# WordBond: An AI-Powered Interactive Tool to Facilitate Shared Reading Between Visually Impaired Parents and Sighted Children


Zhaopeng Zhu
Shanghai Jiao Tong University
Shanghai, China
Eindhoven University of Technology
Eindhoven, The Netherlands
zhuzhaopeng@sjtu.edu.cn

Helena Mossel
Eindhoven University of Technology
Eindhoven, The Netherlands
h.t.mossel@student.tue.nl

Jesse Strijker
Eindhoven University of Technology
Eindhoven, The Netherlands
j.j.strijker@student.tue.nl

Jun Hu*
Eindhoven University of Technology
Eindhoven, The Netherlands
j.hu@tue.nl


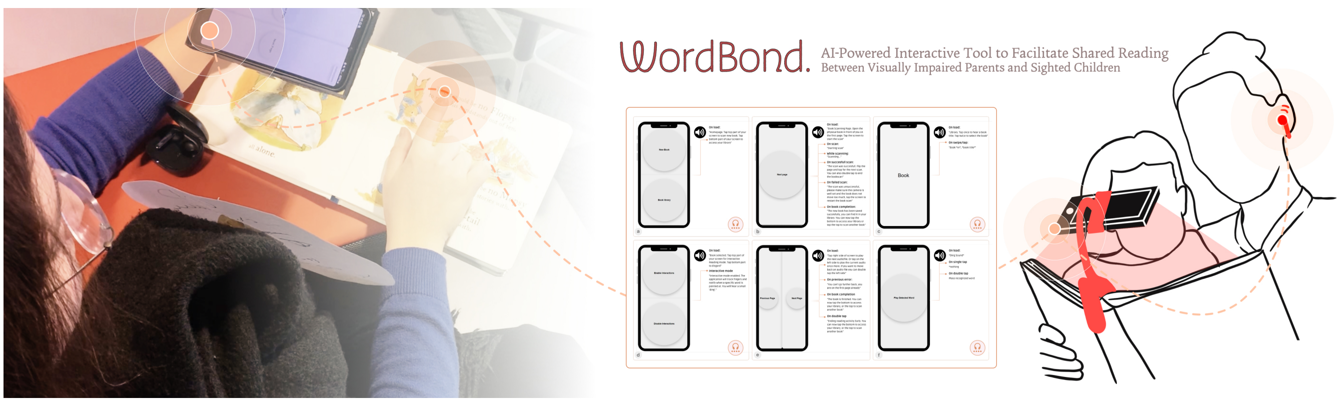


**Figure 1: WordBond Facilitates Shared Reading Between Visually Impaired Parents and Sighted Children.**


## Abstract

This paper introduces WordBond, an innovative AI-driven assistive tool designed to support visually impaired parents (VIPs) in shared reading with their sighted children. Existing assistive tools often fail to preserve the intimate parent-child voice bond, forcing VIPs to rely on impersonal synthetic voices. WordBond addresses this by leveraging a suite of technologies—including low-light optimized OCR (EasyOCR), an LLM-based correction module, and interactive hand-tracking—to detect content pointed to by the child and provide reading assistance in the parent's own voice. We present a functional prototype supported by an iterative user study (pilot and final evaluation, n=7 blindfolded participants simulating VIPs). This evaluation demonstrates the system's ability to facilitate prompt and personalized responses to user interaction, promoting natural and unscripted storytelling flow. WordBond's human-centered design approach showcases the potential to reduce cognitive load and enhance the quality of the shared reading experience, while highlighting critical areas for future usability optimization.



*Corresponding author




*ICHEC '25, Singapore*



ACM ISBN 978-1-4503-XXXX-X/2025/xx
https://doi.org/XXXXXXX.XXXXXXX



## CCS Concepts

• **Human-centered computing** → **Accessibility design and evaluation methods**.

## Keywords

Accessibility, Visually Impaired, Shared Reading Interaction

**ACM Reference Format:**
Zhaopeng Zhu, Helena Mossel, Jesse Strijker, and Jun Hu. 2025. WordBond: An AI-Powered Interactive Tool to Facilitate Shared Reading Between Visually Impaired Parents and Sighted Children. In *Proceedings of International Conference on Human-Engaged Computing (ICHEC '25)*. ACM, New York, NY, USA, 14 pages. https://doi.org/XXXXXXX.XXXXXXX

# 1 Introduction

Shared reading is a well-established critical practice that fosters early literacy, enriches vocabulary, and stimulates imagination in children. Beyond these cognitive and linguistic benefits, shared reading carries profound affective and social value, serving as a primary mechanism for strengthening parent-child bonds through shared experiences and emotional closeness [24]. Parental involvement in this activity promotes lasting engagement with books and nurtures deep relational connection [23].

However, these crucial developmental opportunities are not equally accessible to all families. *Visually impaired parents (VIPs)* face significant systemic barriers to engaging in spontaneous shared reading with their sighted children. These challenges include the limited availability of accessible tactile formats (e.g., braille), documented low braille literacy rates among the VIP population, and crucially the functional shortcomings of existing assistive technologies [28].

Existing solutions such as screen readers and digital audiobooks often prioritize *information access* over *relational intimacy*. Although effective for independent consumption, these tools rarely preserve the natural voice of the parents, which is widely considered central to the intimate, relational experience of reading aloud [26]. The reliance on impersonal synthetic voices fundamentally undermines the emotional connection and natural flow of storytelling that parents actively seek to maintain [9]. Furthermore, while current AI-powered accessibility tools, such as Microsoft's *Seeing AI* [16] for object recognition and *WeWALK* [5] for navigation, are highly effective in structured domains, they remain ill-suited for the dynamic low-light environment of a shared reading context, often struggling with varied fonts for children's books and fluctuating lighting conditions [31].

