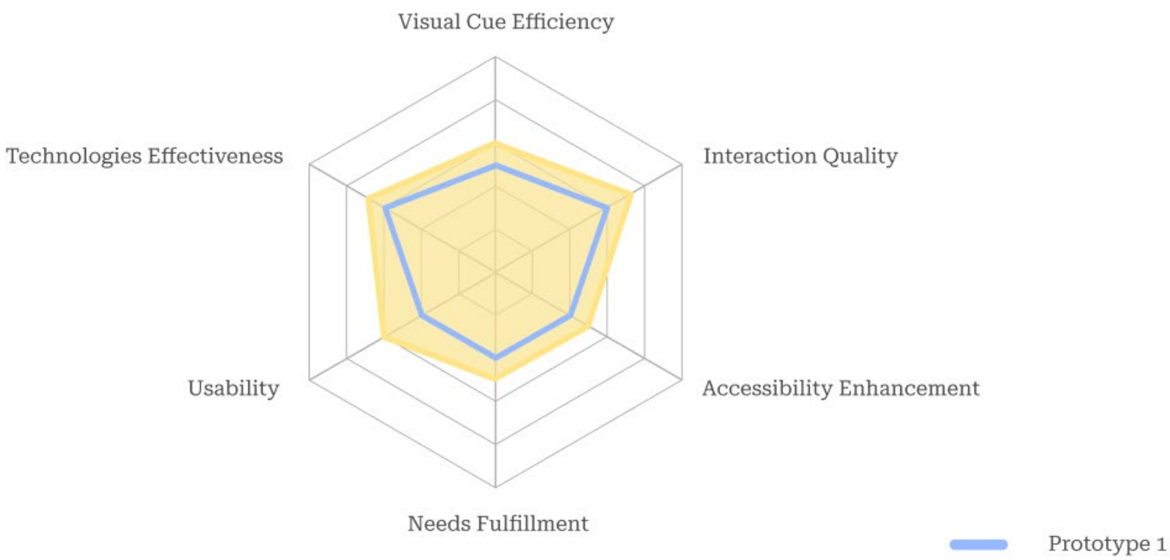


Figure 9. Radar chart for prototype 2.

Visual Cue Efficiency (3/5): This score acknowledges Prototype 2's effective use of the circular slider bars within the AR interface to provide visual feedback, enhancing interaction and user experience by visually representing adjustments. While this feature marks a positive step towards intuitive visual communication, the rating suggests the potential for expanding the range and depth of visual cues to more comprehensively address the needs of DHH users.

Interaction Quality (3.5/5): This score recognizes Prototype 2's ability to transcend spatial limitations through the use of an HMD, significantly improving user mobility and interaction within the home devices. However, it also underscores the need for optimizing interaction quality and more convenient and precise interaction methods.

Accessibility Enhancement (2.5/5): Prototype 2 shows promise in improving accessibility through the use of HMDs by overcoming spatial constraints. However, it lacks clarity on features specifically addressing DHH users' needs, indicating a need for more targeted accessibility features for this group.

Needs Fulfillment (2.5/5): While Prototype 2 advances in mobility and supports non-verbal commands, it lacks clarity in how it caters to the unique needs of DHH users.

Usability (3/5): Although gesture recognition shows limitations, the direct selection menu demonstrates its potential, offering an intuitive method for users to interact with their environment without a learning curve.

Technologies Effectiveness (3.5/5): Prototype 2 shows the effective use of AR and IoT combination, which merges immersive AR visuals with IoT device connectivity, allowing for intuitive control and real-time feedback, thus enhancing the smart home experience. Yet, there's potential for better utilizing AR to offer targeted visual information for DHH users, an aspect also highlighted in the Accessibility Enhancement assessment.

4.2 Stage Two: Integration of AR, AI, and IoT Technologies

Building on the foundational phase, this section of the prototype iteration marks the comprehensive integration of AR, AI, and IoT technologies. Moving beyond the initial scope of facilitating intuitive device control, this phase broadens the system's capabilities to address a wider range of identified needs from the needs finding section. This progression is characterized by a purpose-driven strategy, focusing on creating technological solutions tailored to bridge specific gaps in environmental awareness, safety, communication, and home device management for DHH individuals.

Prototype 3: Augmenting Environmental Perception through the Fusion of AR, AI, and IoT

Prototype 3 Overview

Prototype 1 established a foundation by showcasing the combined potential of IoT and AI to foster intuitive interactions with home devices. Following this, Prototype 2 utilized AR to transcend spatial constraints, thus enhancing device control and broadening the spectrum of user interactions within the IoT framework. Building upon these achievements, Prototype 3[10] seeks to synergize AR, AI, and IoT, concentrating on leveraging these technologies to address specific user needs identified from the need findings gained from both the literature review and online survey, with a particular emphasis on sound awareness.

The following table outlines how Prototype 3 addresses the identified needs and advances beyond the capabilities of Prototypes 1 and 2.

Category	Needs Finding from Literature Review and Online Survey	How Prototype 3 Addresses the Needs	Comparison with Prototypes 1 and 2
Sound Awareness and Environmental Information Perception	Need for enhanced **awareness of environmental** sounds for DHH individuals	**1.** Uses YAMNet sound classification AI model for **sound-to-text alters** in **real-time**. **2.** Employs **AR** to **enhance** DHH individuals' **awareness** of their surroundings	Prototypes 1 and 2 did not include the sound awareness system, as they were more focused on non-verbal interaction

Table 7. Prototype 3 Overview. Comparative advancements of Prototype 3 over Prototype 1 and 2 in addressing the identified needs and limitations of DHH individuals.

The focus of Prototype 3 is on creating a method to interpret sound information to DHH users, thereby enhancing their awareness of their surroundings. Given the complexity of everyday scenarios that the sound detection function needs to handle, I find it most practical to

[10] Prototype 3 Video Link: https://youtu.be/zBOJxOSs4lY

leverage pre-existing AI models from platforms like TensorFlow Hub[11]. These platforms offer a range of sound classification models trained on extensive datasets that align perfectly with my needs as an independent researcher without many resources.

My approach involves identifying sound classification models that are efficient, compact, and resource-friendly, making them suitable for devices with limited computational capabilities, such as Raspberry Pi. This strategy ensures not only the models' performance but also their seamless and cost-effective integration into the everyday lives of DHH users.

Therefore, I have added an additional phase before developing Prototype 3, which involves investigating and testing the capabilities of the sound classification AI models. The insights gained from this testing will guide the decision-making process, helping to select the most suitable model and design for integration into Prototype 3.

AI Model Assessment: Sound Classification with YAMNet Model

In this assessment, the AI model under examination is YAMNet[12], which is available for download from TensorFlow Hub. This model functions as an audio event classifier and has been trained on the AudioSet dataset. It is capable of making individual predictions for 521 audio events, as defined by the AudioSet[13] ontology.

Starting with a Python script on a Windows PC, I utilize the Lite version of the YAMNet model, which is compact and only takes up around 3MB. This design choice aligns with my intention to deploy the model on resource-constrained devices such as the Raspberry Pi, ensuring the system remains lightweight and integrates seamlessly into the DHH user's daily life.

[11] TensorFlow Hub: https://www.tensorflow.org/hub
[12] YAMNet Sound Classification Model: https://www.tensorflow.org/hub/tutorials/yamnet
[13] AudioSet: https://research.google.com/audioset/

The script captures audio from the microphone in real-time using the PyAudio[14] library, classifies environmental sounds with the YAMNet Lite model, and displays the text of the three most probable classifications along with their confidence scores (ranging from 0 to 1, reflecting the likelihood of each sound being correctly identified) through a graphical user interface (GUI) created with customtkinter[15] library, continuously updating these results to the user (refer to Figure 10).

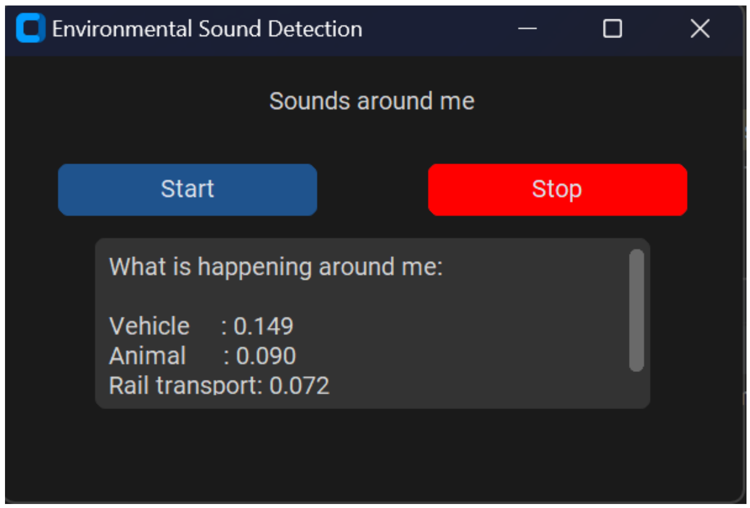


Figure 10. Screenshot of the sound classification application running on Windows PC.

As part of the TensorFlow ecosystem, the pre-trained YAMNet model offers significant savings in terms of developer time and resources, and it's relatively straightforward to integrate into applications. In my testing process, I use a set of audio samples from my home environment, including everyday sounds like door knocks, music, and conversations. By comparing the YAMNet model's predictions with the actual labels of these samples, I calculate the model's accuracy at approximately 85% in my home setting, demonstrating its potential for real-world applications. To assess the model's efficiency on devices with limited resources, I conduct

[14] PyAudio: https://pypi.org/project/PyAudio/
[15] Customtinker Library: https://pypi.org/project/customtkinter/0.3/

evaluations on a Raspberry Pi 4 Model B with 4GB RAM[16], paired with a ReSpeaker 4-Mic Array[17]. This board, which features four microphones, significantly enhances audio capture quality and able to detect sounds within a 3-meter radius (refer to Figure 11).

In Prototype 3, the Raspberry Pi functions as a key intermediary, handling the reception, analysis, and transmission of text-based, lightweight data. This data is then processed by the AR application to generate visual cues. Therefore, I have removed all GUI components to focus solely on outputting prediction results (refer to Figure 12). The application takes approximately 7 seconds to initialize before starting detection. Once operational, it performs in real-time, delivering consistent and stable results. Based on these factors, I suggest that adopting the YAMNet model and the Raspberry Pi presents a viable and practical approach for developing Prototype 3.

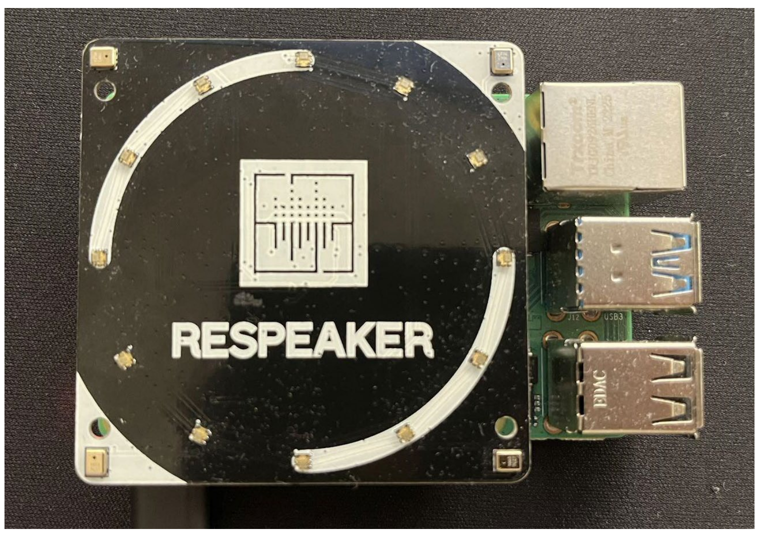


Figure 11. A ReSpeaker 4-Mic Array expansion board attached to a Raspberry Pi.

[16] Raspberry Pi 4 Model B with 4GB RAM: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
[17] ReSpeaker 4-Mic Array for Raspberry Pi: https://www.seeedstudio.com/ReSpeaker-4-Mic-Array-for-Raspberry-Pi.html

```
Speech      : 0.932
Inside, small room: 0.029
Child speech, kid speaking: 0.024

Speech      : 0.766
Ding        : 0.084
Inside, small room: 0.041

Speech      : 0.729
Animal      : 0.066
Inside, small room: 0.040

Speech      : 0.558
Inside, small room: 0.096
Child speech, kid speaking: 0.016
```

Figure 12. The output of the environmental sound detection application is displayed in the Raspberry Pi system terminal.

Hardware Setup and Design of Prototype 3

The hardware setup for Prototype 3 requires a Raspberry Pi 3 or 4 model series powered by a power bank, to which a webcam with a microphone is attached (see Figure 13). Due to the Raspberry Pi's voltage being insufficient to power the servos stably, an additional servo driver module was added (Monk, 2013).

Prototype 3 captures audio using the PyAudio library and utilizes the YAMNet model on a Raspberry Pi for real-time environmental sound detection. Simultaneously, it streams webcam video to a localhost website with the Flask library and MediaPipe. The YAMNet processes these audio data, converting them into textual alerts, identifies the top five sounds and transmits this information to the AR application via User Datagram Protocol (UDP)[18].

On the AR application end, developed in Unity using the Meta XR Interaction SDK, webcam video streamed from a localhost website is accessed and displayed on the UI canvas via

[18] User Datagram Protocol (UDP): https://www.sciencedirect.com/topics/computer-science/user-datagram-protocol

the UnityOculusAndroidVRBrowser[19] package. Simultaneously, the application receives text data over UDP from Raspberry Pi using the same port number.

In the proposed system design, the integration of text alerts into the AR environment enriches the visual experience with auditory information, providing significant benefits for surveillance systems tailored to DHH users. By translating environmental auditory signals into visual representations, the system significantly enhances situational awareness, crucial for recognizing emergency situations.

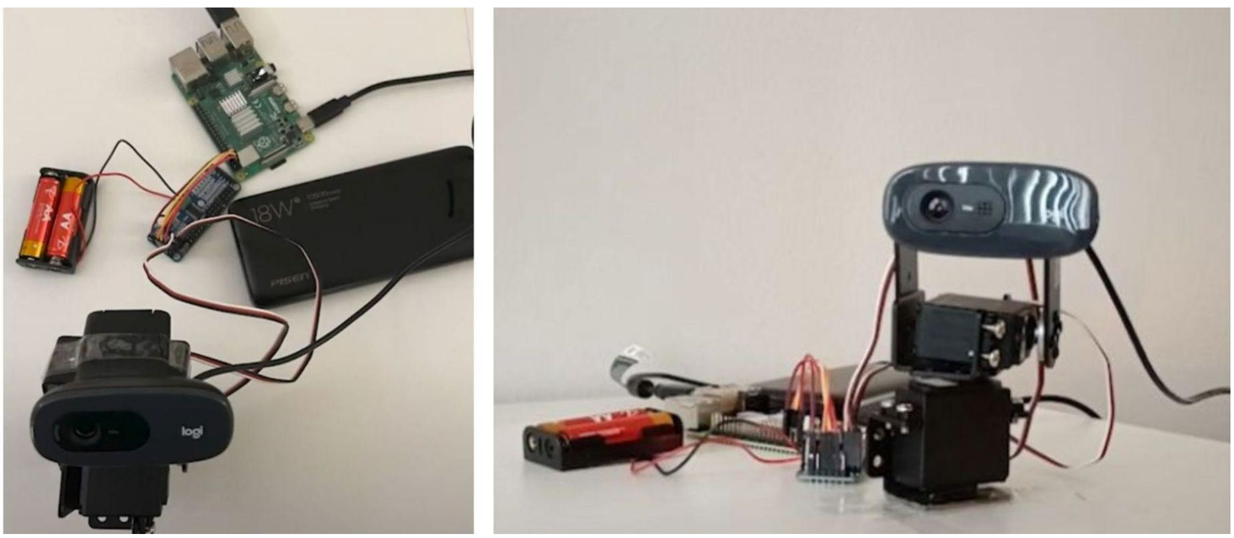


Figure 13. Prototype 3 Hardware Setup. The setup includes a Raspberry Pi Model 4B powered by a power bank. It also features a servo kit, which includes a servo driver that is powered by two AAA batteries and connected to both the Raspberry Pi and two servo motors.

Process of Prototype 3

Prototype 3 converts environmental sounds into text, providing DHH users with continuous access to critical auditory information for enhanced safety and security beyond their direct field of view. Furthermore, by leveraging AR's capabilities to display sound information and relay camera footage on HMD, DHH individuals can remain aware of auditory events

[19] UnityOculusAndroidVRBrowser GitHub Link: https://github.com/IanPhilips/UnityOculusAndroidVRBrowser

through the security camera feed, closely emulating the experience of those with hearing abilities. To further improve user experience and boost their situational awareness and security, I've incorporated direction control buttons into the system. These buttons, accessible via a touch menu, allow users to adjust the camera's orientation as needed (see Figures 14 and 15).

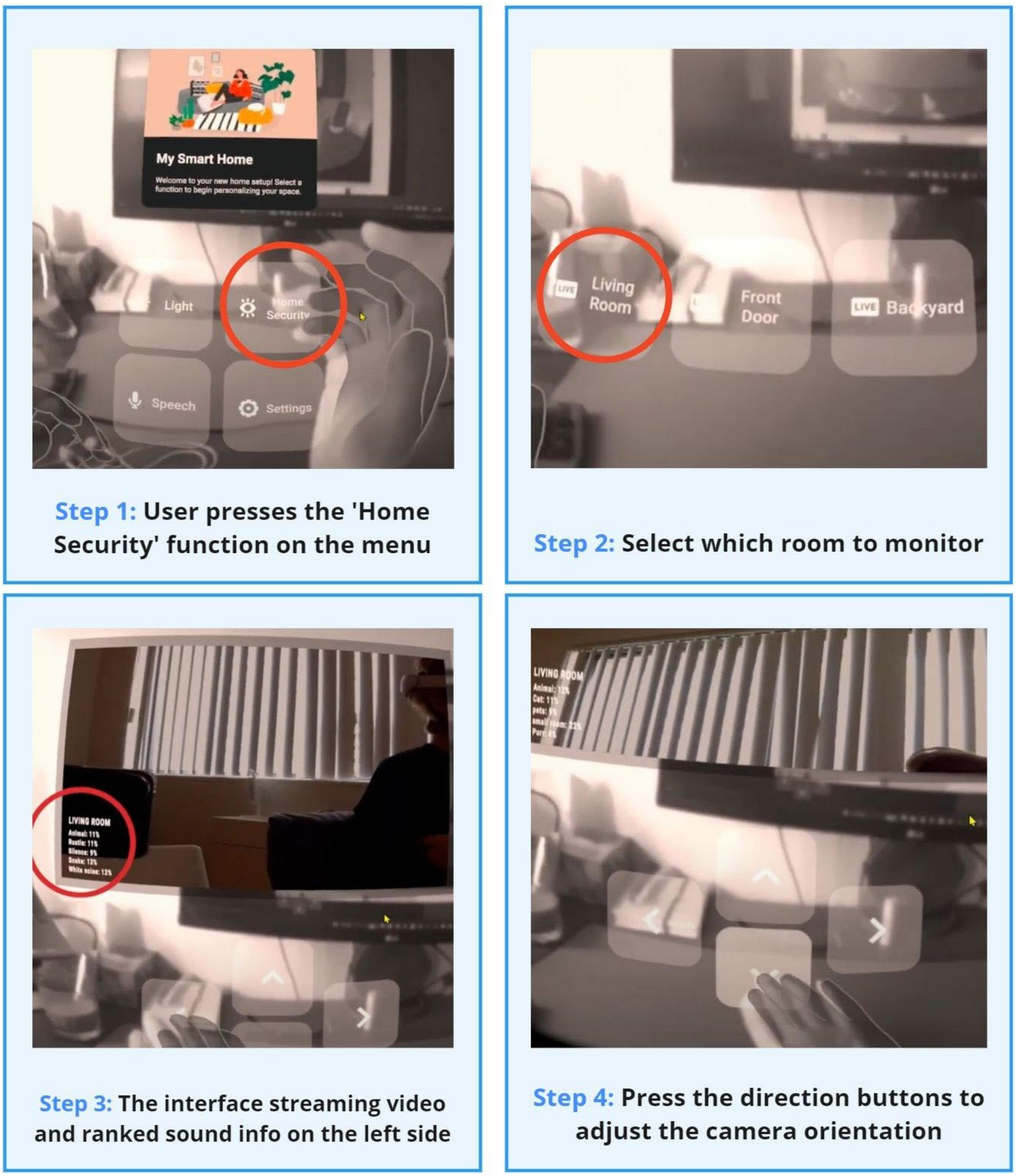


Figure 14. User Journey Map of Prototype 3.

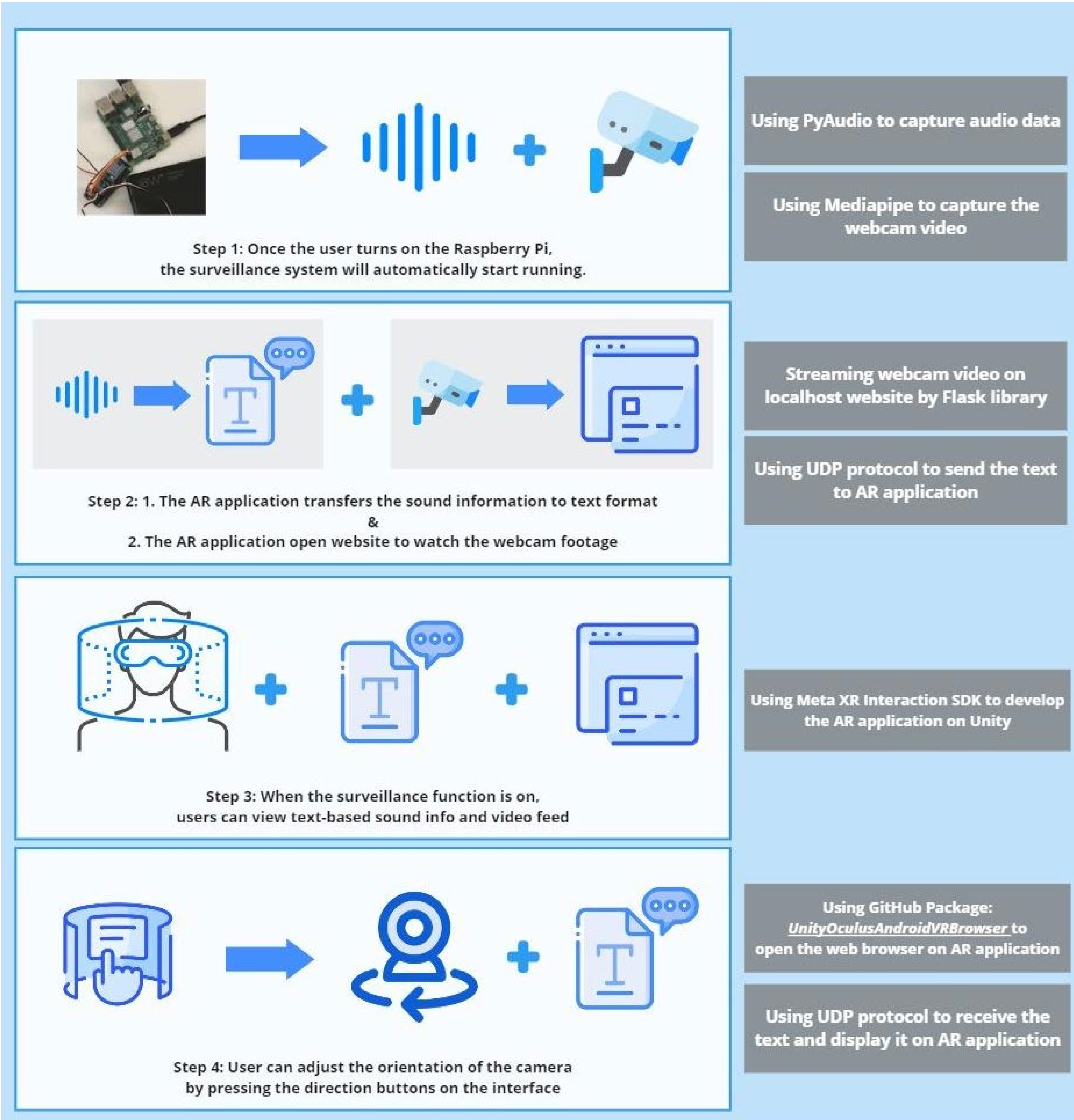


Figure 15. Prototype 3 system workflow.

Reflection on Prototype 3

Prototype 3 represents a significant step forward in this research by effectively combining AR, AI, and IoT technologies. It is primarily designed to augment sound awareness for DHH individuals, tackling a major challenge they encounter in their daily lives. Unlike its predecessor, which focused on gesture recognition, this prototype adopts a straightforward direct selection menu that utilizes hand interactions as the primary method. This change simplifies the user interface, enhancing its usability and ease of learning. Importantly, Prototype 3 transforms environmental sounds into text alerts, addressing a crucial need within the DHH community and showcasing the potential of technology to address real-world challenges effectively.

Nonetheless, this prototype faces several limitations that warrant further development. First, as identified during the needs-finding phase, the representation of sound information predominantly in text form introduces cognitive challenges that need mitigation. Second, while hand interactions offer an approachable and seamless input method, enhancing both self and social presence, they also introduce challenges. For instance, virtual objects fail to provide the physical tactile feedback of real-world interactions, and designing hand gestures that prevent unintended activations is challenging due to the wide variety of hand movements people use in daily life. Furthermore, as Prototype 3 relies on UI interfaces to guide users through different functionalities, there's a pressing need to redesign these interfaces. The goal is to create interfaces that are more intuitive and cognitively engaging, taking full advantage of AR's capability to enhance user interaction and cognitive accessibility through innovative menu designs.

Subjective Evaluation of Prototype 3

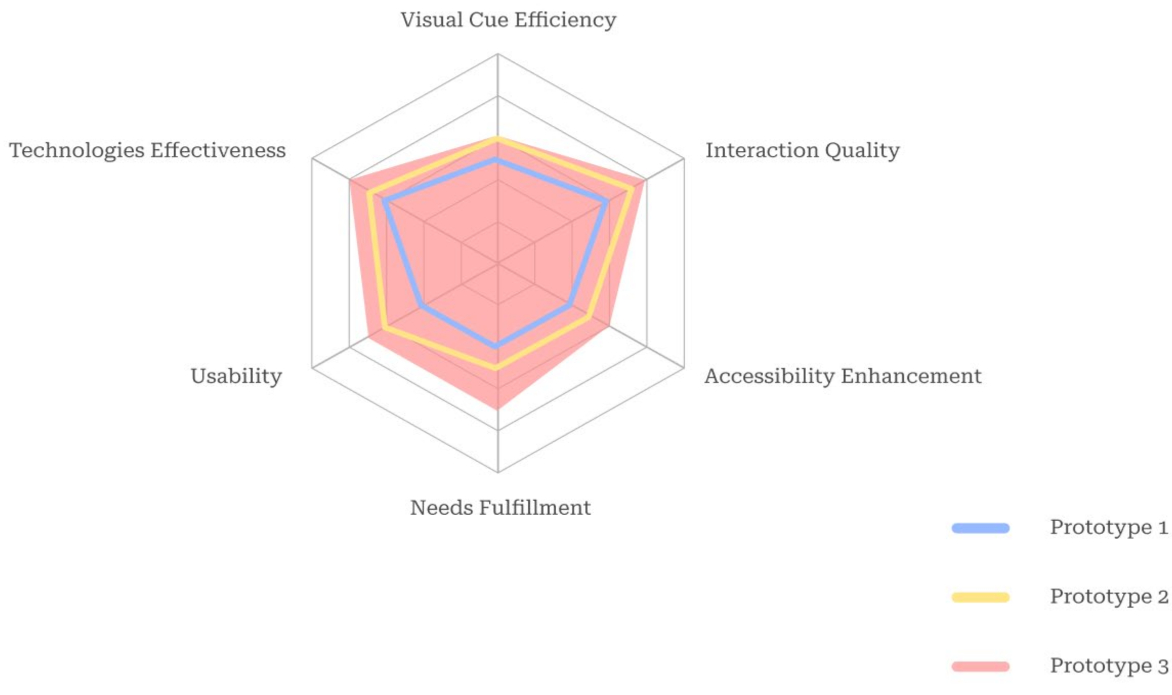


Figure 16. Radar chart for Prototype 3.

Visual Cue Efficiency (3/5): While Prototype 3 transforms environmental sounds into text alerts, enhancing DHH users' awareness, it still relies heavily on textual information. This approach, despite its effectiveness, introduces cognitive challenges by overwhelming users with text, suggesting a need for more diversified and cognitively accessible visual cues.

Interaction Quality (4/5): The shift to a direct selection menu and hand interaction simplifies the user interface, making interactions more intuitive and reducing the learning curve. This development positively impacts the overall interaction quality, though further refinement could enhance the experience.

Accessibility Enhancement (3/5): Prototype 3 enhances accessibility by incorporating a direct selection menu and prioritizing hand interaction. It also empowers users with control over additional device features, such as the orientation of the camera. However, despite these improvements, it still has limitations and cannot be considered a comprehensive solution for addressing the needs of the DHH community.

Needs Fulfillment (3.5/5): With its objective of augmenting sound awareness, Prototype 3 makes a significant contribution to addressing the needs of the DHH community. However, it also points to the importance of addressing a broader spectrum of needs.

Usability (3.5/5): Prototype 3 demonstrates improved usability with its user-friendly interface and simplified interaction mechanisms. However, as Prototype 3 primarily concentrates on the sound awareness system, the usability of other interaction methods and broader home interactive features remains under investigation.

Technologies Effectiveness (4/5): Prototype 3 effectively combines AR, AI, and IoT, demonstrating the potential of these technologies in addressing challenges faced by DHH individuals. AI enables efficient and real-time information interpretation, facilitated by IoT and displayed via an AR HMD application. However, further optimization of these technologies to reduce cognitive load and enhance user engagement is a crucial next step.

Prototype 4: Enhanced Social Inclusive and Engagement - Integration of AR, AI, and IoT

Prototype 4 Overview

While Prototype 3 focuses on addressing the challenges of sound awareness for the DHH community through exploring environmental sound classification models, the goal for Prototype 4 is to determine the optimal pre-trained automatic speech recognition (ASR) model for live captioning, a crucial feature for its integration into the AR application to meet the communication needs of the DHH community. ASR technology, which transforms spoken language into written text using complex algorithms and machine learning techniques (Alharbi et al., 2021), led to the next phase of AI model evaluation. This phase was specifically aimed at assessing the performance of ASR models to select the most suitable one for Prototype 4's design.

The table below outlines how Prototype 4 addresses specific needs and limitations identified in prior studies, comparing its advancements to those of previous prototypes.

Category	Needs Finding from Literature Review and Online Survey	How Prototype 4 Addresses the Needs	Comparison with Previous Prototypes
Visual Representation of Sounds	Preference for visual cues to convey sound information	Offers **text-based** visual alters via AR for **both sounds and speech**	Advances the visual aspects compared to earlier prototypes
Communication Barriers	Need for equal access to visual communication tools	Uses **AI models and AR** to **transcribe** and **display live captions**, giving DHH users equal access to communication and information	Prototype 4 introduces real-time captioning, which was not addressed in previous prototypes

Table 8. Prototype 4 Overview.

AI Model Assessment: Compassion of Speech Services

During my initial tests, I tried to connect an external microphone to the Meta Quest 3 headset using a USB-C adapter to assess the input audio quality. Unfortunately, the headset does not support external microphones, resulting in poor audio input quality. To overcome this, I utilize a Raspberry Pi 4B with a ReSpeaker 4-Mic expansion board for audio capture. Given the large size of the language model and the limited computational resources of the Raspberry Pi, I explore the option of utilizing speech services through an Application Programming Interface

(API)[20], a set of protocols that enable different software applications to interact and share data, offered by Microsoft, Google, and OpenAI. My evaluation criteria for these services include their accuracy, cost-effectiveness, efficiency, and connection stability.

Experiment 1: Testing Whisper-Tiny ASR model

The first experiment uses the Whisper-Tiny model via the Hugging Face Inference API[21]. Hugging Face[22] is a machine learning (ML) community that assists users in building, deploying, and training ML models, while Whisper[23] is an ASR model that can both do transcription and translation, it can transcription multiple languages and translate them into English. The Whisper-Tiny model used in this case is notable for its small size, which only takes 39MB, making it ideal for standalone devices like Raspberry Pi and HMD. I implemented two functions to test its capabilities: one calls the Whisper-Tiny model via Hugging Face's Inference API, and the other downloads the model directly to the project file to assess its performance.

Testing Whisper-Tiny Model via Hugging Face API

In this experiment, Unity version 2022.3.0 was utilized. A package is downloaded from the Hugging Face's Unity API[24] repository on GitHub and imported into a Unity project. Upon successful installation, users need to input their Hugging Face API token directly into the Unity inspector. Additionally, they can copy the link of the desired model from the Hugging Face website and paste it into the inspector panel, specifically in the Task Endpoint section. In this instance, the model of choice was Whisper-Tiny[25].

---

[20] Application Programming Interface (API): https://www.ibm.com/topics/api
[21] Hugging Face Inference API: https://huggingface.co/docs/hub/en/models-inference
[22] Hugging Face: https://huggingface.co/
[23] Whisper ASR Model by OpenAI: https://github.com/openai/whisper
[24] HuggingFace-Unity Package: https://github.com/huggingface/unity-api
[25] Whisper-Tiny Model on Hugging Face: https://huggingface.co/openai/whisper-tiny

In this experiment testing, the Whisper-Tiny model demonstrates basic English speech recognition capabilities, although the results vary. The results are easily affected by accent and the microphone quality, I attempted to slow down my speech and enunciate clearly. However, the model's accuracy in real-world implementation was quite low. For example, when I said, *'I can see a cat is making dinner in a magical kitchen, where full of mushrooms and flowers,'* the model detected it as, *'I can see a cat is the mechanism in a magical kitchen, where full of mushrooms and flowers.'*.

In addition, the frequent occurrence of HTTP Error 503 (Service Unavailable) messages from the Hugging Face API protocol suggests issues with the calling frequency and connection stability. Therefore, the method of utilizing Whisper-Tiny through the Hugging Face API might not be feasible for practical deployment.

Built-in Whisper-Tiny Model in Unity Project

Given the compact size of the Whisper-Tiny model, suitable for running on a Raspberry Pi and HMD, I decided to further explore its capabilities. To eliminate the need for dependency on an Internet connection, I download the Whisper-Tiny model from the Whisper GitHub repository [26] and import both the model and the Whipser-Unity package[27] into Unity.

The results show better accuracy in detecting my words, achieving one hundred percent accuracy, compared to when using Hugging Face's API. However, when I increase the speed of my speech to mimic real-world speech pace, the accuracy rapidly declines. I also test the model's translation feature by speaking Chinese with the same content. The translation results were far from what I was saying. Notably, the processing speed, displayed at the bottom left corner,

[26] Whisper Github repository: https://github.com/openai/whisper
[27] Whipser-Unity Package: https://github.com/Macoron/whisper.unity

increased but remained within an acceptable range, taking approximately 2.7 seconds for Chinese, compared to approximately 1.6 seconds for English.

Although the built-in method demonstrates stability when the model is directly imported into a project, the high inaccuracy rate leads to failure in real-world implementation. Real-world communication scenarios present a complex landscape, with individuals showcasing a variety of accents, speech tones, and speeds. In these situations, the Whisper-Tiny model may not deliver optimal performance. Therefore, future experiments should focus on identifying more sophisticated models that are better equipped to handle the intricacies of real-world cases.

Experiment 2: Testing Whisper-Large V2 ASR Model via OpenAI API

The second experiment utilizes the API provided by OpenAI, which offers access to the Whisper-Large V2 model—an advanced version known for its improved accuracy in transcription tasks. This method is efficient and straightforward, requiring only a few lines of code to implement the API (variations may occur depending on the programming language used).

The results generally exhibit high accuracy in both transcription and translation. However, there are rare instances where the translation function fails, resulting in a blank display on the panel. Additionally, it is important to consider the costs[28] associated with using OpenAI's API service.

A crucial aspect of the Whisper model as of March 11th, 2024, is its lack of support for real-time transcription. Audio inputs or clips must be segmented into several seconds and sent to the service for analysis, which can disrupt sentence continuity and confuse users about the context. Thus, a real-time speech service will be necessary for the next experiment.

[28] OpenAI API price: https://openai.com/pricing

Experiment 3: Real-time Speech-to-Text Transcription via Different APIs

In the real-time speech detection experiment, APIs provided by Google Cloud[29], Assembly AI[30], Microsoft Azure[31], and Amazon AWS[32] demonstrated high accuracy, low latency, and stable connections. These services are user-friendly and can be easily integrated by downloading their library from the website and only require basic code modifications. However, it's important to note that these services are not free, which might limit accessibility for DHH individuals on a budget.

Summary of ASR Model Experiments

The evaluation of the ASR model begins with the Whisper-Tiny model, known for its compactness and capability in transcription and translation, compatible with Raspberry Pi and HMD headsets. Despite its convenience, the model's accuracy in transcription and translation, when accessed through the HuggingFace API, was found to be insufficient for practical applications. Although integrating the model directly into the project and running it on a standalone device increased stability, it did not sufficiently overcome limitations such as accent recognition and speech speed for real-world use.

Subsequently, I tested the Whisper Large-V2 model via API, which shows remarkable accuracy and low latency. However, it struggles with real-time transcription requirements, which leads me to investigate real-time speech services provided by Assembly AI, Google Cloud, Microsoft Azure, and AWS. These services demonstrated high accuracy, low latency, and stable connections, presenting viable alternatives. Nevertheless, their high cost poses a significant

---

[29] Google Cloud Speech Service: https://cloud.google.com/speech-to-text/docs/transcribe-streaming-audio
[30] AssemlyAI: https://www.assemblyai.com/docs/speech-to-text/real-time
[31] Microsoft Azure Speech Service: https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text
[32] AWS Speech Service: https://aws.amazon.com/transcribe/

challenge, suggesting that future development should consider financial implications when incorporating these services.

Prototype 4 Development

Core Functionalities of Prototype 4

Prototype 4 centers on two key objectives: developing communication solutions for DHH users and advancing the user interaction design of the AR application's user interface to fully leverage AR capabilities.

Recognizing the complexity of daily communication scenarios, which include mutual conversation needs and participation in group discussions, it is vital to ensure that the designed features support real-time operation. This allows for immediate response and interaction and facilitates speaker identification. Therefore, I revisit the speech services that I previously tested, specifically seeking their real-time speech features. Microsoft Azure stands out for its comprehensive features, such as real-time speech-to-text, text-to-speech, real-time translation, and notably, real-time speaker identification, all with a low and straightforward setup requirement, making it the superior choice for diverse communication needs.

User Interface Design of Prototype 4

Initially, Prototype 4 features a traditional 2D text-led menu, adhering to conventional text-based 2D user interface practices (refer to Figure 17).

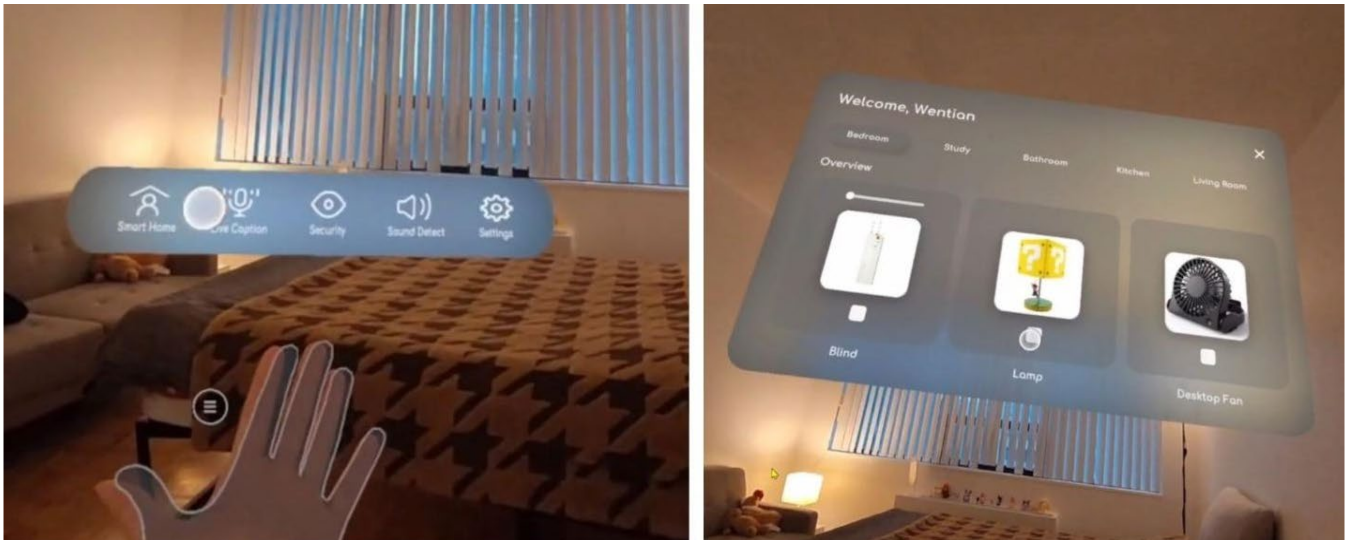


Figure 17. Initial 2D interface design for smart home control. This figure illustrates the initial traditional 2D text-led menu for smart home device interaction, which was explored during the early design stages of Prototype 4.