In response to this critical gap, this paper introduces ***WordBond***, an AI-powered interactive system designed to empower visually impaired parents to engage in personalized, real-time storytelling with their sighted children. WordBond integrates *optical character recognition (OCR)* optimized for poor lighting, a corrective *Large Language Model (LLM)*, and real-time interactive hand-tracking to detect content pointed to by the child. This novel combination bridges accessibility barriers while deliberately prioritizing the intimacy of shared reading. By enabling parents to rehearse, recall, and deliver stories in their own voices, WordBond goes beyond mere functional accessibility to cultivate a richer and more emotionally resonant parent-child experience.

Through an iterative design process informed by human-centered principles, we developed and evaluated a functional prototype with blindfolded sighted participants simulating VIP use. Our findings reveal both the promise and the interactional challenges of AI-enhanced shared reading, pointing toward design refinements such as segmented audio feedback and dedicated rehearsal phases to support a more natural and less cognitively demanding storytelling experience. Ultimately, this work contributes to the design of inclusive assistive technologies that emphasize not only usability but also the affective and relational dimensions of family interaction, extending previous research on accessible co-reading [9].

# 2 Background and Related Work

This section provides the necessary background for our work by first establishing the affective and cognitive importance of shared reading within the family unit *(Section 2.1)*. We then detail the unique accessibility challenges faced by visually impaired parents (VIPs) and their current coping mechanisms *(Section 2.2)*. Next, we review the landscape of accessible AI and assistive technologies *(Section 2.3)*. Finally, we articulate the critical gap in existing solutions that fail to support intimate, parent-led co-reading, thereby motivating the design of WordBond *(Section 2.4)*.

## 2.1 The Affective and Cognitive Value of Shared Reading

Reading aloud is a practice of significant developmental importance, fostering a child's cognitive and linguistic growth. Studies consistently demonstrate that shared reading enhances language and vocabulary skills while accelerating the development of pre-literacy abilities essential for reading and writing [23, 26, 36]. Furthermore, it is shown to activate brain regions associated with mental imagery and narrative comprehension [18, 24, 28].

Beyond the cognitive domain, research highlights the profound affective and relational benefits of parental involvement. When parents employ strategies such as connecting the story with the child's life or asking open-ended questions, it significantly improves the development of language and literacy of the child and cultivates a lasting interest in reading [1, 2]. Crucially, shared reading is a powerful mechanism for strengthening the parent-child relationship, creating a focused shared activity that deepens emotional bonds and closeness [8].

## 2.2 Accessibility Challenges and Coping Strategies for Visually Impaired Parents (VIPs)

Despite the profound benefits, these intimate moments are not universally accessible. Visually impaired parents (VIPs) face significant structural barriers when trying to read with their sighted children. Only a small fraction of children's books are available in accessible formats (e.g., braille or audio), and limited braille literacy among VIPs further restricts access to these resources [9, 14, 31].

Existing research, though limited, offers valuable insights into the needs of this user group. VIPs consistently place a high value on the intimate experience of reading to their children and overwhelmingly prefer to use their own voices to maintain a strong emotional connection, often viewing their voice as central to the relationship [6].

The literature outlines several coping strategies currently employed by VIPs:

- **Braille Books**: A standard but problematic solution, as utility is restricted by low braille literacy rates in many geographies [27].
- **Child-Led Reading**: Encouraging older children with stronger reading skills to take the lead; however, this postpones engagement and shifts the parental role [7].
- **Memorization**: VIPs often prepare by reading accessible versions in advance to memorize the text, enabling them to

recite the story later, This method is highly labor-intensive and limits spontaneous interaction [10].

### 2.3 Accessible AI and Assistive Technology for Visual Impairment

The field of accessibility has seen rapid growth with innovative technology designed to enhance independence for visually impaired individuals. Solutions fall primarily into two categories:

**General Accessibility Tools**: These focus on mobility and daily living. Examples include *BlindSquare* [34], a navigation app that combines GPS with public data sources (such as Foursquare and OpenStreetMap) to provide a real-time environmental context. Similarly, *Dot Watch* [11] uses a tactile braille interface for communication, and *Be My Eyes* [3] connects blind users with sighted volunteers through live video for immediate visual assistance.

**AI-Powered Perception Tools**: The introduction of Artificial Intelligence (AI), specifically deep learning models, has enabled context-aware applications with personalized features. Solutions such as the *WeWALK Smart Cane 2* [5, 38] integrate sensors and a voice assistant for real-time surroundings information. The *Envision App and Glasses* [37] leverage AI to describe objects, text, and scenes. Most notably, *Seeing AI* [12, 16] utilizes Optical Character Recognition (OCR) and text-to-audio models to provide rapid auditory descriptions of text, objects, and people.

### 2.4 The Gap: Limitations of Current AI in Relational Contexts

While the AI solutions described in *Section 2.3* are powerful tools for independent task completion, literature shows that the target group has a decidedly unfavorable opinion on using them for shared reading activities [20]. The core issue is a misalignment between technology design and relational need.

Specific limitations include: The OCR models in these applications typically struggle with the decorative fonts, non-linear layouts, and complex imagery common in children's books [21]. Their performance and accuracy are highly dependent on ideal lighting conditions, which is often not the case during bedtime reading [22]. Most critically, these applications provide direct, synthesized audio feedback. This functionality does not meet the requirement of reading with the parent's own voice, which is highly valued by VIPs for maintaining emotional closeness [4].