However, reflecting on insights from previous prototypes, I shift towards a more innovative approach by integrating a 3D model-based interface. The employment of 3D models [33][34] as the primary visual element, supported by textual descriptions, crafts an interface that is both more intuitive and engaging. This shift acknowledges the static nature and interaction limitations of 2D menus. In contrast, a 3D interface fosters exploration and interaction within a virtual environment, aligning more closely with natural human perception.

Notably, these 3D models incorporate motion design; when users hover over an option and remain, the model animates, increasing in size to indicate the current selection and popping up the text description of the chosen functionality. The combination of these design elements enhances user feedback and interaction, making the selection process clear and intuitive. Evolving from a 2D text-based user interface to a 3D-designed user interface with dynamic

[33] House model: https://assetstore.unity.com/packages/3d/props/pandazole-simple-game-low-poly-pack-210274
[34] Characters model: https://assetstore.unity.com/packages/3d/characters/city-people-lite-260446

motion cues, this design choice reflects my dedication to enhancing user interaction and intuitiveness by leveraging AR's potential (refer to Figure 18).

Figure 18. 3D user interface design of Prototype 4.

Hardware Setup and System Workflow of Prototype 4

To optimally capture audio, the Quest 3 headset is equipped with a Raspberry Pi, to which a ReSpeaker Mic Array is attached as an expansion board (see Figure 19). The Raspberry P is powered by a power bank connected with a 1.5-meter cable, facilitating user mobility by allowing the power source to be placed in a pocket. Additionally, this application eliminates the need for controllers, enabling users to directly wear the headset and interact with the system solely with their hands.

Figure 19. Prototype 4 hardware setup. A user wearing a Meta Quest 3 with a Raspberry Pi Model 4B mounted on it (left). The hardware setup for Prototype 4 (right): A Raspberry Pi Model 4B with a ReSpeaker Mic Array, powered by a power bank connected via a 1.5-meter cable (right).

The system workflow begins when a user puts on the Meta Quest 3 headset, activating the AR application (refer to Figure 20). Once the 'Live Caption' feature is selected from the menu, the Raspberry Pi springs into action, initializing speech services and establishing a connection with the Microsoft Azure API for audio processing. Simultaneously, the ReSpeaker Mic Array on the Raspberry Pi captures environmental audio sources, which are then analyzed and transcribed into text by the API. The transcribed text is sent back to the Raspberry Pi and subsequently transmitted to the AR application via UDP. Completing the loop, the AR application, developed using Unity Version 2022.3.18, receives this transcribed text using the Microsoft Speech SDK package[35] and displays it on the headset, providing the user with real-time captions of their surroundings. If the user wishes to respond or communicate, they can manually activate the ‘Text-to-Speech’ function using specific gestures. This will display a

[35] Microsoft Speech SDK package: https://github.com/Azure-Samples/cognitive-services-speech-sdk/tree/master/quickstart/csharp/unity/text-to-speech

keyboard for users to type text, which will then be vocalized through the Meta Quest speaker, thereby facilitating two-way interaction.

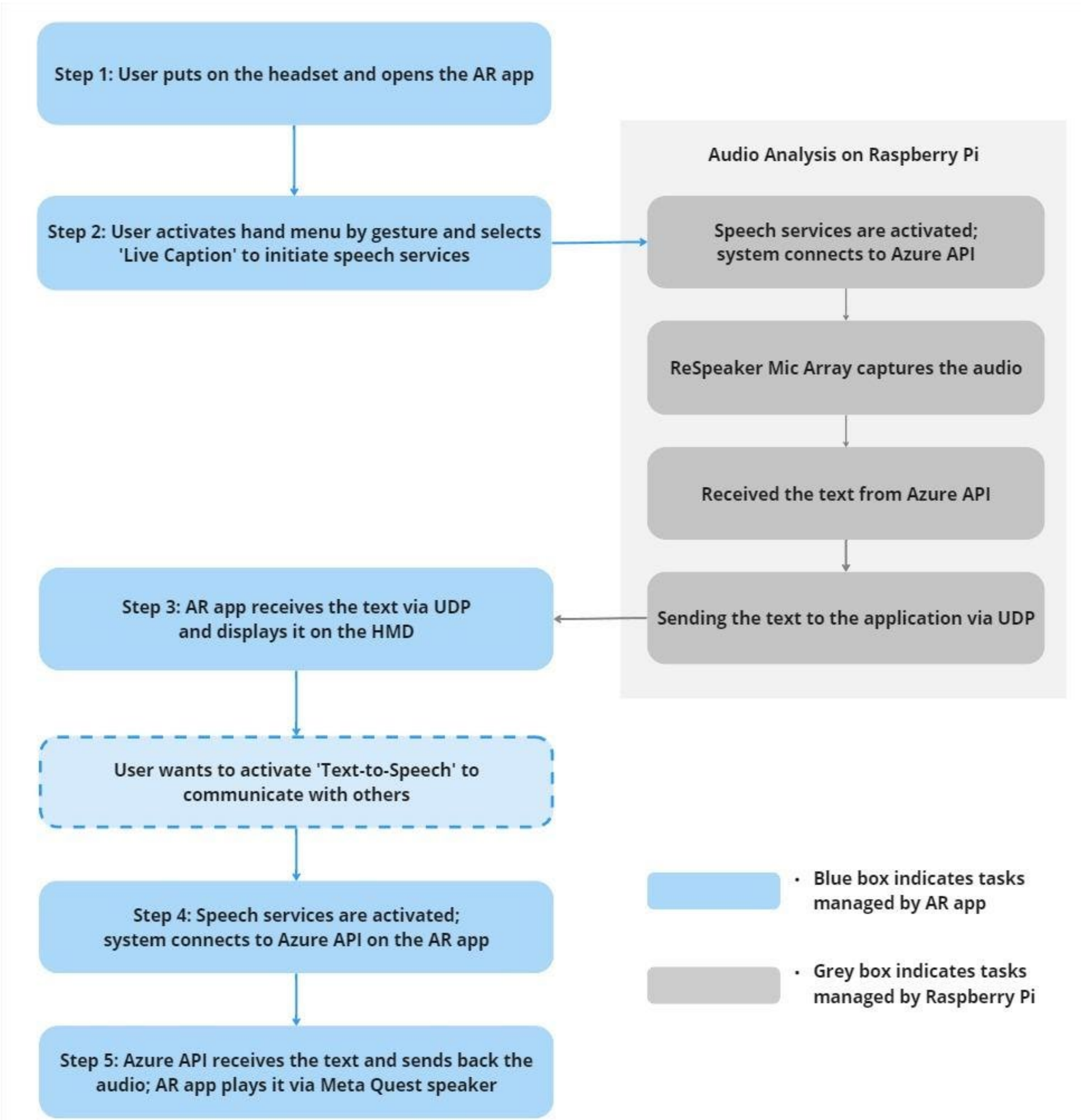


Figure 20. Prototype 4 System architecture and workflow.

Process of Prototype 4

The Raspberry Pi system operates similarly to Prototype 3, executing an application that transmits audio to the Azure API and receives the recognized text in return. This text is then forwarded via UDP protocol on the AR application which runs on HMD for display.

Within the AR application side, when users initiate the app on the HMD, they can activate the menu by moving their left-hand palm towards their face and deactivate it by moving the hand in the opposite direction. This design choice reduces menu interruptions during non-interactive phases, aligning with the application's goal to blend seamlessly into daily life. With the menu activated, users are presented with two clear options: home management and live caption, maintaining a clutter-free system interface to prevent information overload (see Figure 21).

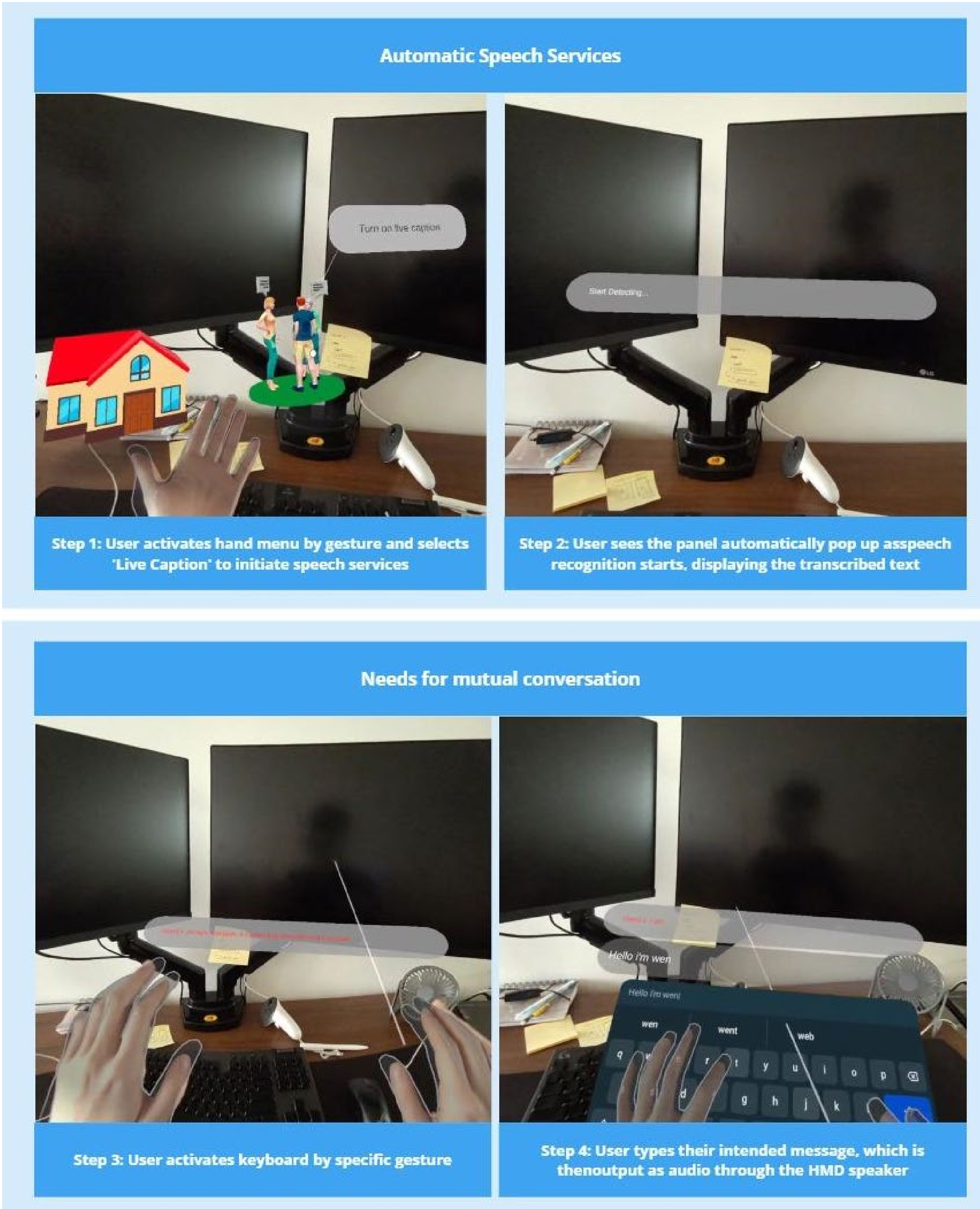


Figure 21. User Journey of Prototype 4. The upper section shows automatic speech service features, while the lower section requires manual 'Text-to-Speech' activation via specific gestures.

Interactions within the user interface utilize a ray casting method based on the hand's position, reducing the necessity for wide-ranging hand movements. This method is designed to prevent accidental physical contact with others and eliminates the need for exact fingertip placement on interface buttons, streamlining the interaction process.

The 'Live Caption' function encompasses several features managed by Azure's speech service, including real-time speech-to-text, text-to-speech, real-time translation, and real-time speaker identification, each designed for different scenarios:

- ***Real-time Speech-to-Text (Automatic):*** Activates automatically upon selecting the live caption function, providing immediate captioning as the user navigates the interface. This feature is designed to accommodate a variety of scenarios, such as viewing videos or listening to podcasts.
- ***Text-to-Speech (Manual):*** Designed for engaging in mutual conversations, this feature activates when a user performs a specific hand gesture, pushing both hands down, which triggers a virtual keyboard and text panel for typing. Leveraging predictive typing, the virtual keyboard minimizes effort and typing time. When a user presses the 'enter' key, the typed text converts into audio and plays through the headset speaker.
- ***Real-time Translation (Automatic):*** Azure's service automatically detects and translates supported foreign languages into English, providing instant translations to the user. This feature is designed for translation needs, such as watching foreign TV episodes or traveling abroad.
- ***Real-time Speaker Identification (Automatic):*** The system automatically identifies speakers, labeling them as "Guest" followed by a number (e.g., "Guest-1", "Guest-2"), to distinguish between different speakers in a conversation. This feature is designed for scenarios such as engaging in group discussions.

Reflection on Prototype 4

Prototype 4 represents a significant advancement in enhancing communication for DHH individuals, leveraging AR interfaces and sophisticated real-time AI speech services. This prototype integrates features such as real-time Speech-to-Text, Text-to-Speech, Translation, and Speaker Identification. Notably, 'Text-to-Speech' is the only feature requiring manual activation via specific gestures, while the others operate automatically without needing user intervention. Such design approach allows users to seamlessly perceive speech, mirroring the experience of hearing individuals. Users activate the 'Text-to-Speech' feature only as needed, ensuring that it integrates smoothly without disrupting other activities.

Also, the transition from static 2D text-based menus to a 3D model-based interface with motion design enhances user experience, making interactions more natural and aligning with human spatial understanding.

However, there are some main limitations that should be noted:

- ***Lack of Customization in Azure's Pre-trained Models:*** Using Azure for speech services offers real-time capabilities but restricts customization, such as generic speaker tags ('Guest-1', 'Guest-2'), making interactions with familiar people feel impersonal and affecting the user experience.
- ***User Interface and Interaction Adaptability:*** The prototype lacks an introductory tutorial and faces challenges with gesture recognition accuracy, leading to occasional unintended activations that could impact usability.
- ***Service Cost:*** Throughout the development of Prototype 4, experimentation remained within the free tier limits of Azure services. Yet, in a real-world deployment, costs could increase significantly.

- ***Personalized Receiving Text Setting:*** The current real-time Speech-to-Text feature operates in full transcription mode, meaning users receive the transcribed text only after the speaker completes a sentence without significant pauses. For future development, it would be beneficial to allow users to customize their text-receiving settings. For instance, a 'partial receiving' option could be introduced, where users receive transcriptions word by word as the speaker talks, providing continuous updates. Although this might increase the cognitive load for some users, it would offer them the flexibility to choose the mode that best suits their preferences.

Subjective Evaluation of Prototype 4

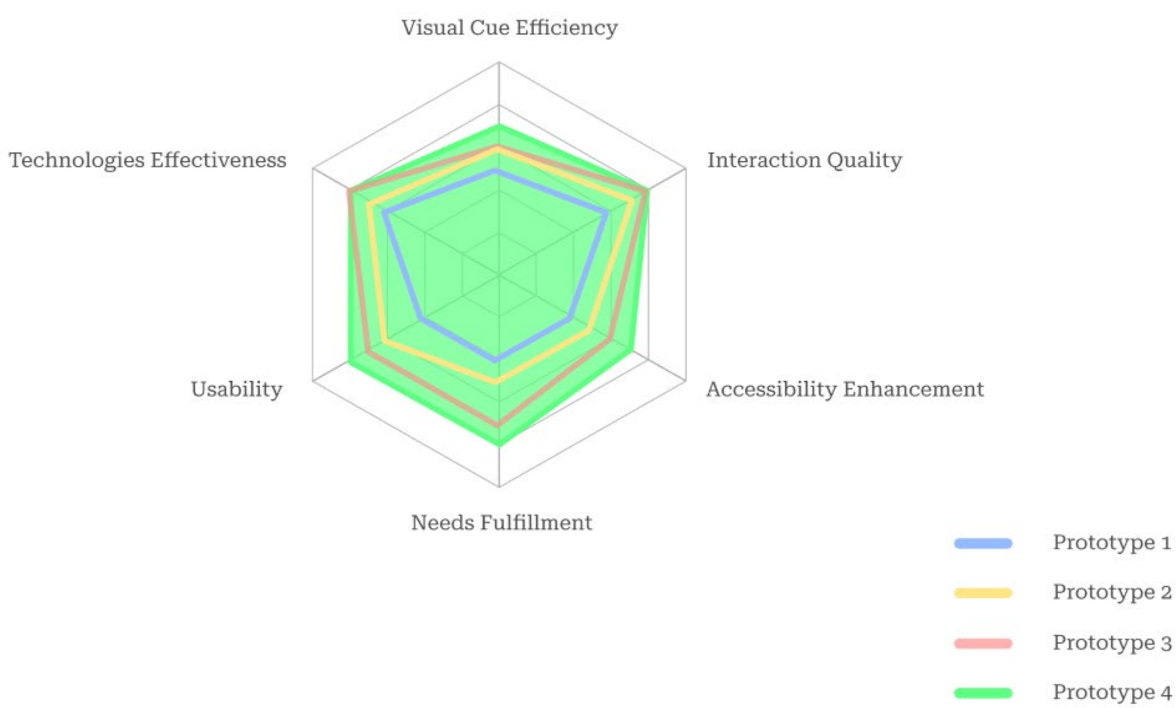


Figure 22. Radar chart for Prototype 4.

Visual Cue Efficiency (3.5/5): Prototype 4's real-time speech-to-text conversion facilitates DHH users' communication, effectively using visual cues to display spoken words.

Incorporating a broader array of visual feedback, such as animated indicators for speaker activity, could further enhance comprehension.

Interaction Quality (4/5): Gesture controls and raycasting in Prototype 4 enhance intuitiveness, cutting down on physical buttons and avoiding unnecessary contact. By fine-tuning gesture accuracy and introducing initial tutorials, alongside minimizing precise button presses, this prototype significantly streamlines interactions.

Accessibility Enhancement (3.5/5): Prototype 4 greatly supports DHH users in communication through vital speech services. However, its approach, particularly the text-to-speech output being captured by the external mic array, can inadvertently trigger the speech-to-text feature. This, combined with restricted customization for speaker identification, underscores a need for refining the system's personalization and operational distinctiveness to better serve user requirements.

Needs Fulfillment (4/5): Addressing critical communication barriers, Prototype 4 makes notable strides in fulfilling the needs of the DHH community with its real-time captioning and speaker identification. Expanding service customization could provide a more tailored experience. Moving forward, broadening speech features to not only bridge hearing gaps but also provide more comprehensive information, enhancing both understanding and engagement, will further the prototype's capabilities beyond basic communication needs.

Usability (3.5/5): Prototype 4's user-friendly design, characterized by gesture navigation, makes it accessible and easy to use. Implementing a detailed tutorial for new users and refining gesture recognition to eliminate misinterpretations will further improve the overall user experience.

Technologies Effectiveness (4/5): The integration of AR, AI, and IoT in Prototype 4 provides strong support for DHH individuals in communication contexts. Azure's speech service delivers stability with high accuracy and low latency, significantly enhancing the system's effectiveness. The primary concern remains the cost of utilizing these services.

5. Final Prototype

Final Prototype Overview

The final prototype[36][37] is an AR application that represents the culmination of integrating AR, AI, and IoT technologies to offer a holistic solution aimed at enhancing the home living experience for DHH individuals. The development is driven by the pivotal research question: *'How can the fusion of AR, AI, and IoT technologies within HMD platforms foster new possibilities for enriching the home living experience of DHH individuals?'* Along with addressing this question, the project explores related sub-questions to build a solid foundation for comprehensive solutions.

Drawing on the insights gained from each prototype during the iterative development phase, which was specifically tailored to meet the distinct needs of the DHH community, I continue to identify and integrate new features to address previously unmet needs, thereby enhancing the solution's comprehensiveness.

During the initial phase of the final prototype development, the focus remains on deepening the exploration of user interface intuitiveness to minimize cognitive load, adapting a more user-friendly design. This phase involves creating a representative 3D model of the user's home, mirroring the actual layout, structure, and device placements, in order to facilitate easier device management and reduce the cognitive load. Additionally, I expand system compatibility to include a broader range of devices, leveraging Zigbee[38] for direct device communication and MQTT for lightweight messaging. In addition, users can refer to the Zigbee2MQTT[39] website, which offers a comprehensive list of Zigbee devices supported for MQTT communication. On

[36] Final Prototype video link: https://youtu.be/4YytMwS7ik4
[37] Final Prototype interaction method video link: https://youtu.be/z4Dv3mkz-7o
[38] Zigbee: https://www.digi.com/solutions/by-technology/zigbee-wireless-standard
[39] List of supported Zigbee devices: https://www.zigbee2mqtt.io/supported-devices/

this website, users are able to select devices within their preferred price range, ensuring they find the best fit for their needs. The enhancements in the device range not only improve device management efficiency but also enrich the overall user experience by offering more direct and reliable monitoring and control capabilities in their homes.