Therefore, current commercial and research demos fundamentally fail to support shared reading because they prioritize the delivery of content over the preservation of the parent-child voice bond. This paper addresses this specific, high-value gap by exploring the design of an AI-enhanced system that assists with technical access while foregrounding the intimacy of natural storytelling.

## 3 Design Implementation

### 3.1 System Overview

We developed WordBond, a mobile assistive technology built to fundamentally enhance the shared reading experience for visually impaired parents (VIPs) and their sighted children. Unlike existing accessibility solutions that prioritize information delivery, WordBond's core objective is to foster a meaningful and interactive parent-child connection by emphasizing the use of the parent's own voice. This design choice directly addresses the critical gap identified in the literature regarding the preservation of the emotional voice bond.

WordBond is designed to be versatile and inclusive, operating with any physical children's book regardless of language or the parent's residual vision. Its primary features are realized through an accessible audio-based interface that supports (Fig.2):

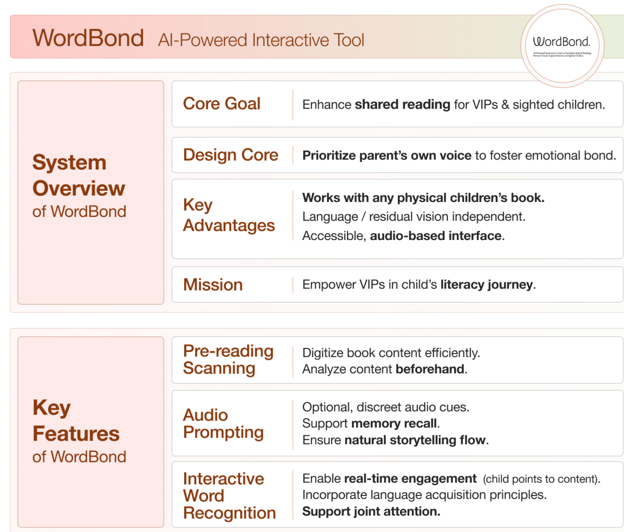


**Figure 2: System Overview and Key Features of WordBond.**

- **Pre-reading Scanning**: Allowing parents to efficiently digitize and analyze a book's content beforehand.
- **Audio Prompting**: Offering optional, discreet audio cues during reading to support memory recall, ensuring a natural storytelling flow.
- **Interactive Word Recognition**: Facilitating real-time engagement when a child points to content, incorporating essential principles of language acquisition and joint attention.

By combining accessibility with emotional connection, WordBond empowers VIPs to play an active and confident role in their child's literacy journey.

### 3.2 Technical Pipeline

The functional integration of WordBond is based on a robust technical pipeline comprising optical character recognition (OCR), hand tracking, a corrective Large Language Model (LLM) and text-to-speech (TTS) conversion. The system operates in two distinct sequential phases: ***Book Scanning and Analysis Session*** and ***Interactive Reading Session*** (Fig.3).

*3.2.1 Phase 1: Book Scanning and Analysis Session.* The first phase involves preparing a book for future reading. To overcome the challenge of precisely aligning a mobile camera, especially for users with partial or complete blindness, a simple plastic frame accessory has been designed. This frame clips onto the book, securely holding

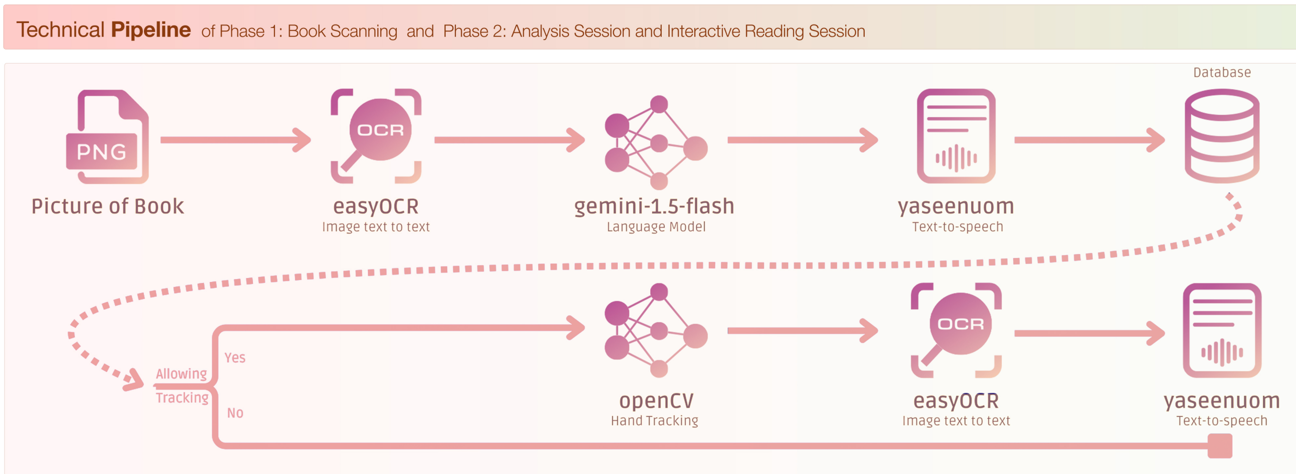


**Figure 3: Technical Pipeline of Phase 1: Book Scanning and Phase 2: Analysis Session and Interactive Reading Session.**

the mobile phone to guarantee the camera is correctly positioned, thereby improving the stability and accuracy of page scans and subsequent recognition.