Subsequently, the Speech Services, originally introduced in Prototype 4 are further refined. These services include ***real-time captioning, translation, speaker identification,*** and ***manual activation of text-to-speech***, directly addressing essential communication needs. These features enhance communication by facilitating one-to-one conversations through real-time speech-to-text and text-to-speech, incorporating speaker identification in group discussions, and providing real-time translation for breaking language barriers.

Thus, the holistic approach of integrating different features for speech services function not only addresses the immediate communication needs of DHH individuals but also greatly enriches their daily interactions, highlighting the prototype's continuous advancement in tackling the diverse challenges faced by the DHH community.

The final prototype, therefore, represents a fusion of the most effective elements from its predecessors, addressing the broad spectrum of needs within the DHH community with a comprehensive and cohesive approach.

The following table summarizes how the final prototype addresses identified needs by leveraging the integrated capabilities of AR, AI, and IoT technologies, and how it represents an evolution from the approaches taken in earlier prototypes.

Category	Needs Finding from Literature Review and Online Survey	How Final Prototype Addresses the Needs	Advancements Over Previous Prototypes
Advancing Home Accessibility Through Technology	Challenges with **auditory-based** smart home devices	Utilizes **direct hand control** and integrates **more devices**, thereby **improving user autonomy** in device control	1. Direct hand control and ray cast interaction methods 2. Broader device selections are tailored to the user's needs, ensuring wider system compatibility
Visual Representation (UI Design)	**Heavy cognitive load** from complex and text-based UI design	Develop a **representative 3D home model** of the user's actual home for a more **intuitive interaction**	The transition from static 2D text-based UI interactions to dynamic, spatially coherent home management through a 3D model-based UI design
Visual Representation (Sound Alerts)	Dependence on **visual aids for alters**	Automated **visual alerts for appliance** status changes and malfunctions	Enhancement on alter system using color and text for immediate recognition
Communication Barriers	Dependence on digital devices presents challenges such as **latency and inconvenience**	**Real-Time Speech Services,** including transcription, translation, speaker identification	Add comprehensive communication features within the AR environment for seamless interaction

Table 9. Final Prototype Overview. Comparative advancements of Final Prototype over previous prototypes in addressing the identified needs and limitations of DHH individuals.

Design of Final Prototype

The final prototype features two primary functions: ***Home management and Speech Services.***

Function 1: Home Management

User Interface Design

To further innovate the concept of 3D model user interface design, I aim to create a more intuitive interaction method by designing a home model based on the actual user's home. Initially, I utilized the ‘Canvas’ application on my iPhone 13 to conduct a scan of my apartment. This app is capable of generating 3D models from camera scans within minutes. However, the scan results in a model with excessive vertex detail, demanding significant computational resources, making it unsuitable for HMD implementation. To address this, I craft a low-poly model of my home, maintaining key furniture, appliances, and the accurate layout and structure of the room, which drastically reduces resource requirements (See Figure 23).

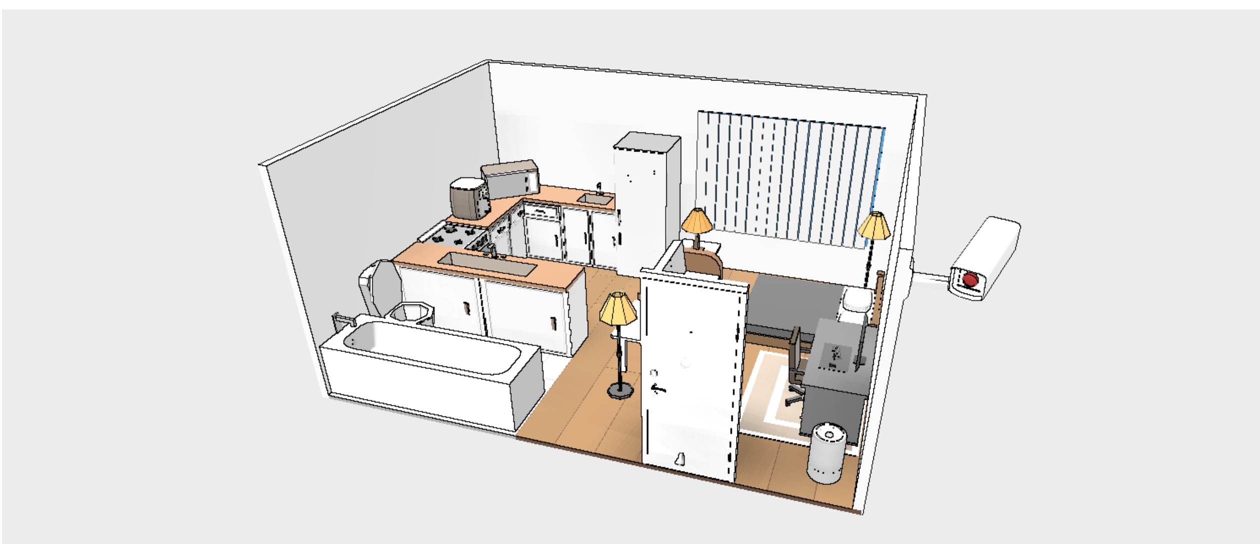


Figure 23. Low-Poly 3D model of the author’s apartment. The figure displays the layout, structure, and device placement, accurately reflecting the real home environment.

Building on this foundation, I enhance the model with visual cues to facilitate room selection. This is illustrated when a user selects a room, such as the bathroom, which is then highlighted in color to stand out from other areas that remain grayed out. This design approach intuitively directs the user's focus to their active interaction within the virtual environment (see Figure 24).

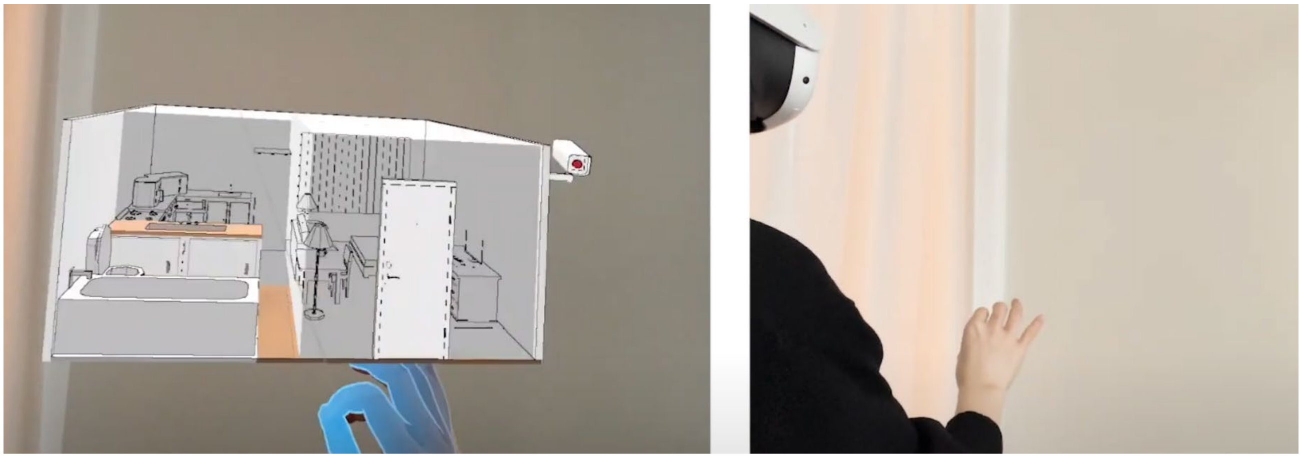


Figure 24. Visual cues for room selection. This figure showcases the bathroom selection, highlighted in colors, with other rooms grayed out to direct user focus.

To improve interaction with the home model, I implemented a feature that allows users to adjust the model's size by dragging the walls inward or outward with both hands. Adjusting the distance between the hands changes the model's size—larger distances increase the model size, and vice versa. Additionally, users can reposition the room model by dragging it to a preferred location, facilitating easier interaction (see Figure 25). This interaction method offers flexibility and control over the interface using straightforward, intuitive gestures, which require only minimal learning and reduce cognitive load.

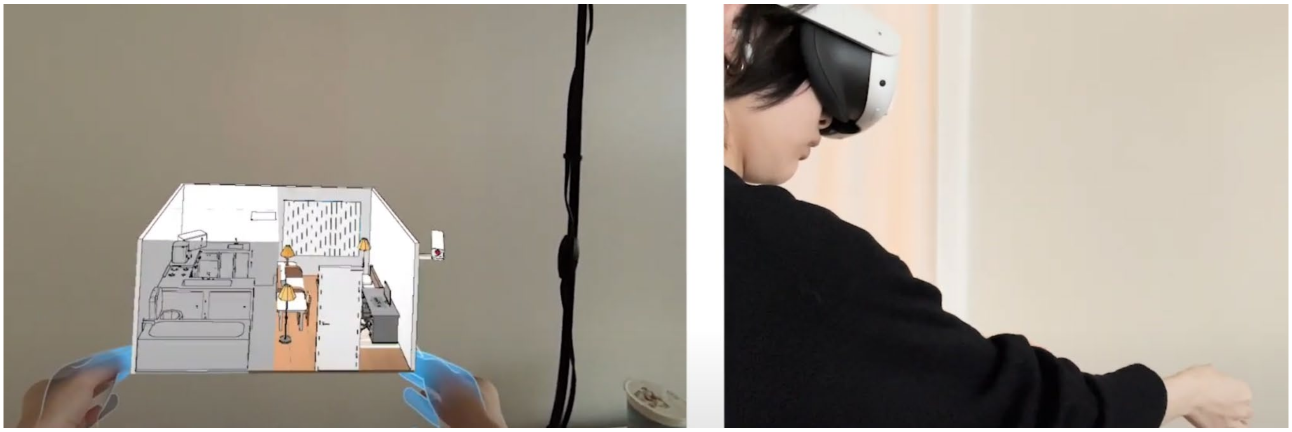


Figure 25. The user interacts with the home model. The figure demonstrates how users can adjust the model's size or reposition it by changing the distance between their hands or employing a drag-and-place gesture, indicated by the hand tracking image turning blue.

Beyond enhancing spatial navigation through intuitive 3D modeling, the system introduces an automated approach to home appliance notifications. The approach is to monitor sharp power level fluctuations—like sudden drops—and promptly inform the user of appliance malfunctions. To better provide users with personalized event categorization, this system designs a Notification System that categorizes events into two levels: non-urgent and urgent. Non-urgent events include scenarios such as someone entering a room or household appliances completing their tasks. Urgent events, for example, might involve the refrigerator door being left open or a bathtub leaking.

For non-urgent events, such as when an appliance finishes its task, the system models the appliance in green accompanied by an animation—for instance, the air fryer's tray extends, along with a pop-up menu informing the user with text that 'the air fryer is ready' (see Figure 26). Similarly, if someone enters a room, the door model in the system triggers an opening animation, and a person entering animation from the doorway is also triggered, accompanied by a pop-up menu to inform the user (see Figure 27). For non-urgent events, the user can acknowledge this information by clicking a checkmark on the pop-up menu, which then returns the model to its original state and makes the pop-up menu disappear, ensuring it does not further disturb the user.

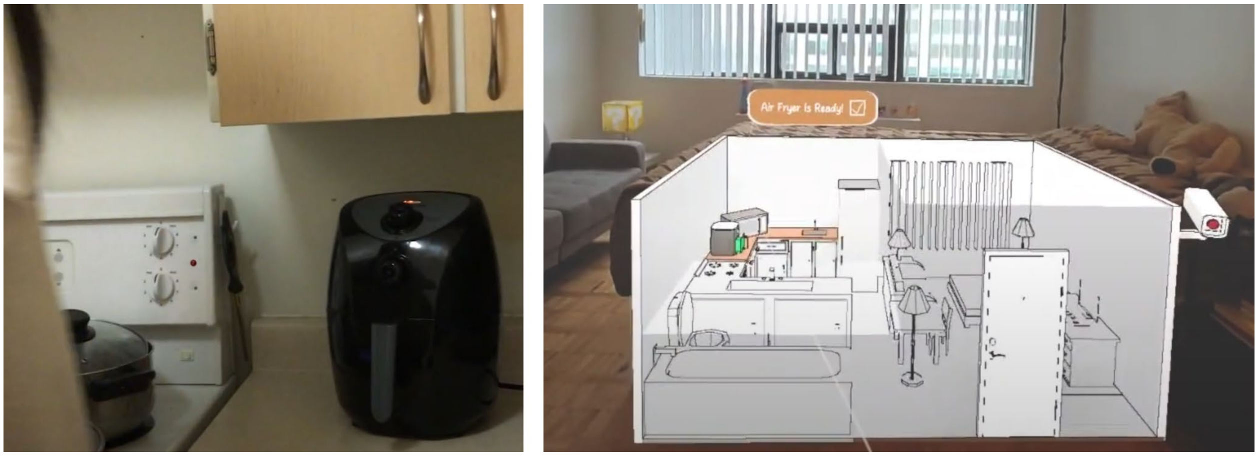


Figure 26. Visual cues for Air Fryer status. The air fryer model turns green and triggers its corresponding animation, and a pop-up menu appears, indicating that the air fryer has finished operating.

Figure 27. Room Entry Visualization. As a person enters, the room's door model animates to open, and a figure animation shows the individual stepping into the room. Simultaneously, a pop-up menu appears to inform the user of the entry.

For urgent events, such as the refrigerator door being left open, the system turns the model red, triggers an animation, and displays a pop-up menu. Differently from non-urgent events, the animation and the model’s state do not revert by clicking a checkmark but require the user to manually resolve the issue before the model returns to normal. For instance, if the refrigerator door is left open, the system detects this status, and the refrigerator model's door turns red and is accompanied by an animation and a menu. The user must manually close the refrigerator door for the model to revert to its original state (see Figure 28).

In conclusion, the advancement in the visual representation of sound alerts and notification system design fulfills the unique needs of DHH users by delivering easily recognizable visual notifications regarding home events and appliances' status, significantly improving interaction clarity and awareness within the home environment.

Figure 28. Refrigerator Alert Visualization. Notification appears if the door is left open and resets only after the user closes it.

Following the automated and visual-based notification system, the design further evolves to enhance user interaction through the home device control feature. Users can directly select device models within the virtual home model to access a control panel brimming with detailed options for operation (see Figure 29). The integration not only simplifies the control process but also enriches the user's ability to manage their home environment effectively, blending the flow of automated alerts with the hands-on device control experience.

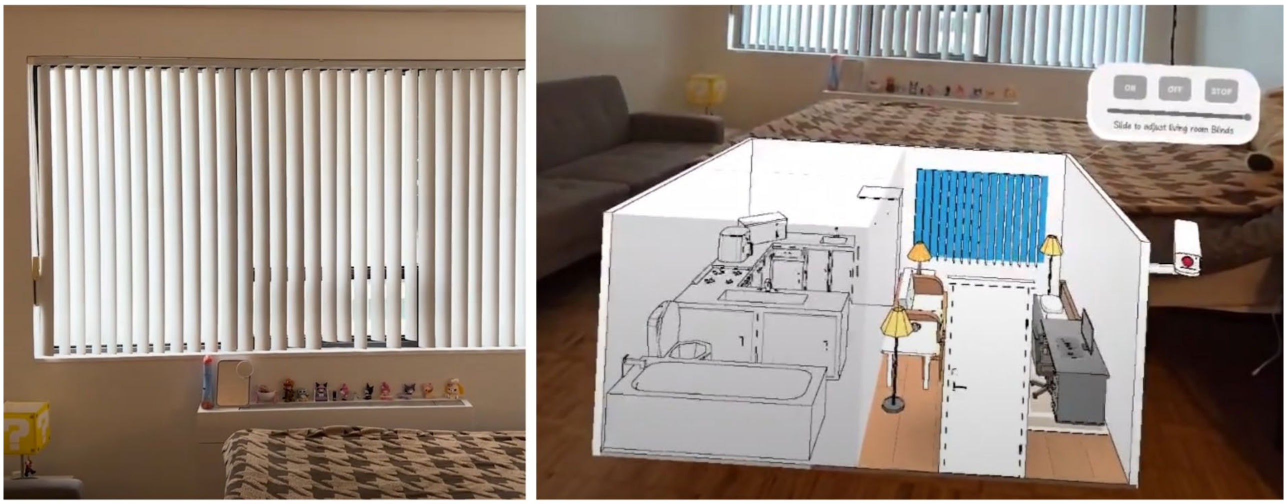


Figure 29. Pop-up menu for the blinds model. A pop-up menu details control options appear upon user selection of the blinds model.

The surveillance system explored in Prototype 3 continues with the same hardware setup but features an upgraded visual design. A Raspberry Pi application streams security camera footage via a local webpage created with the Flask Library, enabling real-time monitoring within the AR application. Simultaneously, the YAMNet model performs sound classification and sends the data to the AR application via UDP. The AR application then displays this sound text beneath the streaming video feed (refer to Figure 30). This integration activates when the user selects the camera model, then a menu appears allowing the user to view the camera footage, thereby providing both visual and auditory information in a unified stream.

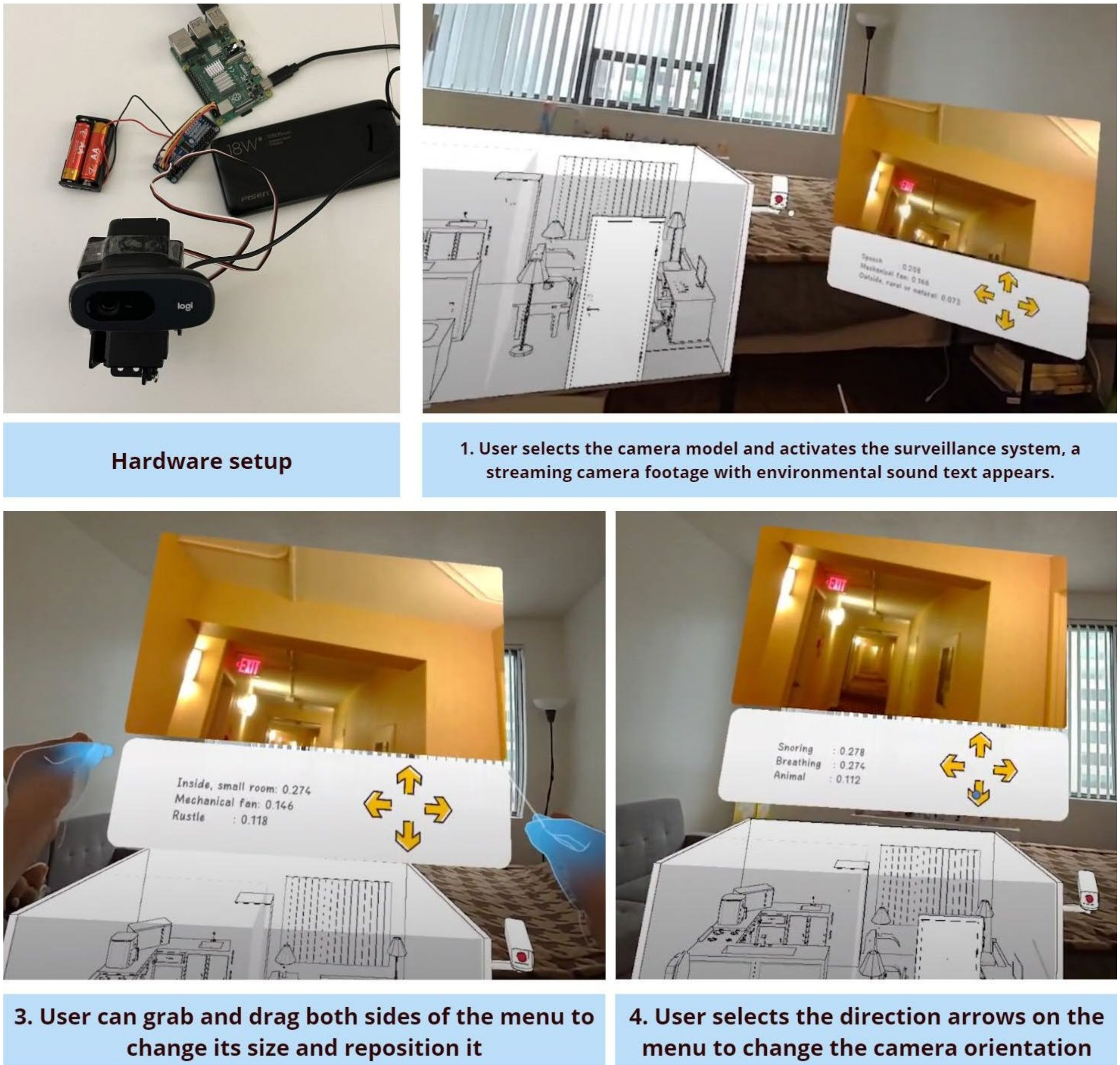


Figure 30. User Journey Map of the surveillance system feature.

Hardware Setup and System Workflow of Home Management

Addressing a critical need for safety and situational awareness among DHH users, the system introduces an alert system for appliance and device statuses. Drawing from the lessons learned in Prototype 3, which faced challenges with sound classification's sensitivity to noise interference, particularly with household appliances requiring precise detection, a shift is made to a more reliable method. The design involves employing Zigbee network and MQTT communication for robust home device monitoring.

The Zigbee network within the prototype ecosystem incorporates a variety of smart devices, each utilizing the Zigbee protocol for seamless communication. 'ThirdReality' smart plugs[40] stand out with their real-time power monitoring capability, providing swift and precise notifications about appliance activity to the system (refer to Figure 31). This ensures users are immediately informed of any operational changes. Alongside these, the 'Tuya' door sensor[41] is employed to monitor the open or closed status of doors, offering versatility in its application, such as confirming if a fridge door or a window is left open, thus illustrating its broad utility (see Figure 32).

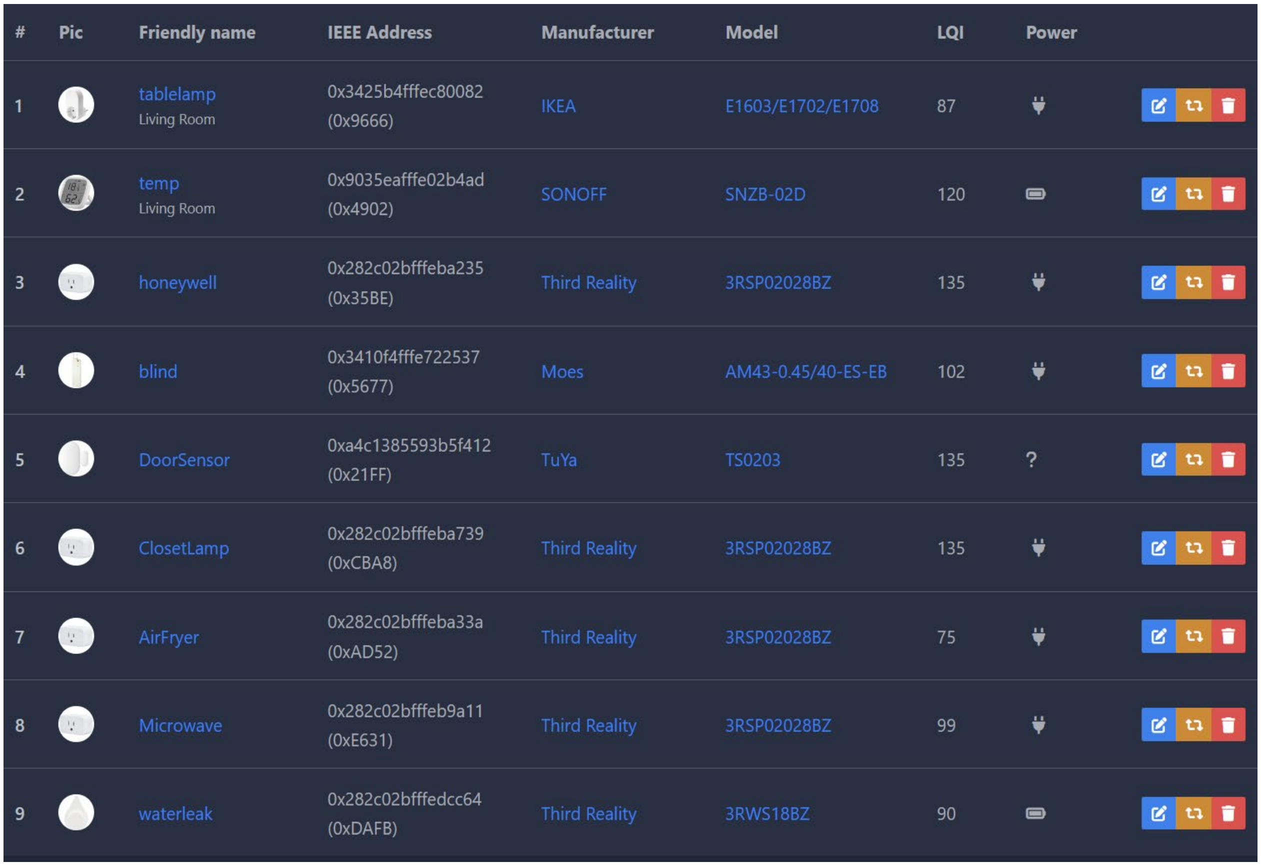
Figure 31. 'ThirdReality' Smart Plug. The smart plug is used for power monitoring.

40 ThirdReality Smart Plug: https://www.zigbee2mqtt.io/devices/3RSP019BZ.html
41 Tuya Door Sensor: https://www.zigbee2mqtt.io/devices/TS0203.html

Figure 32. 'Tuya' Door sensor. A pair of door sensors is attached to the bottom of the door.

The prototype's device suite also includes a temperature and humidity sensor, a motorized controller for blinds, and a water leak sensor, among others (refer to Figure 33). These diverse Zigbee devices, sourced from various manufacturers, interface with the MQTT broker – facilitated by the Raspberry Pi and the SONOFF Zigbee Dongle USB[42] (see Figure 34). This dongle acts as a translation gateway, converting Zigbee signals into MQTT messages, which bind the components into a unified, responsive network. The integration affords real-time communication between the AR application and the networked devices, enabling a seamless user experience (refer to Figure 35).

[42] SONOFF Zigbee Dongle USB: https://sonoff.tech/product/gateway-and-sensors/sonoff-zigbee-3-0-usb-dongle-plus-p/



#	Pic	Friendly name	IEEE Address	Manufacturer	Model	LQI	Power
1		tablelamp Living Room	0x3425b4fffec80082 (0x9666)	IKEA	E1603/E1702/E1708	87	
2		temp Living Room	0x9035eafffe02b4ad (0x4902)	SONOFF	SNZB-02D	120	
3		honeywell	0x282c02bfffeba235 (0x35BE)	Third Reality	3RSP02028BZ	135	
4		blind	0x3410f4fffe722537 (0x5677)	Moes	AM43-0.45/40-ES-EB	102	
5		DoorSensor	0xa4c1385593b5f412 (0x21FF)	TuYa	TS0203	135	?
6		ClosetLamp	0x282c02bfffeba739 (0xCBA8)	Third Reality	3RSP02028BZ	135	
7		AirFryer	0x282c02bfffeba33a (0xAD52)	Third Reality	3RSP02028BZ	75	
8		Microwave	0x282c02bfffeb9a11 (0xE631)	Third Reality	3RSP02028BZ	99	
9		waterleak	0x282c02bfffedcc64 (0xDAFB)	Third Reality	3RWS18BZ	90	

Figure 33. A list of Zigbee devices used in the final prototype. The figure is a screenshot from the Zigbee2Mqtt frontend webpage.

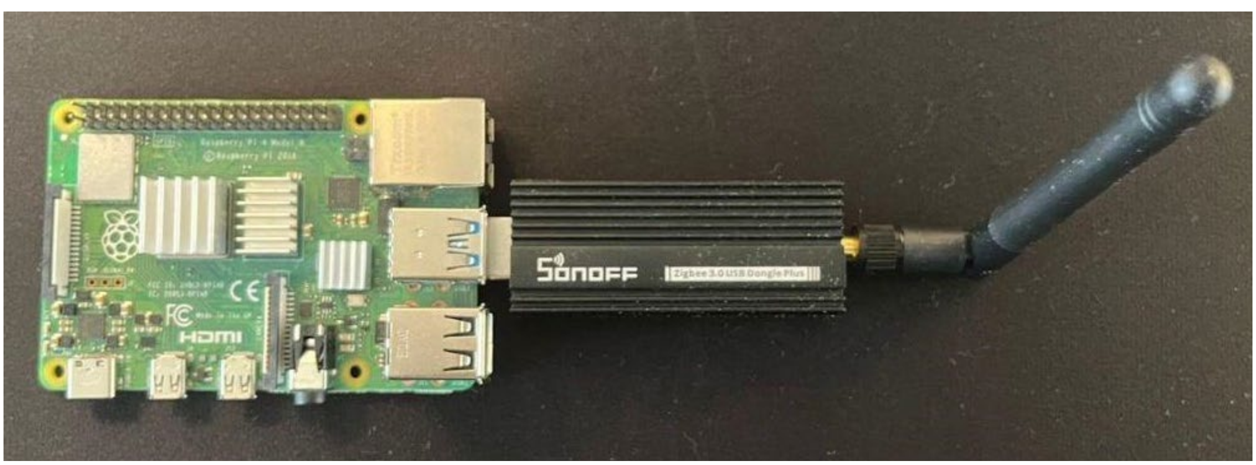


Figure 34. A SONOFF Zigbee 3.0 USB Dongle is attached to a Raspberry Pi.

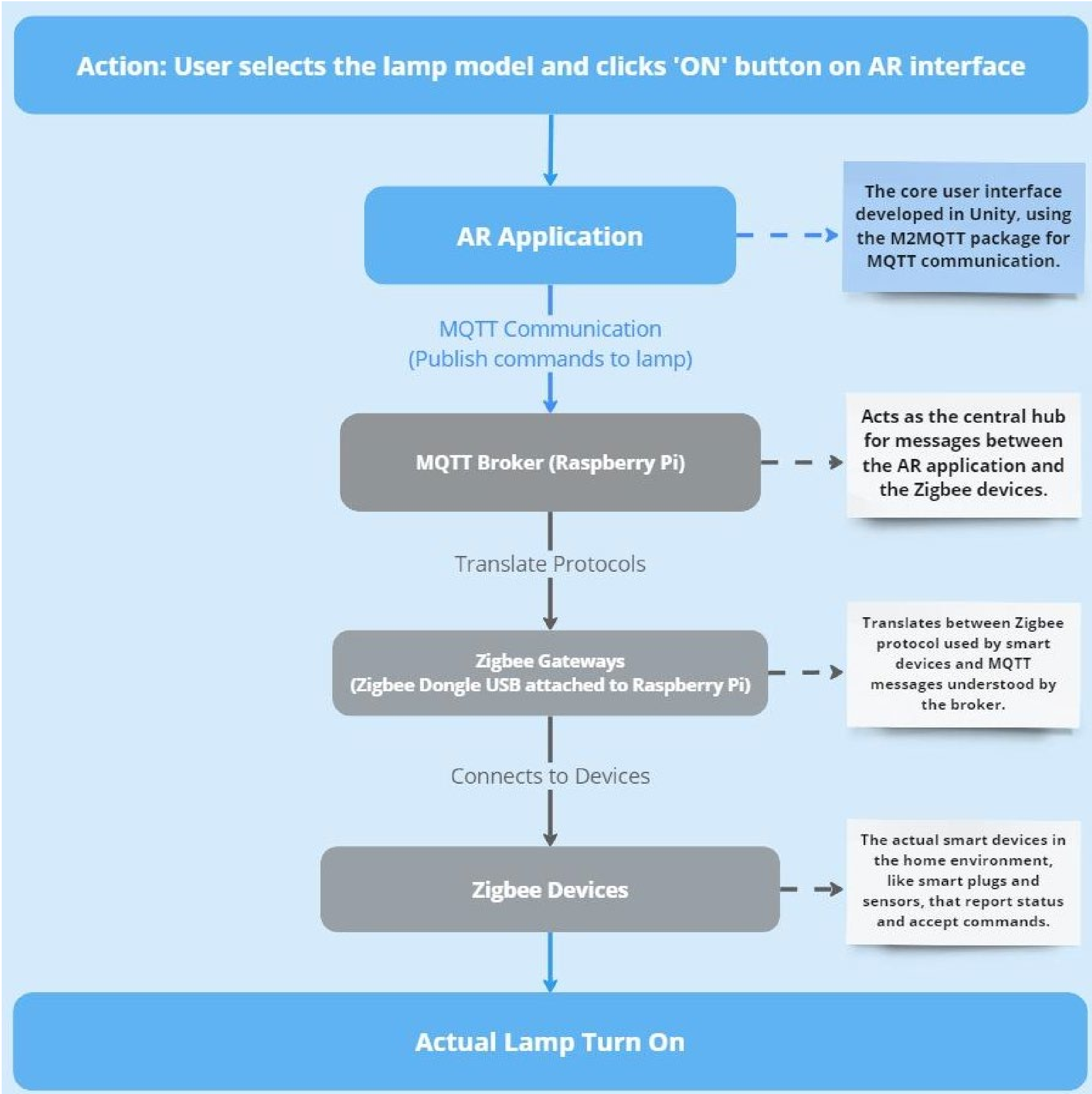


Figure 35. Final Prototype device management system workflow.

Building on the established Zigbee network, the final prototype harnesses the Unity Editor to create an AR application, running on a Meta Quest 3 headset. Leveraging the M2MqttUnity[43] package, it connects the user directly to the device network, enriching the home

[43] M2MqttUnity: https://github.com/gpvigano/M2MqttUnity

living experience. By launching the AR application, users subscribe to the Zigbee devices within their real-time updates and send commands to them (refer to Figure 36). This seamless interaction embodies the convergence of the IoT framework with user-centric control, ensuring a cohesive and intuitive management system (see Figure 37).

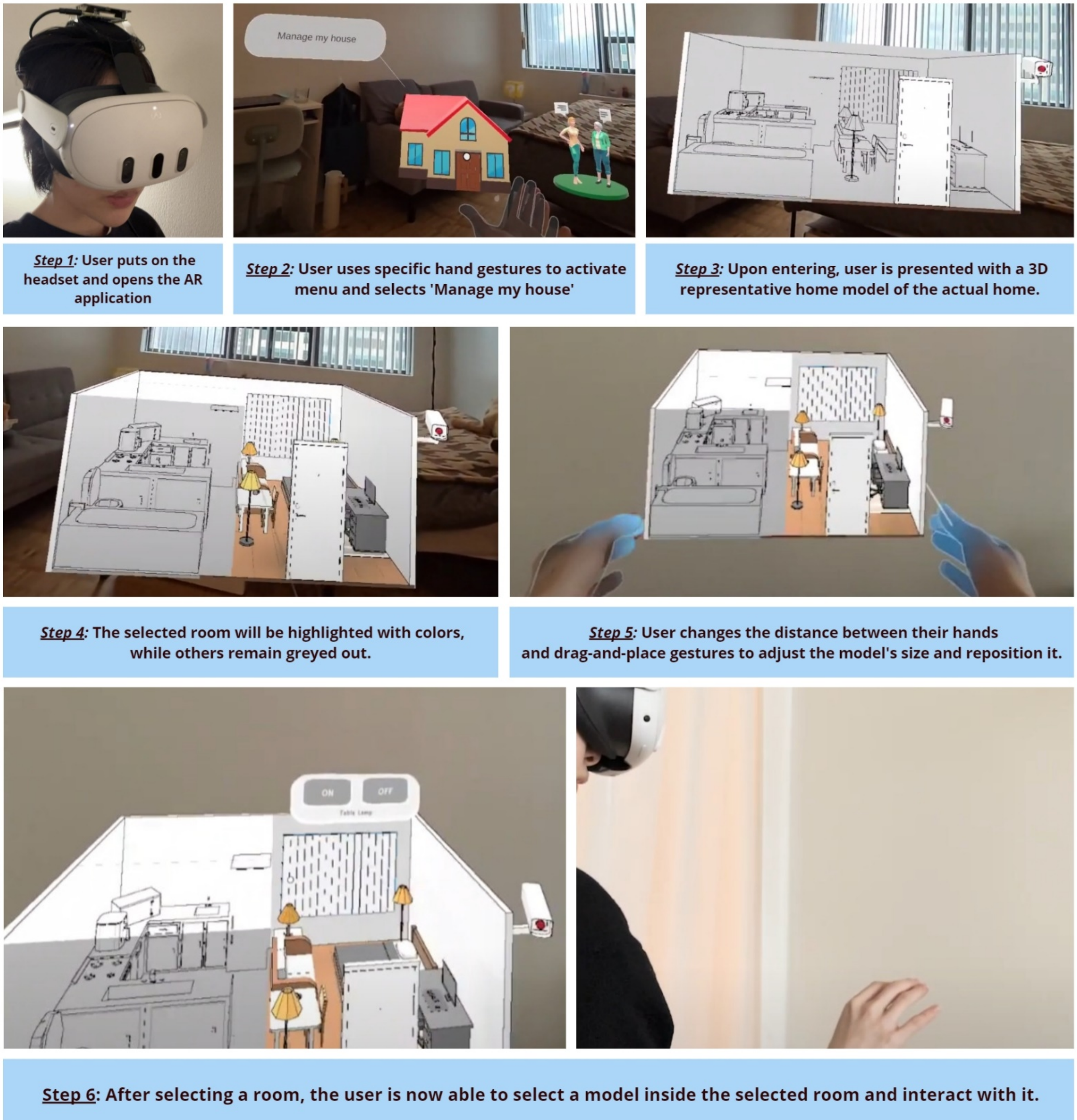


Figure 36. User Journey Map of home management function.

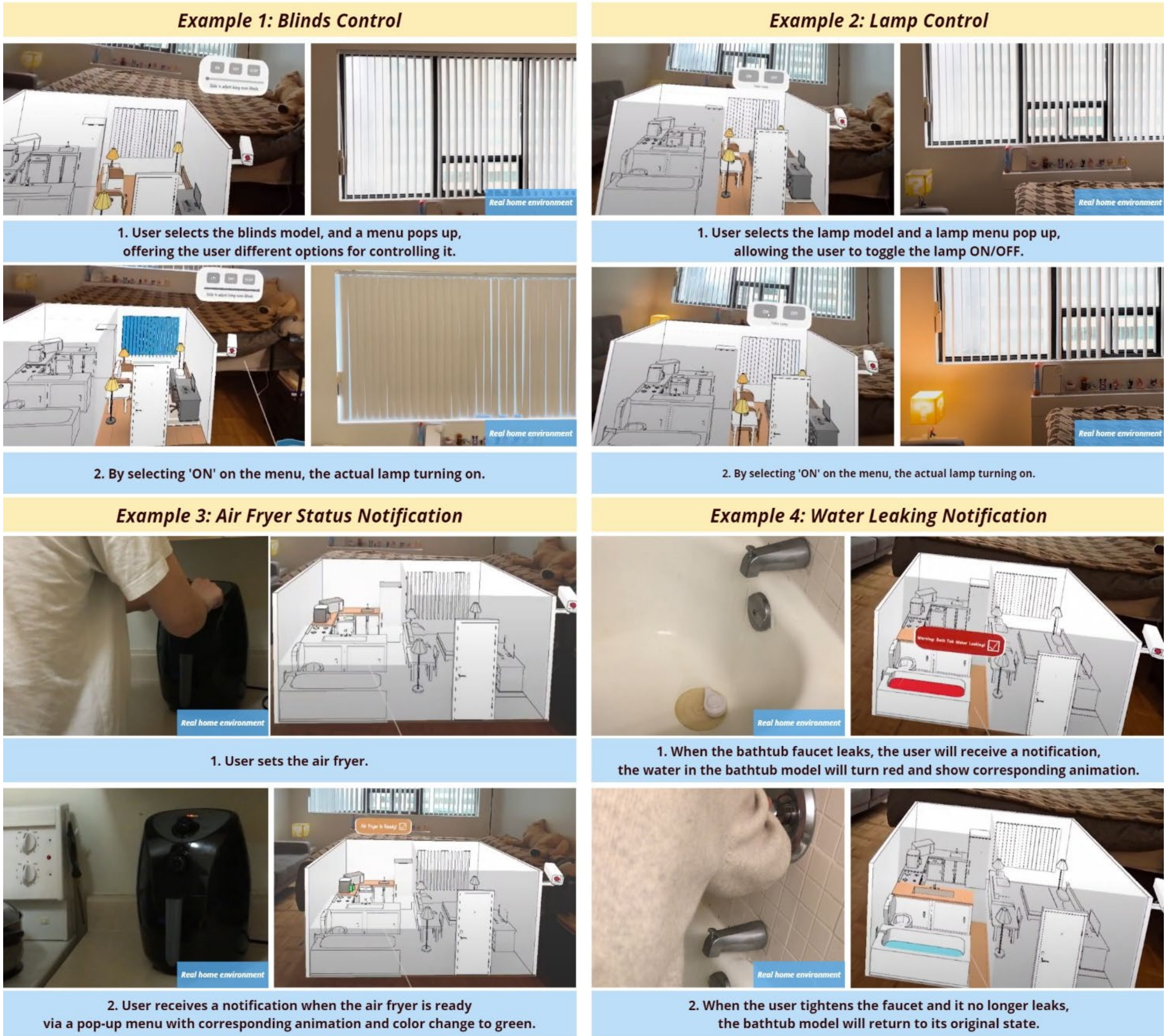


Figure 37. Examples of Home device management and notification system. The figure depicts the user journey map with the interaction examples of Zigbee device controls and the notification system.

Function 2: Speech Services

The speech services and hardware setup for the final prototype maintains the main functionalities of Prototype 4, such as real-time captioning, translation, speaker identification, and manual activation of text-to-speech. These features collaboratively enhance communication for DHH individuals in a variety of settings. The following sections provide a detailed depiction of the user journey, demonstrating the capabilities of the 'Speech Services' across different scenarios.