Users initiate the process by selecting the option to add a new book. The camera captures an image of the open pages, which is then split and converted to gray-scale. The image is processed using the *EasyOCR model* [19, 29]. EasyOCR was chosen over alternatives like PyTesseract [17] (a Python wrapper for Google's Tesseract) due to its superior performance under suboptimal conditions, such as blurry or poorly lit images common in home environments. The recognized text, output as character strings, is then sent to an LLM with a text correction prompt (detailed in Tab.1) to filter out common OCR errors and reconstruct well-formed sentence segments. Finally, these segments are processed by a text-to-speech model, which generates MP3 audio files. The audio files are stored locally, linked to the specific book and page. Users can continue scanning subsequent pages by turning the page and tapping the screen; a double-tap ends this scanning phase and saves the book.

**Table 1: Text Correction Prompt Sent to LLM (Along with EasyOCR-Recognized Text)**

**Firstly:**

• *You will be given a text with errors in some letters: 'text2'.*
• *Correct all errors and provide the intended text as truthfully as possible.*
• *Only reply with the fully corrected text — do not repeat the input or include any additional comments.*

**Important:**

• *Only reply with the final output, formatted with '|' as described.*
• *Do not include any extra words or explanations — only the corrected and split text is allowed in the output.*

*3.2.2 Phase 2: Interactive Reading Session.* The second phase is designed to facilitate the real-time shared reading experience. After navigating through their library via an audio-guided interface, users select a book and choose whether to enable the interactive reading mode.

If enabled, the system activates the camera and employs Mediapipe [15, 25] for real-time hand-tracking. The system is configured to detect a pointing gesture, specifically when the index finger is extended while other fingers are folded. When this gesture is recognized, the system, at a set interval, captures a snapshot of a small text area directly above the fingertip. This snapshot is processed by EasyOCR. If the text is successfully extracted, the word nearest the physical center of the snapshot is converted into an audio file. A non-intrusive notification sound alerts the parent that a word has been identified. The parent can then either continue reading without interruption or perform a double-tap on the screen to hear the spoken word immediately. This real-time seamless support bridges the gap between physical books and digital assistance, allowing spontaneous, meaningful interaction.

## 3.3 Interaction Design and Prototype

A functional prototype of the WordBond application was developed using a combination of Python, JavaScript, HTML, and CSS for user testing purposes. While a simplified version for Android phones was created for basic replay functionality, the full prototype, maintaining a mobile application's shape and interaction, was tested on a desktop environment to ensure full component integration.

The interface is driven entirely by audio cues and simple tap commands, designed to minimize visual dependence (Fig.4). Upon opening the application, users receive a short, guiding audio prompt (lasting no more than 10 seconds) that explains the purpose of the current page and the available options. The interface logic is based on simplicity: options are typically limited to two buttons per screen, and any action requiring confirmation (e.g., confirming a selection) is accompanied by an additional voice message.


**Figure 4: Interaction Design of WordBond: Application User Interfaces and Core User Flow**

The core user flow (Fig.4) revolves around two main actions on the landing page: adding a new book or selecting an existing one. For a detailed account of the user flow, please refer to Appendix A.

**Adding a Book**: This triggers the Phase 1 process *(Section 3.2.1)*, guiding the user through capturing pages and generating accompanying audio for each sentence segment.

**Reading Session**: Selecting a book from the library immediately initiates the reading session. If the interactive mode is enabled, the system monitors the pointing gestures. When a child points at a word, the parent hears a discreet alert. The parent can ignore this prompt or perform a *double-tap gesture* on the screen. This confirmation triggers the pronunciation of the recognized word and pauses any background narration, allowing the parent to confidently communicate the word and engage the child.

This tap-based, audio-first design ensures accessibility while facilitating the natural, shared interaction that is central to the WordBond philosophy.

## 4 Pilot Study

The pilot study served as a crucial step in our iterative design process, intended to validate the initial prototype's usability and feasibility before proceeding to a formal evaluation with the target user group. The primary objectives were to filter out unexpected system behaviors, assess potential cognitive overload imposed by the novel interaction model, gather preliminary feedback on design choices, and collect necessary data to refine our technical algorithms and study protocol for the final user study. Both the pilot study and the final user study were approved by the Ethical Review Board of Eindhoven University of Technology, The Netherlands.

### 4.1 Study Set Up and Participants

Due to the early stage of the initial prototype and the practical constraints associated with recruiting the target population, we adopted a simulated user approach. We recruited four participants (P1–P4), aged 18 to 25, through network sampling. All participants had a background in Design, which allowed us to gather valuable exploratory feedback on design considerations from peers familiar with user experience principles.

We opted for simulated testing *(using blindfolded sighted participants)* as a practical measure due to the significant time and logistical resources required to recruit visually impaired parents (VIPs). This approach is commonly accepted in early-stage HCI prototyping to establish technical viability and address initial usability hurdles before engaging with a vulnerable user group. Testing with a representative sample of VIPs is planned for the subsequent main evaluation phase, once the system achieves greater stability.

## 4.2 Study Protocol and Data Collection

To simulate the target context, participants were blindfolded throughout the test. They were introduced to the audio-based interface and guided through the following two key tasks (see Fig.5):

- **Preparation Task**: The participants selected a book and used the application to scan its pages. The resulting text was processed to generate the accompanying audio version of the story. This step was designed to familiarize participants with the application's setup and the book's content.
- **Reading Task**: Participants read the book aloud to one of the researchers, who actively assumed the role of a child, prompting interaction by asking questions about the text or pointing to words. Participants used WordBond's interactive features to assist their responses.