Scenario 1: Communication with others - Use case of 'Real-Time Speech-to-Text' and 'Text-to-Speech'

Scenario 1 illustrates the use of 'Real-Time Speech-to-Text' and 'Text-to-Speech' functions within the AR application, facilitating seamless one-to-one communication between a DHH user and another individual (see Figure 38). This use case demonstrates how the application transcribes spoken language in real-time for the DHH user, while also allowing them to respond via text that is converted into spoken words through the Text-to-Speech feature.

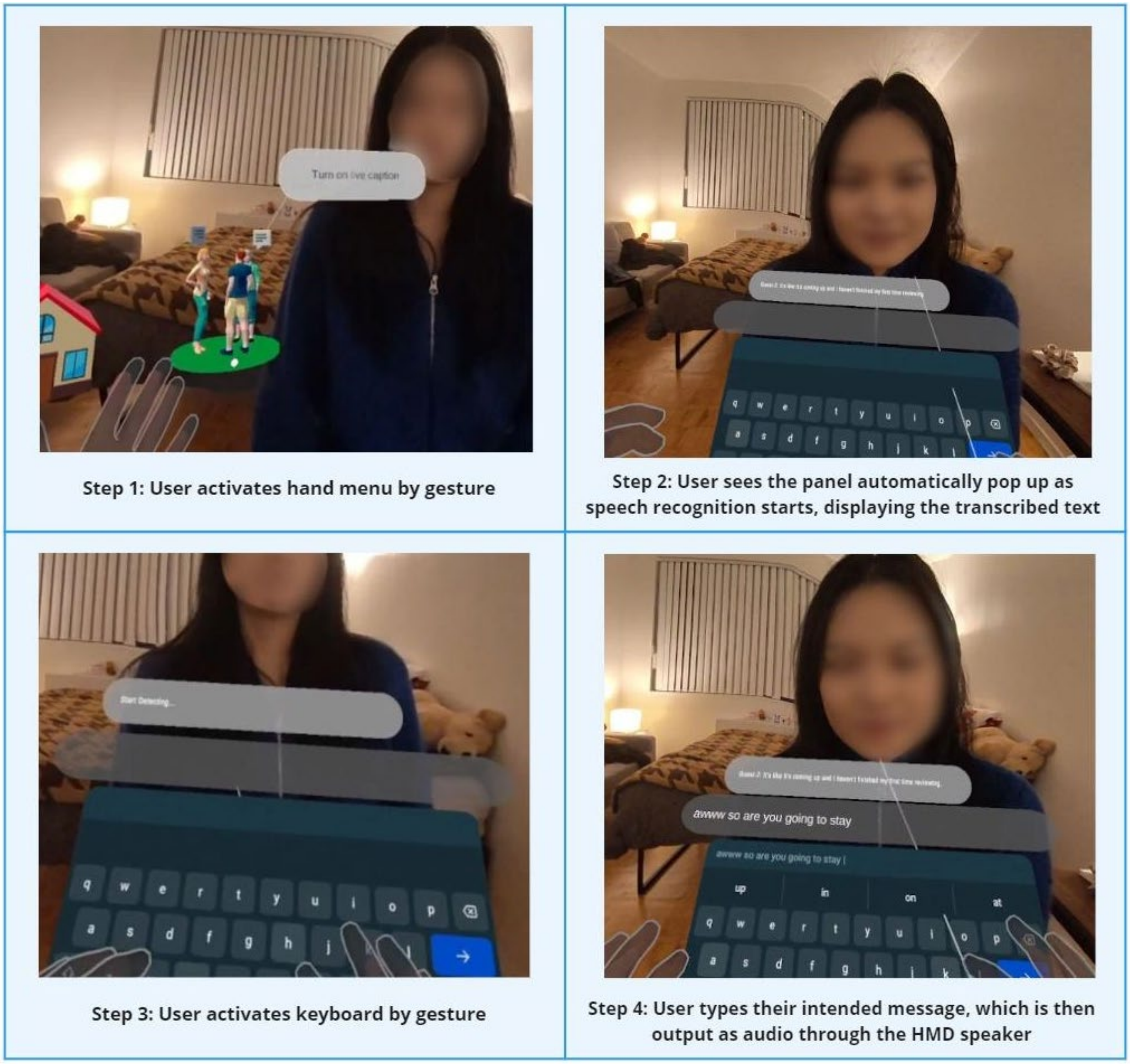


Figure 38. User Journey of Communication use case. The figure illustrates that the user is communicating with only one person. Note that the participant's face is blurred to protect privacy and maintain anonymity.

Scenario 2: Watching a Foreign Drama - Use case of 'Real-Time Translation'

In this scenario, a DHH user is watching a foreign drama. The dialogue, in a language the user does not understand, presents an additional challenge beyond their hearing impairment. This is where the AR application's 'Real-Time Translation' feature becomes invaluable. As the characters in the drama speak, the application transcribes and translates the spoken language in real-time. The translated text then appears in the AR display, allowing the DHH user to follow along with the plot despite the language barrier (see Figure 39).

This use case demonstrates the application's potential to break down language barriers and enhance the entertainment experience for DHH users. Whether it's a drama, a movie, or a live performance in a foreign language, the 'Real-Time Translation' feature can make it accessible and enjoyable for all users, including those who are DHH.

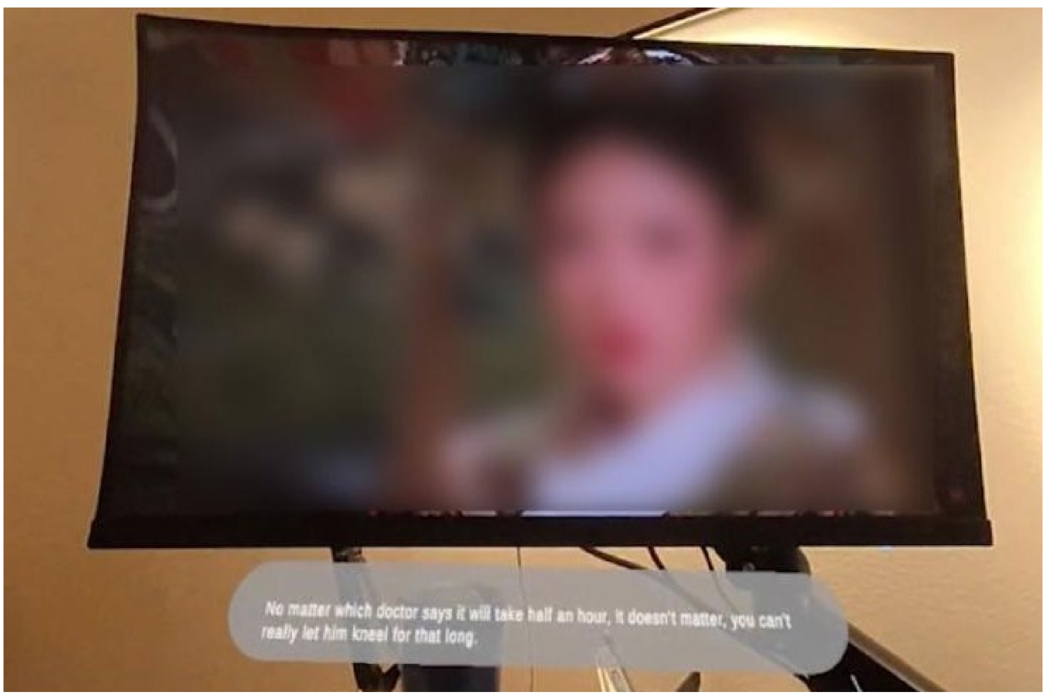


Figure 39. Screenshot of real-time translation features of the 'Speech Service' function. This screenshot image depicts the AR application's real-time translation from Chinese to English. The foreign drama has been blurred to respect copyright restrictions.

Scenario 3: Listening to Podcasts- Use case of 'Real-Time Speech-to-Text' and 'Real-Time Speaker-Identification'

In this scenario, a DHH user is listening to a podcast (see Figure 40). The AR application's 'Real-Time Speech-to-Text' feature transcribes the podcast into text. The 'Real-

Time Speaker-Identification' feature distinguishes between speakers, labeling their transcribed speech. This allows the DHH user to follow the podcast content and know who is speaking at any given time.

Figure 40. Screenshot of 'real-time speech-to-text and speaker identification features. The screenshot depicts the AR application's real-time Speech-to-Text and Speaker Identification. The podcast's cover image has been blurred to respect copyright restrictions.

System Architecture and Workflow of Final Prototype

The architecture of the Final Prototype blends the functionalities of home management and speech services into an intuitive AR application (refer to Figure 41).

First, for home device management, the system uses an MQTT Broker—a publish-subscribe-based messaging protocol. The AR application subscribes to the broker, receiving status updates and sending control commands to manage devices within the home environment. Simultaneously, the web browser in the AR application directs users to the page streaming video footage and transmits environmental sound text to the user.

Second, speech services are managed through Microsoft Azure's API, which both receives audio from the user and sends transcribed or translated text back to the AR application in real-time. This ensures users receive live captioning and speaker identification during their interactions.

The application's coherent workflow includes gesture recognition for menu navigation, enhancing the user experience by making it more hands-on and interactive. This system design prioritizes real-time information processing for DHH individuals and enriches their interaction with the home environment, offering control and awareness through a single, cohesive AR platform.

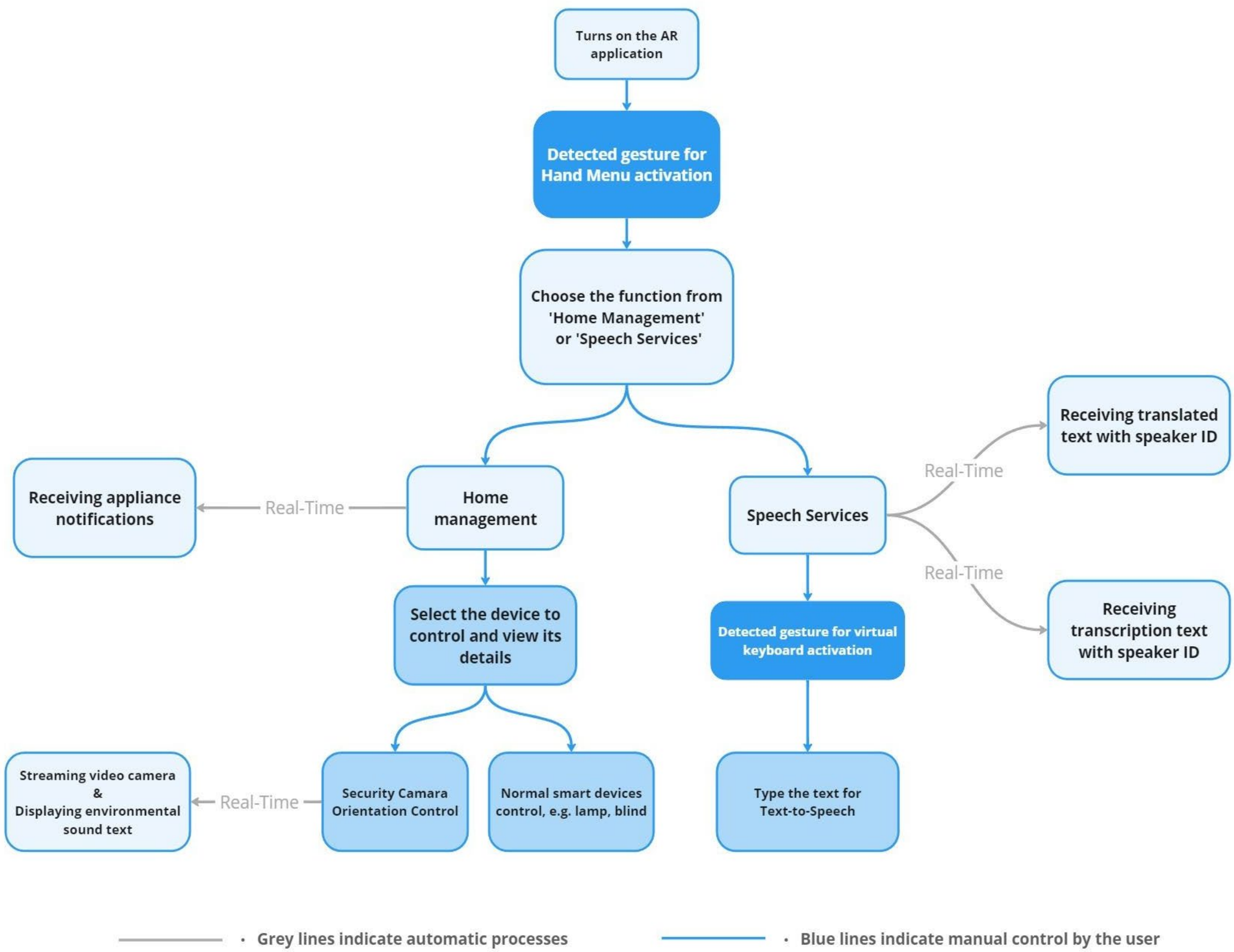


Figure 41. Final Prototype System architecture diagram.

Simulate Scenarios - A Day in the Life: The Impact of the Final Prototype on DHH Users

This section provides a glimpse into the daily life of a DHH individual, highlighting how the Final Prototype influences their home living experience. Through simulated scenarios, the effectiveness of the Final Prototype is illustrated:

Starting the Day: Upon waking, the user puts on their HMD to access the AR application. Through intuitive gestures within the app, the user turns on the coffee machine directly from the AR interface. As the coffee brews and becomes ready, the coffee machine model in the kitchen brightens with vibrant colors and displays an animation on the application, signaling the coffee is prepared.

Morning Routine: Energized by the coffee, the user moves on to prepare a more substantial breakfast. Utilizing the AR interface, they can now control other kitchen appliances with ease. For instance, adjusting the settings of the kitchen lights or preheating the oven is as simple as selecting their models in the virtual space. The toaster model glows softly to visually signify the toast is ready, providing a clear cue without relying on sound. This seamless integration of technology enhances the user's daily routine.

Managing Home Devices: Throughout the day, the user relies on the prototype for home management tasks. The AR interface alerts them when laundry is done, based on power consumption changes monitored through the smart plug. An outline and color change on the appliance model within their AR view signals the completion, complemented by a textual notification.

Social Interactions: In the evening, the user invites friends over. Laughter and chatter fill the room. Thanks to the AR application's live captioning, every word spoken by friends is transcribed in real-time and displayed within the user's field of vision. This allows the user to follow along and respond without missing a beat, fully participating in the lively exchange.

Wanting to share a story, the user activates the virtual keyboard with a simple gesture. They type their story onto the AR interface, and upon pressing 'send,' their typed message is converted into speech, and broadcast to the room. The seamless transition from text to voice ensures the user's story is heard and appreciated by everyone, fostering a truly inclusive environment. As the night deepens into a discussion on various topics, the AR application distinguishes between speakers, labeling them with customizable tags in the user's view. This feature allows the user to not only follow who is speaking but also understand the flow of the conversation, enhancing the clarity of group interactions.

Evening Wind Down: As evening turns to night, the user engages the AR application to ensure their home is secure before bedtime. A quick glance through the AR interface reveals the status of doors and windows, all monitored by Zigbee sensors, confirming that the home is safely locked down. This efficient and intuitive verification process brings a comforting sense of security, allowing the user to end their day with peace of mind.

Evening Entertainment: As the user settles into their evening routine, they decide to unwind by watching a drama on their favorite streaming platform. They find an intriguing foreign drama but realize it doesn’t have English subtitles. Undeterred, they activate the speech service function on their AR application. The application listens to the dialogue, translates it in real-time, and displays English subtitles directly on its AR interface. This allows the user to immerse themselves in the drama without language barriers, enhancing their viewing experience and providing a relaxing end to their day.

Midnight Security Check: Later, in the stillness of midnight, a sense of restlessness prompts the user to conduct a security check. With their HMD at the ready, they activate the AR application to view live feeds from the security cameras stationed around the perimeter—the backyard, front door, and other critical areas. The detailed AR overlays not only show a live visual but, utilizing the advanced features of the application, confirm the absence of unusual activity. Reassured by the comprehensive surveillance at their fingertips, the user feels a renewed sense of tranquility, facilitating a return to restful sleep, secure in the knowledge that their home remains safe.

Checklist Table for Final Prototype

The table below provides a comprehensive checklist that aligns the final prototype with the limitations identified in related works and the needs discovered from both the literature review and online survey.

Needs or Limitations from Literature Review and Online Survey	Final Prototype	If the Need or Limitation being Addressed	Notes
Auditory to Visual Conversion Mechanism	Visual representation of sounds by color and animation	***YES***	Transforms appliances' sound alerts into visual notifications for better awareness.
Communication Accessibility Features	Automatic real-time speech services	***YES***	Displays real-time, text-based transcriptions for various scenarios.
Environmental Awareness	YAMNet Environmental sound classification	***YES***	Integration of text-based real-time environmental sound classification into the surveillance system.
Customizable Alert Settings	Notification system categorizes events as urgent and non-urgent	***PARTIAL***	The design of notification system lacks depth in personalization for settings and modes.

Sign Language Support	No feature supports sign language	***NO***	Not included in this study.
Multisensory Feedback Mechanisms	Gesture and Visual-based system	***PARTIAL***	No haptic feedback; only visual and auditory cues were provided.
Advancing Home Accessibility Through Technology	Various home device control via an AR application	***YES***	Provides a gesture-based device control system for the entire home, eliminating the need for voice commands.
Intuitive Interaction Technology	3D User Interface design to reduce cognitive load	***YES***	Facilitates direct and intuitive user interface of home device control via a representative 3D home model.
Text-Heavy User Interface	User interface based on visual design	***PARTIAL***	Although most parts of the system design focus on visual feedback, the speech service primarily relies on text-based interaction.

Table 10. Final Prototype Checklist. A checklist showing how the final prototype addresses the identified needs and limitations from the literature review and online survey.

Reflection on Final Prototypes

The journey to creating the Final Prototype has been both challenging and immensely rewarding. As I navigate through the integration of AR, AI, and IoT technologies to build an AR application within an HMD platform, my goal is always clear: to enrich the home living experience of DHH individuals. This journey, driven by a series of iterative developments and the exploration of specific sub-questions, has culminated in a solution that I believe the final prototype resonates with the needs of the DHH community.

Reflecting on the process, the decision to innovate within the realm of 3D model user interface design stands out as a pivotal moment. Transitioning from a high-detail model to a low-poly version not only solved the computational challenge but also underscored the importance of adaptability and creative problem-solving in technology design. Similarly, the implementation of Zigbee and MQTT for home device management represents a significant advancement, moving away from unreliable sound-based alert detection to a more dependable, direct monitoring system, and successfully broadening the device range in pushing the boundaries of technology to better serve the needs of our users.

Maintaining the speech services from Prototype 4, underscored AR applications' capacity to dismantle communication barriers. This continuity in features reinforces their role in fostering accessible and inclusive social interactions for DHH individuals, demonstrating the enduring value of these services in the prototype's evolution.

However, the prototype has not yet fully addressed all limitations identified during the needs-finding phase. A pivotal area for enhancement is the development of a sophisticated filter system, which would enable users to customize notification preferences and select specific sound types for alerts. This feature aims to align the system more closely with individual needs and preferences, thereby enhancing user experience. In conjunction, the reliance on text-based notifications for environmental sounds emerges as another potential issue, as it could inadvertently increase cognitive load for users. This observation suggests an imperative need to explore alternative, more intuitive notification methods to alleviate potential strain.

Furthermore, privacy concerns associated with leveraging external services, such as Microsoft Azure's API, have been acknowledged. These concerns illuminate a broader research avenue focusing on data security and bolstering user trust, which is crucial for the widespread acceptance and success of the technology.

Compounding these limitations is the current design of the HMD. Its significant size and the degree to which it obscures the user's face could potentially undermine the prototype's objective of fostering social inclusion. This issue is predicated on the assumption that future iterations of HMD technology will be lighter and less obtrusive, thus minimizing any negative impact on social interactions and furthering the prototype's goal of enhancing the living experience for DHH individuals.