Data collection was multi-modal, including direct observation, quantitative questionnaires, and semi-structured interviews (see Appendix B). We selected the following instruments:

- **NASA Task Load Index (TLX)** [13]: This instrument was chosen specifically to quantify the cognitive, temporal, and physical demands associated with the novel, multi-sensory interaction model, helping us pinpoint sources of user effort.
- **User Experience Questionnaire (UEQ)** [30, 35]: The UEQ was administered to assess both Pragmatic Quality (usability, efficiency) and Hedonic Quality (novelty, attractiveness), providing a balanced view of the user experience.

Qualitative data from observations and interviews were analyzed using Thematic Analysis, while quantitative questionnaire data were statistically analyzed to inform necessary prototype adjustments.

## 4.3 Pilot Study Results

The quantitative results (see Fig.6) highlighted a core contradiction in the prototype's user experience. While the overall user impression was positive, reflected in the UEQ *Attractiveness* score of 1.08 (classified as "low good"), the *Hedonic Quality* dimensions (Stimulation, Novelty) demonstrated stronger performance. The participants expressed favorable opinions about the "very novel" (P2) concept and its perceived usefulness in achieving the goal of fostering a "connection between parent and child" (P1).

However, significant challenges were identified in the *Pragmatic Quality* dimensions. The Perspicuity (clarity/learnability) dimension scored a low 0.06, confirming that users found the initial prototype difficult to operate. The high scores in the NASA TLX (specifically Mental Demand, Temporal Demand, and Effort) corroborated this, suggesting that the reading activity required significant cognitive effort to maintain accuracy.

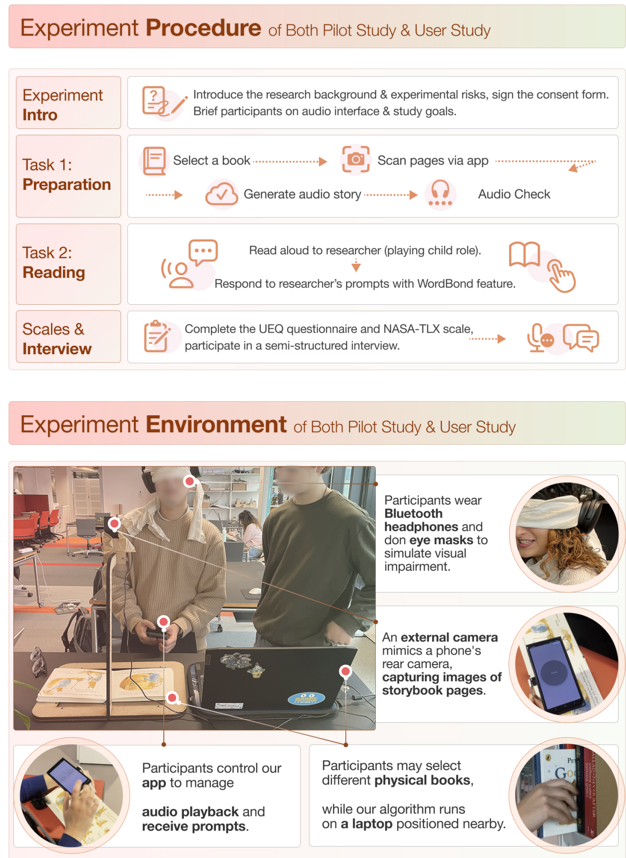


**Figure 5: Experiment Procedure and Environment of Both Pilot Study and User Study.**

Observations and semi-structured interviews revealed specific usability bottlenecks:

**Instructional Overload**: Although the auditory instructions were clear, they often were delivered too quickly and sometimes too long, making them difficult to recall (P1, P2, P4). Participants frequently requested reminders regarding the core navigation gestures (e.g., double-tap).

**Recitation Difficulty**: The initial design provided audio playback in long segments (entire pages or long paragraphs). Participants struggled to maintain structured, natural sentences when trying to recite the text after hearing it, often forgetting the exact wording in their attempt to be completely accurate.

## 4.4 Prototype Improvements and Algorithmic Refinements

The feedback gathered from the pilot study provided clear, actionable insights for design improvements. Participants advised caution regarding the heavy use of audio, noting that it is a dense information medium that can quickly overwhelm users (P2). They universally suggested making audio instructions shorter, slower, and enabling the option to repeat them. They also proposed adding clear

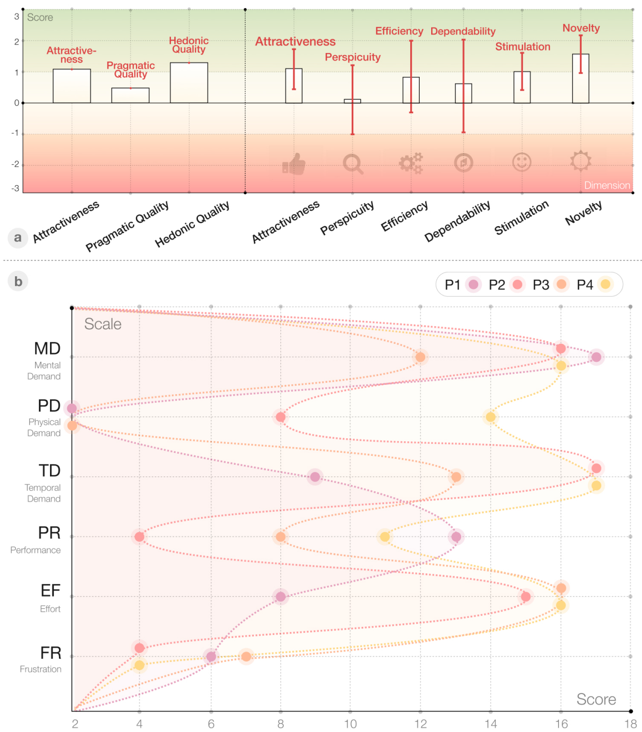


**Figure 6: Subjective Evaluation Results of Pilot Study: (a) User Experience Questionnaire (UEQ) , (b) NASA Task Load Index (TLX).**

confirmation cues and implementing pauses between sentences in the book audio, with one participant (P1) suggesting the ability to repeat each sentence segment "two to three times."

Based on this critical feedback, the following list of iterative changes was immediately implemented in the prototype to significantly reduce cognitive load and improve usability.

- **Audio Segmentation**: Book replay files were shortened from entire pages to single, naturally segmented sentences.
- **Audio Control**: The user can now easily play, repeat and move back through previously read sentence audio files.
- **Instructional Clarity**: All navigation instructions were shortened, clarified, and given a dedicated replay function.
- **Feedback Cues**: Additional auditory and tactile indications were added throughout the app to signal successful button presses and process loading states.

A key technical refinement was the integration of a Corrective Large Language Model (LLM) into the scanning pipeline. Initial testing of the OCR model (EasyOCR) revealed instability, with environmental factors like lighting causing frequent errors (e.g., replacing 'c' with '['). While a simple spell-checker could fix single-character errors, it failed to address multi-character mistakes, lookalike characters, or, crucially, context, often skipping small but necessary words (like 'a' or 'the').

To resolve this and ensure high-quality audio segments, the LLM was introduced to perform two chained tasks (Tab.2): error correction and natural sentence breaking. The prompt provided to the LLM is structured to ensure output accuracy and flow, generating the sentence segments required for an improved audio experience.

**Table 2: Refined Text Correction Prompt Sent to LLM (Along with EasyOCR-Recognized Text)**

**Firstly:**

• *You will be given a text with errors in some letters:* `text2`.
• *Correct all errors and provide the intended text as truthfully as possible.*
• *Only reply with the fully corrected text — do not repeat the input or include any additional comments.*

**Secondly:**

• *Take this corrected text and break it into small sentence parts, separating each part with a '|' character.*
• *Splits should occur at natural pause points such as periods ('.'), commas (','), or appropriate places in long sentences where a speaker would naturally pause.*
• *Use logical splits that maintain normal sentence flow — meaning it should still sound natural if spoken aloud, with a '|' marking where a brief pause would be.*

**Important:**

• *Only reply with the final output, formatted with '|' as described.*
• *Do not repeat or duplicate text from the input.*
• *Do not include any extra words or explanations — only the corrected and split text is allowed in the output.*

These iterative changes, derived directly from the pilot findings, address the core usability and cognitive load issues, preparing the refined prototype for the subsequent formal user study.

## 5 User Study

The formal user study was conducted to rigorously evaluate the user experience (UX) of the refined WordBond prototype. Specifically, this study aimed to quantify the reduction in cognitive load after iterative improvements based on the pilot study, and to formally assess the usability and hedonic qualities of the system in a simulated reading environment. The data collected provides validation for the current design and provides critical insights for future deployment with the target user group.

### 5.1 Participants

We recruited seven participants (P1–P7), aged 18 to 25, through network sampling from a technical university environment. Similar to the pilot study, the participants were sighted students, primarily with backgrounds in design and engineering. Our decision to utilize a simulated user approach (blindfolded sighted participants) for this formal study, rather than recruiting visually impaired parents (VIPs), was based on two primary factors:

**Ethical Imperative and Risk Mitigation**: The pilot study indicated significant challenges with the complexity of the initial prototype and documented a high cognitive load *(Section 4.3)*. Deploying a prototype with known high cognitive demands in the target population carries an ethical risk of causing frustration and potential harm to users with functional limitations. This simulation allowed us to achieve greater design stability before engaging the vulnerable target group. By controlling the visual variable, the study could focus solely on evaluating the efficacy of the revised audio-guided interface, the segmented audio feedback, and the core technical pipeline (OCR/Hand-tracking), which are the essential components developed since the pilot study.

Furthermore, for safety and ethical expediency, real sighted children were not involved. A researcher assumed the role of the child confederal, interacting spontaneously with the book and prompting the participant, a method validated in the pilot to effectively simulate the interactive stimuli required for system evaluation.

## 5.2 Study Protocol

The protocol was designed to be highly realistic while ensuring that all interactions were logged and controlled. Upon arrival, participants were informed about the purpose of the application (assisting VIP parents) and the two main expectations of the test: (1) Scan a book and (2) Read the book to a researcher.

Participants were explicitly told that the audio-guided interface of the application was designed to provide all necessary instructions, and the researchers provided no further assistance after the initial briefing. Then they were asked to blindfold themselves and handle the mobile phone with the installed application. This protocol successfully simulated the user experience as intended—forcing reliance solely on audio prompts, touch gestures, and system functionality via a mobile device. Additionally, the experimental procedure of the current user study is consistent with that of the pilot study (see Fig.5).