Subjective Evaluation of Final Prototype

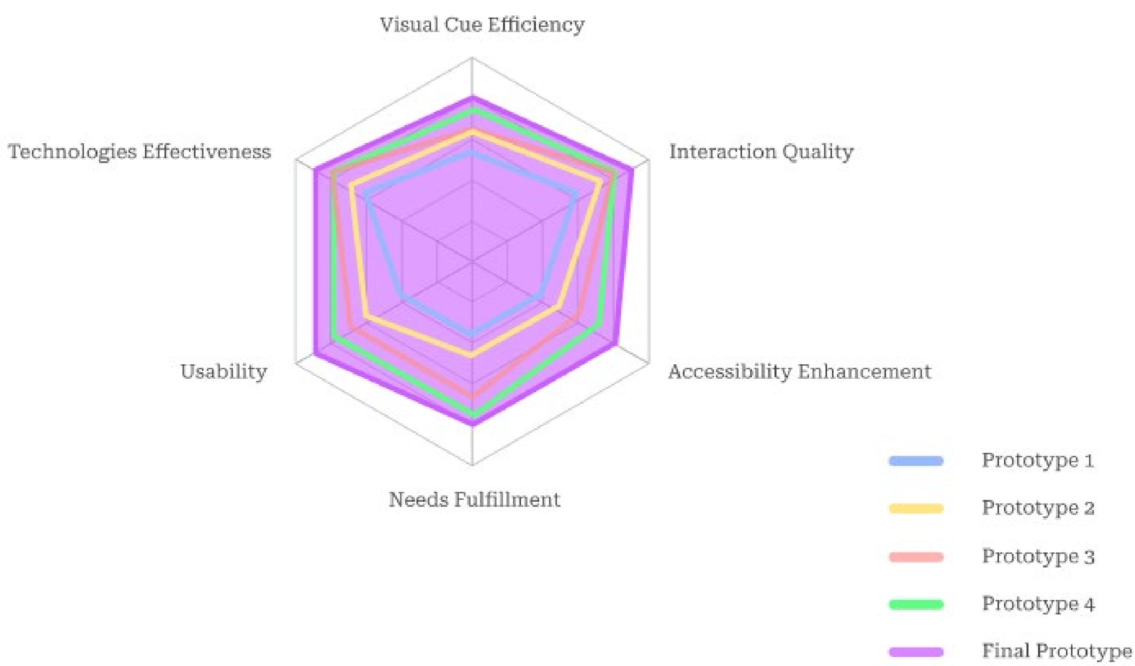


Figure 42. Radar chart for Final Prototype.

Visual Cue Efficiency (4/5): The prototype excelled in translating auditory information into visual formats, with users particularly valuing the intuitive presentation of auditory alerts and environmental sounds. The visual notifications for appliance and device status are anticipated to be notable enhancements, potentially increasing user satisfaction with the visual interface.

Interaction Quality (4.5/5): The ease of use and natural interaction afforded by gesture-based controls and a 3D home model are likely to be perceived as high points of the prototype. The direct control over home devices and the fluid integration of speech services within the AR environment could be seen as significant steps towards a user-friendly experience.

Accessibility Enhancement (4/5): Beyond speech services, the prototype endeavors to improve overall environmental accessibility through AR visualizations and alerts for DHH individuals. The incorporation of visual cues for environmental awareness and device status, combined with speech services, is likely to be beneficial in supporting clear communication and accessibility in daily living.

Needs Fulfillment (4/5): The prototype has been acknowledged for its effectiveness in meeting the crucial needs of DHH users, substantially enriching their home living experience. The 'Home Management' and 'Speech Services' components are specifically engineered to offer seamless and intuitive interaction with their physical environment and facilitate social inclusion. However, there remains an opportunity for further enhancement through the development of a more personalized system to cater to individual preferences and requirements more closely.

Technologies Effectiveness (4.5/5): The effective use of AR, AI, and IoT technologies in the prototype is likely to be regarded as a key strength, enhancing user navigation and interaction within their environment. The integration of these technologies is aimed at providing an intuitive and cohesive user experience.

Usability (4.5/5): The prototype is expected to score highly in usability based on its design, which is intended to be easy to learn and use. The overall experience is anticipated to be positive, with potential suggestions for improvement focusing on streamlining the initial setup process for users, such as Zigbee devices connection, personalized room model setup, etc.

6. Discussion & Future Work

6.1 Reflection of the Research

In this research, my main focus is on augmenting the home living experience for DHH individuals by combining AR, AI, and IoT technologies. The final outcome, an AR application, stands as a concrete representation of this integration. The design choices throughout this journey are carefully tailored to explore intuitive interaction designs and visual representations of auditory cues with a light cognitive load. These choices are driven by the dual objectives of enhancing interaction and increasing accessibility in home living experiences for the DHH community.

The final prototype boasts features that successfully address key needs and limitations identified through literature reviews and the online survey. It transforms auditory cues into vibrant visual formats, integrates environmental sound detection and classification, and facilitates direct interaction through a 3D user interface and a representative 3D home model, all of which reduce cognitive load and aim at enhancing accessibility. The implementation of real-time speech services also marks a leap forward in communication accessibility, providing text-based transcriptions for various scenarios.

However, the journey has revealed areas for growth. Customizable alert settings and sign language support were not fully realized within the scope of this project, pointing to future enhancements. The partial reliance on text-heavy interfaces also indicates an opportunity to further refine the balance between visual and textual information to support varied user preferences.

The insights gained from this thesis are set to steer future development, with a steadfast focus on incorporating feedback from DHH individuals into the application's testing and evolution. Aiming for a more comprehensive user experience, the endeavor is to craft a solution that resonates with their specific needs and benefits the whole DHH community.

6.2 Future Work

Engaging DHH Participants in Development: Future work should include closer collaboration with DHH participants for feedback and testing. This direct engagement aims to ensure the application meets real user needs and preferences, refining its effectiveness and usability.

Advanced Customization and User Personalization: Future developments should focus on creating more sophisticated systems for user customization and personalization. This includes enhancing the prototype to allow DHH users to tailor notification settings, choose specific alert types, and define their own categories of environmental sounds for more personalized interaction experiences.

Sophisticated Notification Systems: Develop a more advanced filter system that allows users to customize their notification preferences. This system would enable users to select specific sound types for alerts, ensuring that notifications are both relevant and tailored to individual preferences.

Environmental Sound Interpretation: Enhance the environmental sound recognition feature to move beyond text-based notifications. Future iterations could explore more intuitive and less cognitively demanding methods of notification, such as using symbolic or pictorial representations.

Integration of Additional Devices and Sensors: To further enrich the home environment for DHH individuals, integrating a wider range of smart home devices and sensors is essential. Future work could explore the inclusion of more diverse types of sensors, such as temperature, humidity, and air quality sensors, and how these can be seamlessly integrated into the AR interface for a more comprehensive home management system.

Enhanced Speech Services: While the prototype already incorporates speech-to-text and text-to-speech services, there is room for improvement in speech recognition accuracy, especially in noisy environments or group discussions. Future iterations could explore advanced noise cancellation techniques and more sophisticated AI models for speaker identification and separation.

Privacy and Security Considerations: Given the prototype's reliance on external services like Microsoft Azure's API for speech services, future research should investigate the privacy and security implications of such dependencies. Developing end-to-end encrypted communication channels and exploring decentralized approaches to data processing could address privacy concerns while maintaining service efficacy.

User-Centered Design Process: Continue to employ and refine the user-centered design process, involving DHH individuals more closely in the development and testing phases. This approach can ensure that the prototype not only meets technical specifications but also resonates with the needs and preferences of the intended user base.

Addressing Unmet Needs and Limitations: Finally, future work must continue to identify and address unmet needs and limitations discovered through this research. This includes exploring solutions for sophisticated environmental sound interpretation to reduce cognitive load and developing more intuitive notification methods beyond text-based alerts.

Cost Considerations and Scalability: Addressing the financial implications of using external services and hardware is critical for the prototype's scalability and accessibility to a broader audience. Future work could include finding cost-effective alternatives or developing a business model that subsidizes costs for end-users.

Social Inclusion and HMD Design Considerations: Despite the strengths demonstrated by the speech service, the current design of the HMD may inadvertently foster a sense of social isolation due to its scale and opaque design. Future iterations should consider the social implications of HMD design, exploring ways to make the device more discreet and transparent. The goal is to reduce any perceived barriers to social interaction, ensuring that the technology not only serves functional purposes but also supports broader social inclusion objectives.

7. Conclusion

This research was sparked by witnessing my grandparents' struggle with age-related hearing loss, shedding light on the DHH community often marginalized by the technological shift towards voice-command systems. Such systems, inherently excluding those who are DHH, underscored the urgent need for discussions on technological inclusivity.

The research process involves a detailed examination of existing technological gaps, identifying unmet needs through literature reviews and online surveys. The foundation of the project facilitates a two-stage prototype iteration. The first stage aligns with existing works, employing dual-domain AR, AI, and IoT technologies to establish a functional base. The second stage, an innovative exploration, merges all three technologies to address the identified challenges and insights gained from earlier prototypes. This integration sets the stage for innovative solutions, steering the project towards its final design.

While initially focused on elevating the domestic quality of life for individuals within the DHH community, the project's deeper objective transcends this boundary, aiming to forge a more accessible environment that enriches their overall life experiences. Acknowledging the varied and nuanced needs of the DHH community, this initiative does not pursue a universal solution. Instead, it aspires to establish a comprehensive foundation, weaving together AR, AI, and IoT technologies. This foundational work is designed not only to address immediate accessibility challenges but also to lay the groundwork for future endeavors in crafting a more inclusive world.

7.1 Revisiting Goals and Objectives

Reflecting on the initial objectives, the development of the AR application is rooted in enhancing the interaction and accessibility of DHH individuals with their home environment through intuitive visual cues. This goal has been successfully met; the final prototype features an intuitive, user-friendly interface that streamlines access to device control and promotes an accessible home experience. The incorporation of ASR AI models, with the help of Microsoft Azure speech services, aims to improve social integration. This goal is realized through the prototype's capabilities, fostering inclusivity by allowing DHH individuals to engage more fully in social interactions. The implementation of visual alerts for safety and convenience effectively compensates for the absence of auditory cues, directly aligning with the core aim to enhance the living experience of DHH individuals within their homes.

7.2 Outcomes and Contributions

The key contributions of this project lie in the successful integration of AR, AI, and IoT technologies to create an AR application that offers comprehensive solutions to the everyday challenges faced by the DHH community. The prototype not only facilitates communication and interaction with the environment but also stands as a testament to the potential of Mixed Reality in enhancing the quality of life for DHH individuals. It represents a significant step forward in making home living spaces more accessible and interactive, setting a precedent for future technological interventions in the realm of assistive devices.

7.3 Research Limitations

Despite the advancements, this research encounters limitations that provide avenues for future exploration. The absence of user study evaluations due to time constraints is a gap that future work must address to validate and refine the prototype's effectiveness. Budgetary limitations restrict access to a broader range of technological devices, which could further enrich the prototype's capabilities. Privacy, security, and safety considerations remain paramount and require ongoing attention to safeguard users in the ever-evolving landscape of home technology.

7.4 Future Development

In future enhancements of the AR application for DHH individuals, paramount importance will be placed on involving DHH users in prototype testing. This direct engagement is essential for fine-tuning the application's functionality to better meet their needs and preferences. The focus will be on refining customization features based on user feedback, advancing notification systems for personalized alerts, and exploring intuitive methods for environmental sound interpretation. Efforts will also extend to integrating a broader range of smart devices and improving speech service accuracy. Privacy, security, and the social inclusiveness of the design, especially regarding head-mounted displays, will be carefully considered. Central to this iterative development process is a user-centered approach, ensuring the technology truly enhances the living spaces of DHH individuals by making them more accessible and interactive.

7.5 Final Remarks

This project sets out to integrate AR, AI, and IoT technologies to enhance the home environment for DHH individuals. The development of the AR application represents a significant step forward in making technology more inclusive. The prototype showcases how Mixed Reality can be utilized to improve accessibility, demonstrating the practical application of these technologies to meet the needs of the DHH community.

Despite challenges such as the lack of user testing and the need for ongoing enhancements in privacy and security, the work has laid a solid foundation for future improvements. The next steps involve deeper community engagement and technological refinement based on DHH feedback to ensure the solution is both effective and user-friendly. The lessons learned throughout this research process underscore the importance of dedication, creative thinking, and user-centric design. The ultimate goal is to continue evolving this application into a tool that not only facilitates communication and interaction for the DHH community but also contributes to creating a more accessible and inclusive world.

Bibliography

IBM Data and AI Team. (2023, July 6). *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?* IBM. https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/

Adi, E., Anwar, A., Baig, Z., & Zeadally, S. (2020, June 30). *Machine learning and data analytics for the IoT*. ArXiv. https://doi.org/10.48550/arXiv.2007.04093

Alahi, M. E. E., Sukkuea, A., Tina, F. W., Nag, A., Kurdthongmee, W., Suwannarat, K., & Mukhopadhyay, S. C. (2023). Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends. *Sensors*, *23*(11). https://doi.org/10.3390/s23115206

Almutairi, A., & Al-Megren, S. (2017). *Augmented Reality for the Literacy Development of Deaf Children*. https://doi.org/10.1145/3132525.3134789

Anderson, N. S., Norman, D. A., & Draper, S. W. (1988). User Centered System Design: New Perspectives on Human-Computer Interaction. *The American Journal of Psychology*, *101*(1), 148. https://doi.org/10.2307/1422802

Asakura, T. (2023). Augmented-Reality Presentation of Household Sounds for Deaf and Hard-of-Hearing People. *Sensors*, *23*(17). https://doi.org/10.3390/s23177616

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Bleakley, A., Rough, D., Roper, A., Lindsay, S., Porcheron, M., Lee, M., Nicholson, S. A., Cowan, B. R., & Clark, L. (2022). Exploring Smart Speaker User Experience for People Who Stammer. *The 24th International ACM SIGACCESS Conference on Computers and Accessibility*. https://doi.org/10.1145/3517428.3544823

Bragg, D., Huynh, N., & Ladner, R. E. (2016). *A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users*. https://doi.org/10.1145/2982142.2982171

Brown, T. (2008, June). *Design Thinking*. Harvard Business Review. https://hbr.org/2008/06/design-thinking

Chaudhry, M. A., & Kazim, E. (2021). Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. *AI and Ethics*, *2*(2), 157–165. https://doi.org/10.1007/s43681-021-00074-z

Cheng, J., & Kunz, T. (2009). *A Survey on Smart Home Networking*.

Chin, J., Callaghan, V., & Allouch, S. B. (2019). The Internet-of-Things: Reflections on the past, present and future from a user-centered and smart environment perspective. *Journal of Ambient Intelligence and Smart Environments*, *11*(1), 45–69. https://doi.org/10.3233/ais-180506

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The Past, Present, and Future of Virtual and Augmented Reality Research: a Network and Cluster Analysis of the Literature. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02086

Cosio, L. D., Buruk, O. 'Oz', Fernández Galeote, D., Bosman, I. D. V., & Hamari, J. (2023). *Virtual and Augmented Reality for Environmental Sustainability: A Systematic Review*. https://doi.org/10.1145/3544548.3581147

Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., & Williams, I. (2023). *Inclusive AR/VR: accessibility barriers for immersive technologies*. https://doi.org/10.1007/s10209-023-00969-0

Després, C. (1991). THE MEANING OF HOME: LITERATURE REVIEW AND DIRECTIONS FOR FUTURE RESEARCH AND THEORETICAL DEVELOPMENT.

Journal of Architectural and Planning Research, *8*(2), 96–115. https://www.jstor.org/stable/43029026

Dokhnyak, B., & Vysotska, V. (2021). Intelligent Smart Home System Using Amazon Alexa Tools. In *Semantic Scholar*. https://www.semanticscholar.org/paper/Intelligent-Smart-Home-System-Using-Amazon-Alexa-Dokhnyak-Vysotska/41df199384089c1f9bc9edcdfd30fafc1fe70c4f

Fang, W., Zhang, T., Chen, L., & Hu, H. (2023). A survey on HoloLens AR in support of human-centric intelligent manufacturing. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-023-02247-5

Findlater, L., Chinh, B., Jain, D., Froehlich, J., Kushalnagar, R., Lin, A., & Carey, A. (2019). *Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies*. https://doi.org/10.1145/3290605.3300276

Gao, J., Hu, H., & Yao, L. (2020). The role of social engagement in the association of self-reported hearing loss and health-related quality of life. *BMC Geriatrics*, *20*(1). https://doi.org/10.1186/s12877-020-01581-0

Gaver, W. (2012). *What should we expect from research through design?* https://doi.org/10.1145/2207676.2208538

Ghorbani, F., Ahmadi, A., Kia, M., Rahman, Q., & Delrobaei, M. (2023). A Decision-Aware Ambient Assisted Living System with IoT Embedded Device for In-Home Monitoring of Older Adults. *Sensors*, *23*(5). https://doi.org/10.3390/s23052673

Ghule, S., & Chavaan, M. (2021). *RETRACTED CHAPTER: Implementation of Hand Gesture Recognition System to Aid Deaf-Dumb People*. 183–194. https://doi.org/10.1007/978-981-15-8391-9_14

Giaccardi, E. (2019). *Histories and Futures of Research through Design: From Prototypes to Connected Things*. International Journal of Dsign. https://www.ijdesign.org/index.php/IJDesign/article/view/3192/875

Goodman, S. M., Liu, P., Jain, D., McDonnell, E. J., Froehlich, J. E., & Findlater, L. (2021). *Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing*. 1–23. https://doi.org/10.1145/3463501

Guo, R., Yang, Y., Kuang, J., Bin, X., Jain, D., Goodman, S., Findlater, L., & Froehlich, J. (2020). *HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-mounted Display*. https://doi.org/10.1145/3373625.3418031

Heine, C., & Browning, C. J. (2009). Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions. *Disability and Rehabilitation*, *24*(15), 763–773. https://doi.org/10.1080/09638280210129162

IBM. (n.d.-a). *What Is Artificial Intelligence (AI)?* IBM. Retrieved March 2, 2024, from https://www.ibm.com/topics/artificial-intelligence

IBM. (n.d.-b). *What is the internet of things?* IBM. Retrieved March 5, 2024, from https://www.ibm.com/topics/internet-of-things

Ifrim, A., Moldoveanu, F., Moldoveanu, A., & Grădinaru, A. (2021). *LibrARy – Enriching the Cultural Physical Spaces with Collaborative AR Content*. 10.1007/978-3-030-77599-5_43

Inomata, S., Iwase, K., Komiya, K., & Nakajima, T. (2020, February 23). *AR Smart Home: a Smart Appliance Controller Using Augmented Reality Technology and a Gesture Recognizer*. https://personales.upv.es/thinkmind/MMEDIA/MMEDIA_2020/mmedia_2020_1_10_50014.html

Intille, S. S. (2002). *Designing a home of the future*. IEEE Xplore. https://ieeexplore.ieee.org/document/1012340

Iswahyudi, S. R., Ediyanto, Efendi, M., Sunandar, A., Indreswari, H., & Rahman, D. H. (2021, November 26). *Augmented Reality Technology as a Learning Media to Help Parents in the Academic Functional Education of Children with Hearing Impairments*. Atlantis Press. https://doi.org/10.2991/assehr.k.211126.060

Jain, D., Findlater, L., Gilkeson, J., Holland, B., Duraiswami, R., Zotkin, D., Vogler, C., & Froehlich, J. E. (2015). *Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing*. https://doi.org/10.1145/2702123.2702393

Jain, D., Lin, A., Guttman, R., Amalachandran, M., Zeng, A., Findlater, L., & Froehlich, J. (2019). *Exploring Sound Awareness in the Home for People who are Deaf or Hard of Hearing*. https://doi.org/10.1145/3290605.3300324

Jain, D., Mack, K., Amrous, A., Wright, M., Goodman, S., Findlater, L., & Froehlich, J. E. (2020). *HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users*. 1–12. https://doi.org/10.1145/3313831.3376758

Jain, K. (2015). *Design Thinking in Research*. Princeton Correspondents on Undergraduate Research. https://pcur.princeton.edu/2015/10/design-thinking-in-research/

Kafle, S., Glasser, A., Al-khazraji, S., Berke, L., Seita, M., & Huenerfauth, M. (2020). *Artificial intelligence fairness in the context of accessibility research on intelligent systems for people who are deaf or hard of hearing*. https://doi.org/10.1145/3386296.3386300

Kerdvibulvech, C. (2019). *A Review of Augmented Reality-Based Human-Computer Interaction Applications of Gesture-Based Interaction*. https://doi.org/10.1007/978-3-030-30033-3_18

Kopytko, V., Shevchuk, L., Yankovska, L., Semchuk, Z., & Strilchuk, R. (2018). Smart Home and Artificial Intelligence as Environment for the Implementation of New Technologies. *Path of Science*, *4*(9), 2007–2012. https://doi.org/10.22178/pos.38-2

Liberatore, M. J., & Wagner, W. P. (2021). *Virtual, mixed, and augmented reality: a systematic review for immersive systems research*. https://doi.org/10.1007/s10055-020-00492-0

Liu, S., Zhou, Z., Du, J., Shangguan, L., Jun, H., & Xin, W. (2017). *UbiEar: Bringing Location-independent Sound Awareness to the Hard-of-hearing People with Smartphones*. 1–21. https://doi.org/10.1145/3090082

Mahroo, A., Greci, L., & Sacco, M. (2019). HoloHome: An Augmented Reality Framework to Manage the Smart Home. *Lecture Notes in Computer Science*, 137–145. https://doi.org/10.1007/978-3-030-25999-0_12

Maiorana-Basas, M., & Pagliaro, C. M. (2014). Technology Use Among Adults Who Are Deaf and Hard of Hearing: A National Survey. *Journal of Deaf Studies and Deaf Education*, *19*(3), 400–410. https://doi.org/10.1093/deafed/enu005

Matthews, T., Fong, J., Ho-Ching, F. W.-L., & Mankoff, J. (2006). Evaluating non-speech sound visualizations for the deaf. *Behaviour & Information Technology*, *25*(4), 333–351. https://doi.org/10.1080/01449290600636488

Mielke, M., & Brueck, R. (2015, November 5). *Design and evaluation of a smartphone application for non-speech sound awareness for people with hearing loss*. IEEE. https://ieeexplore.ieee.org/document/7319516

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *Engineering, Computer Science*. https://www.semanticscholar.org/paper/A-Taxonomy-of-Mixed-Reality-Visual-Displays-Milgram-Kishino/f78a31be8874eda176a5244c645289be9f1d4317

Milman, Naralie. B. (2018). Defining and Conceptualizing Mixed Reality, Augmented Reality, and Virtual Reality. *Distance Learning*, *15*(2), 55–58.