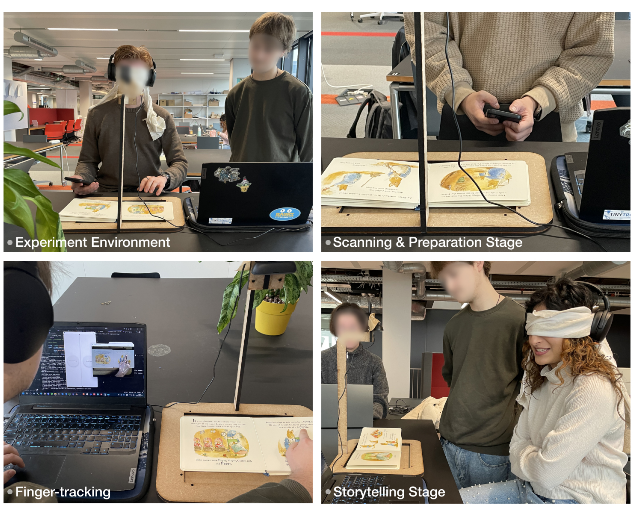


**Figure 7: Demonstration of the Experimental Process.**

## 5.3 Data Collection and Measures

Data collection was carried out using a mixed-method approach to ensure the triangulation of qualitative insights with quantitative metrics of user experience and performance.

*5.3.1 Quantitative Data.* We used two validated instruments commonly used in Human-Computer Interaction research: *User Experience Questionnaire (UEQ)* [30, 35]: This questionnaire uses a 7-point semantic differential scale (ranging from −3 to +3) to provide a direct and standardized evaluation of user experience. It yields scores across six dimensions, providing insight into Pragmatic Quality (Perspicuity, Efficiency, Dependability) and Hedonic Quality (Stimulation, Novelty, Attractiveness). *NASA Task Load Index (TLX)* [13]: The TLX was utilized to quantify the subjective workload and cognitive effort associated with the use of the application. It employs 21 gradations across six subscales (e.g., Mental Demand, Temporal Demand, Effort), which were critical for validating the success of the pilot-driven design modifications aimed at reducing cognitive load.

*5.3.2 Qualitative Data.* To gather in-depth qualitative insights into user experiences with WordBond, this study employed two research methods: Real-Time Observations and Semi-Structured Interviews. Real-Time Observations: Researchers log interactions, specific nonverbal cues, and moments of struggle or success to provide rich contextual data on user interaction patterns with WordBond. Semi-Structured Interviews (see Appendix B): After the test, the interviews explored the general experience of the participants, the perceived ease-of-use, the utility of specific features, the opinions on responsible use of AI and suggested areas for future improvement.

## 5.4 Data Analysis

The collected data was analyzed using distinct, rigorous methods for both quantitative and qualitative datasets.

*5.4.1 Quantitative Data Analysis.* Data from the UEQ and NASA TLX were subjected to quantitative analysis. The UEQ data was processed using the official UEQ Excel Analysis Tool [30], which automatically calculates and visualizes the scores against the benchmark data. NASA TLX data were subjected to standard statistical analysis (e.g., calculation of means and standard deviations) to compare workload between different tasks and against the findings of the pilot study.

*5.4.2 Qualitative Data Analysis.* Qualitative data, derived from semi-structured interviews and observation logs, was analyzed using Thematic Analysis [33]. All interviews were audio-recorded and transcribed verbatim using Microsoft Word's transcription tool. The analysis followed established steps: initial review of the transcripts and logs to identify recurring topics and generate initial codes; subsequent organization of these codes into broader, meaningful themes; and final interpretation to capture overarching interaction patterns and user feedback. This process facilitated the triangulation of qualitative findings with quantitative scores from the UEQ and TLX.

# 6 Results

## 6.1 User Experience Questionnaire (UEQ)

The results of the User Experience Questionnaire (UEQ) indicate a positive overall impression of the system, with all six dimensions scoring above the threshold 1.0 (representing the "above average" user experience). However, the results highlight a crucial discrepancy between engagement and usability (Fig.8).

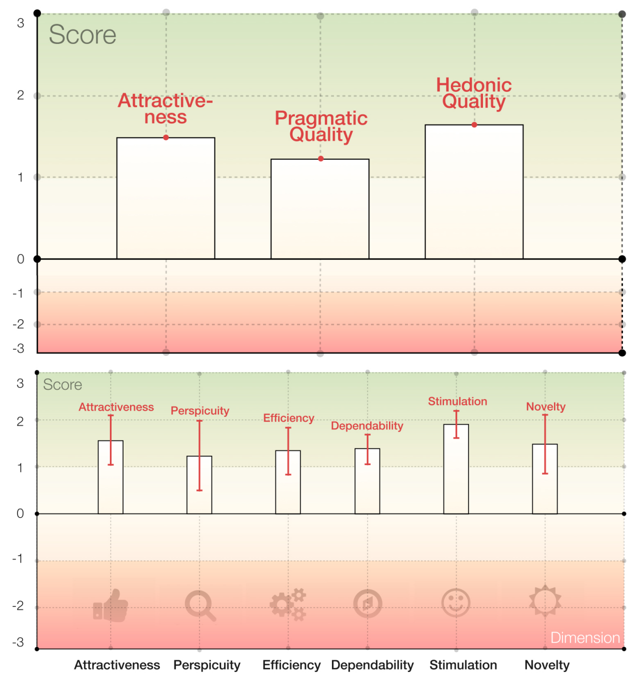


**Figure 8: User Experience Questionnaire (UEQ) Scores of User Study.**

The Hedonic Quality dimensions scored significantly higher ($\bar{x}$ = 1.68) than the Pragmatic Quality dimensions ($\bar{x}$ = 1.30).