Minaee, S., Liang, X., & Yan, S. (2022). Modern Augmented Reality: Applications, Trends, and Future Directions. *ArXiv*. https://doi.org/10.48550/arXiv.2202.09450

Mirzaei, M., Kán, P., & Kaufmann, H. (2020, February 13). *EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People*. IEEE. https://ieeexplore.ieee.org/document/8998298

Nakao, Y., & Sugano, Y. (2020). *Use of Machine Learning by Non-Expert DHH People: Technological Understanding and Sound Perception*. https://doi.org/10.1145/3419249.3420157

Nguyen, L. T., Schicktanz, F., Stankowski, A., & Avramidis, E. (2021, June 30). *Evaluating the translation of speech to virtually-performed sign language on AR glasses | IEEE Conference Publication | IEEE Xplore*. IEEE. https://ieeexplore.ieee.org/abstract/document/9465430

NIH. (2021). *Quick Statistics About Hearing, Balance, & Dizziness*. NIH. https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

Nimmalapudi, V. V., Hashmi, M. F., & Keskar, A. G. (2022). *Machine Learning and IoT-Based Messaging Device for Blind, Deaf, and Dumb People*. https://doi.org/10.1007/978-981-19-0019-8_15

Parekh, P., Patel, S., Patel, N., & Shah, M. (2020). Systematic review and meta-analysis of augmented reality in medicine, retail, and games. *Visual Computing for Industry, Biomedicine, and Art*, *3*(1). https://doi.org/10.1186/s42492-020-00057-7

Peng, Y.-H., Tang, H.-H., Chen, M. Y., Hsi, M.-W., Taele, P., Lin, T.-Y., Lai, P.-E., Hsu, L., Chen, T., Wu, T.-Y., & Chen, Y.-A. (2018). *SpeechBubbles: Enhancing Captioning*

Experiences for Deaf and Hard-of-Hearing People in Group Conversations. https://doi.org/10.1145/3173574.3173867

Pieniak, M., Lachowicz-Tabaczek, K., Masalski, M., Hummel, T., & Oleszkiewicz, A. (2020). Self-rated sensory performance in profoundly deaf individuals. Do deaf people share the conviction about sensory compensation? *Journal of Sensory Studies*, *35*(4). https://doi.org/10.1111/joss.12572

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior*, *133*. https://doi.org/10.1016/j.chb.2022.107289

Ricquebourg, V., Menga, D., Durand, D., Marhic, B., Delahoche, L., & Loge, C. (2006). The Smart Home Concept : our immediate future. *2006 1ST IEEE International Conference on E-Learning in Industrial Electronics*. https://doi.org/10.1109/icelie.2006.347206

Rodolitz, J., Gambill, E., Willis, B., Vogler, C., & Kushalnagar, R. S. (2019). *Accessibility of Voice-Activated Agents for People who are Deaf or Hard of Hearing*. CSU. https://scholarworks.calstate.edu/concern/publications/qj72pb35f

Saifan, R. R., Dweik, W., & Abdel-Majeed, M. (2018). A machine learning based deaf assistance digital system. *Computer Applications in Engineering Education*, *26*(4), 1008–1019. https://doi.org/10.1002/cae.21952

Shi, H., Cao, G., Ma, G., Duan, J., Bai, J., & Meng, X. (2022). *New Progress in Artificial Intelligence Algorithm Research Based on Big Data Processing of IOT Systems on Intelligent Production Lines*. https://doi.org/10.1155/2022/3283165

Suzuki, R., Karim, A., Xia, T., Hedayati, H., & Marquardt, N. (2022). *Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces*. https://doi.org/10.1145/3491102.3517719

TensorFlow. (2023, October 6). *Audio classification*. TensorFlow. https://www.tensorflow.org/lite/examples/audio_classification/overview

Tigwell, G. W., Xu, J., Luna, S. M., & Papangelis, K. (2023). *Communication and Collaboration Among DHH People in a Co-located Collaborative Multiplayer AR Environment*. https://doi.org/10.1145/3597638.3614479

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June 12). *Attention Is All You Need*. ArXiv. https://arxiv.org/abs/1706.03762

Wald, M. (2021). AI Data-Driven Personalisation and Disability Inclusion. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.571955

Wang, X., Wan, Z., Hekmati, A., Zong, M., Alam, S., Zhang, M., & Krishnamachari, B. (2024, January 5). *IoT in the Era of Generative AI: Vision and Challenges*. ArXiv. https://doi.org/10.48550/arXiv.2401.01923

Wang, Z., Bai, X., Zhang, S., Billinghurst, M., He, W., Wang, P., Lan, W., Min, H., & Chen, Y. (2022). *A comprehensive review of augmented reality-based instruction in manual assembly, training and repair*. https://doi.org/10.1016/j.rcim.2022.102407

Wojtanowski, G., Gilmore, C., Seravalli, B., Fargas, K., Vogler, C., & Kushalnagar, R. S. (2020). *"Alexa, Can You See Me?" Making Individual Personal Assistants for the Home Accessible to Deaf Consumers*. CSU. https://scholarworks.calstate.edu/concern/publications/qf85nf68r

Yağanoğlu, M., & Köse, C. (2018). Real-Time Detection of Important Sounds with a Wearable Vibration Based Device for Hearing-Impaired People. *Electronics*, *7*(4). https://doi.org/10.3390/electronics7040050

Yeung, P., Alonzo, O., & Huenerfauth, M. (2020). Interest and Requirements for Sound-Awareness Technologies Among Deaf and Hard-of-Hearing Users of Assistive Listening Devices. *Universal Access in Human-Computer Interaction. Applications and Practice*, *12189*. https://doi.org/10.1007/978-3-030-49108-6_11

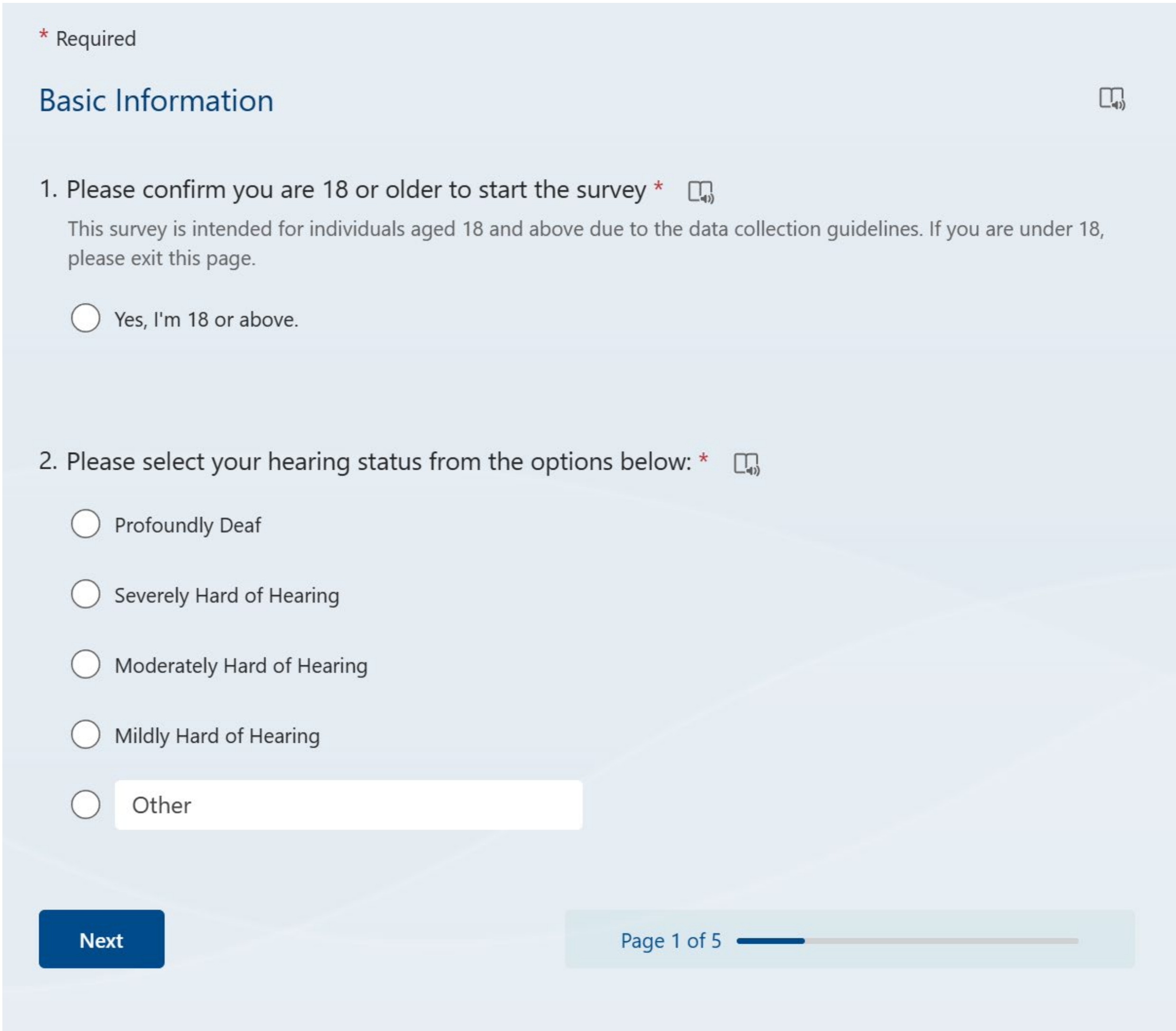
Žilak, M., Car, Ž., & Čuljak, I. (2022). A Systematic Literature Review of Handheld Augmented Reality Solutions for People with Disabilities. *Sensors* , *22*(20). https://doi.org/10.3390/s22207719

Zimmerman, J., & Forlizzi, J. (2008). The Role of Design Artifacts in Design Theory Construction. *Issues in Design Research and Design Research Methods*, *2*(1). https://doi.org/10.1080/17493460802276893

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). *An analysis and critique of Research through Design: towards a formalization of a research approach*. https://doi.org/10.1145/1858171.1858228

Appendices

Appendix A. Online Survey Questionnaire



* Required

Basic Information

1. Please confirm you are 18 or older to start the survey *

This survey is intended for individuals aged 18 and above due to the data collection guidelines. If you are under 18, please exit this page.

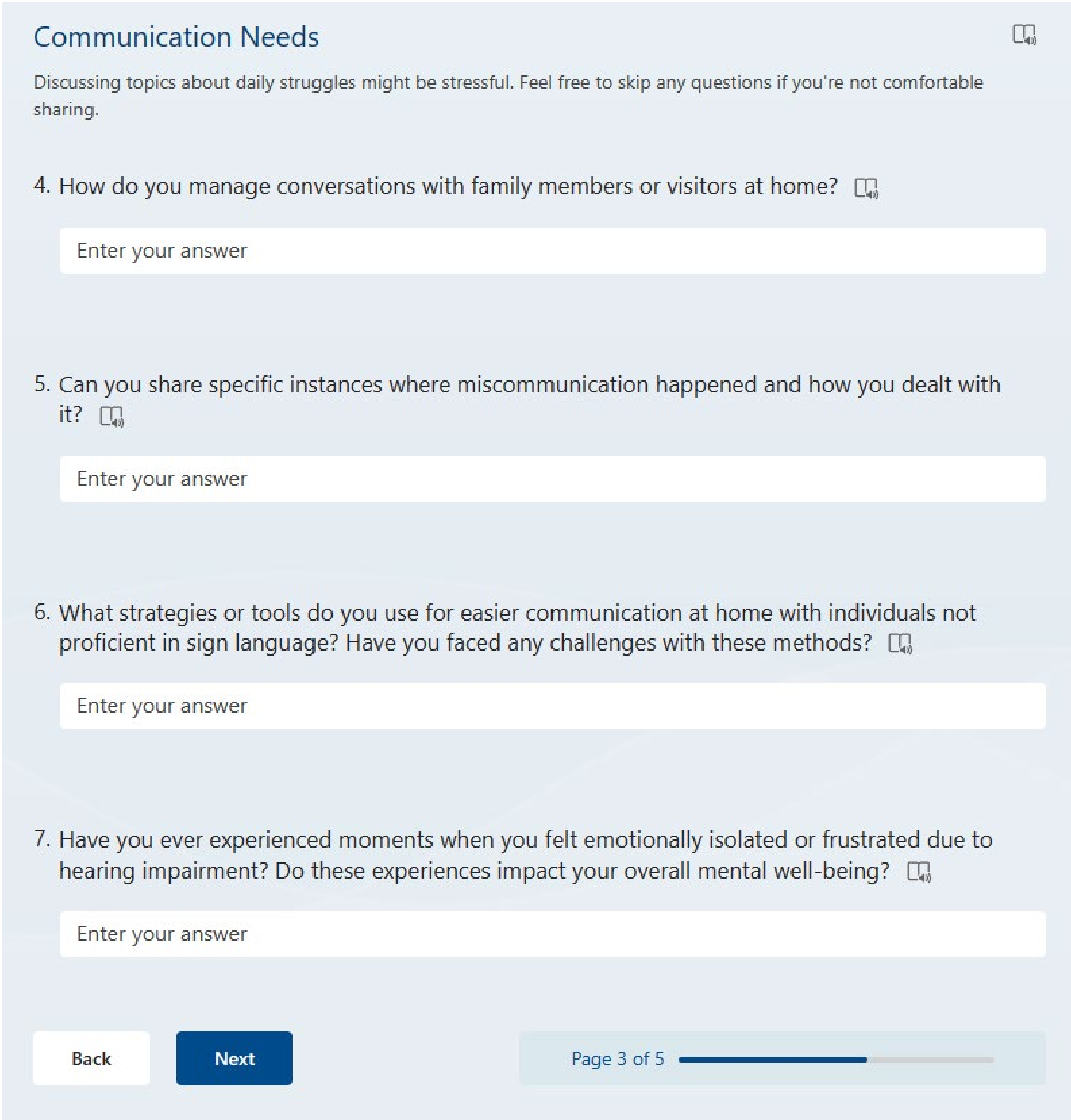
- Yes, I'm 18 or above.

2. Please select your hearing status from the options below: *

- Profoundly Deaf
- Severely Hard of Hearing
- Moderately Hard of Hearing
- Mildly Hard of Hearing
- Other

Next

Page 1 of 5



Communication Needs

Discussing topics about daily struggles might be stressful. Feel free to skip any questions if you're not comfortable sharing.

4. How do you manage conversations with family members or visitors at home?

Enter your answer

5. Can you share specific instances where miscommunication happened and how you dealt with it?

Enter your answer

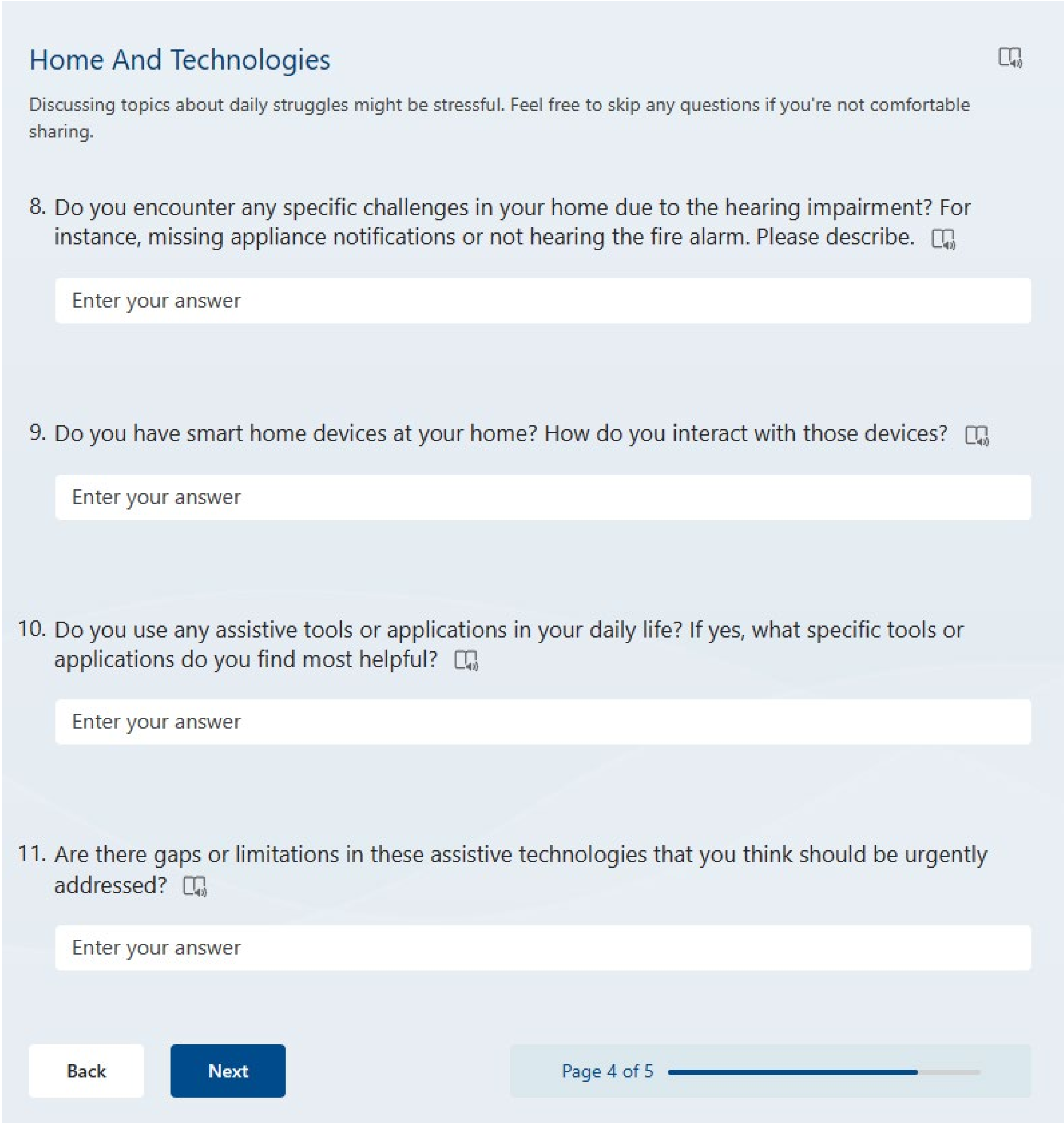
6. What strategies or tools do you use for easier communication at home with individuals not proficient in sign language? Have you faced any challenges with these methods?

Enter your answer

7. Have you ever experienced moments when you felt emotionally isolated or frustrated due to hearing impairment? Do these experiences impact your overall mental well-being?

Enter your answer

Back | Next | Page 3 of 5



Home And Technologies

Discussing topics about daily struggles might be stressful. Feel free to skip any questions if you're not comfortable sharing.

8. Do you encounter any specific challenges in your home due to the hearing impairment? For instance, missing appliance notifications or not hearing the fire alarm. Please describe.

Enter your answer

9. Do you have smart home devices at your home? How do you interact with those devices?

Enter your answer

10. Do you use any assistive tools or applications in your daily life? If yes, what specific tools or applications do you find most helpful?

Enter your answer

11. Are there gaps or limitations in these assistive technologies that you think should be urgently addressed?

Enter your answer

Back Next Page 4 of 5

Future Vision: Assistance Tools in Mixed Reality

Discussing topics about daily struggles might be stressful. Feel free to skip any questions if you're not comfortable sharing.

12. Have you ever seen or used Virtual Reality (VR) or Augmented Reality (AR) applications in use? If yes, how did you feel about these technologies? Were they easy to navigate and understand?

Enter your answer

13. Could you envision using or benefiting from VR or AR technology to assist with challenges you face at home, especially due to hearing impairment? How do you think these technologies could help?

Enter your answer

14. Based on your experience or awareness of VR or AR, what functionalities do you think would be beneficial to better accommodate your daily needs and challenges at home?

Enter your answer

Back | Submit | Page 5 of 5

Appendix B. Participant Invitation Poster & Link to Invitation for Participant Video

a. Sample of Participant Invitation Poster. *Note that all the personal information and contact details have been removed from the poster.*

b. Link to Invitation for Participant video: https://youtu.be/fDky4PHqKd8