- **High Scores (Hedonic)**: Stimulation ($\bar{x}$ = 1.89) and Novelty ($\bar{x}$ = 1.46) received the highest ratings, confirming that users perceived the concept as highly innovative and engaging. Attractiveness ($\bar{x}$ = 1.55) also indicated a generally favorable impression of the system's design.
- **Lower Scores (Pragmatic)**: The dimensions of pragmatic quality, specifically Perspicuity ($\bar{x}$ = 1.21), Efficiency ($\bar{x}$ = 1.32), and Dependability ($\bar{x}$ = 1.36), scored relatively lower. This suggests that while the system is appealing and stimulating, some users found the prototype less intuitive and experienced minor friction points in the interaction flow.

## 6.2 NASA Task Load Index (TLX)

The NASA TLX results revealed an Overall Workload Score (average of all sub-scales) of $\bar{x}$ = 9, which is categorized as medium-low workload (see Fig.9). This shows a significant reduction in the cognitive effort required compared to the initial prototype in the pilot study (where the effort was rated high). The scores highlighted several key findings:

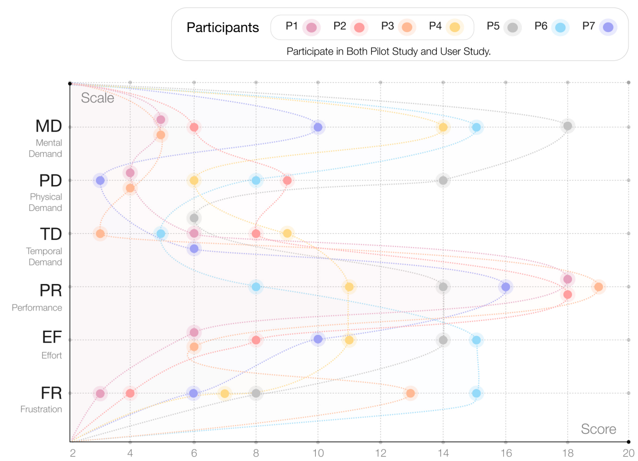


**Figure 9: NASA-TLX Scale Scores of User Study Across Scales.**

- **Low Demand**: Physical Demand and Temporal Demand were consistently low, indicating that the activity was neither physically taxing nor hurried. The frustration scores were also low, suggesting that the participants experienced minimal irritation or stress during the tasks.
- **High Performance**: Participants reported the highest scores on the Performance sub-scale ($\bar{x}$ = 15), generally feeling that they had successfully completed the task.
- **Variability in Cognitive Load**: Both Mental Demand and Effort showed greater variability. Participants with prior experience from the pilot (P1, P2) consistently reported lower scores in these categories, while new participants (P4, P5, P6) recorded higher mental demand scores. This suggests that a learning curve still exists, but the workload can be effectively managed once basic interaction patterns are learned.

## 6.3 Qualitative Findings and Thematic Analysis

The thematic analysis of observations and semi-structured interviews provided context for the quantitative results, particularly explaining the discrepancy between high novelty and moderate usability. Two primary themes emerged: *Friction in Interactive Engagement* and *Monotony in Storytelling Delivery.*

### 6.3.1 *Friction in Interactive Engagement.*

All participants struggled with the purpose and expected response for the Interactive Mode.

Confusing Feedback: When the feature was triggered (e.g., by the child/researcher pointing), participants were uniformly surprised and unsure how to respond to the notification sound ("ding!"). One participant (P4) initially mistook it for a regular phone notification, while another assumed the sound signaled the end of the reading session (P6).

Lack of Context: Participants felt the notification lacked the necessary context, failing to bridge the gap between the sound and the required physical action. P5 suggested clarifying the purpose of the feature by providing an explicit auditory message, such as: "Your child would like to know this word." This qualitative finding directly supports the lower scores observed in the UEQ's Perspicuity dimension.

*6.3.2 Monotony in Storytelling Delivery.* The researchers observed that all participants read in a flat, monotonous tone during the session and that the sentences repeated by the participants often lacked clear semantic coherence. The interviews revealed two key contributors to this issue:

High Focus on Accuracy: Participants (P2, P3, P5, P6) revealed that they were very focused on accurately repeating the audio segments. This intense focus caused them to neglect the overall structure of the story and the meaning of the sentences. This difficulty was attributed to the sometimes unnatural segmentation of sentences, which complicated their ability to apply proper intonation and rhythm when reading aloud.

Environmental Discomfort: A secondary factor was the discomfort caused by the testing environment. Conducting the test in an open public space made participants (P3, P4, P5) feel self-conscious and reluctant to use expressive voices or perform dramatic readings, which is essential for engaging children.

To address the challenge of flat delivery, participants offered two valuable design suggestions: (1) The inclusion of a rehearsal phase to allow users to familiarize themselves with the book's narrative before the reading session. (2) Integration of an image recognition model capable of describing the illustrations in the book (P2, P4, P5, P6). This image description would enable VIPs to fully visualize and engage with the story's visual components, thus enhancing their overall experience and delivery.

## 6.4 Comparative Analysis: Success of Design Iterations

The qualitative feedback from the returning participants (P1 and P2) and the quantitative shift in the UEQ and TLX results unequivocally validate the success of the prototype improvements implemented after the pilot study *(Section 4.4)*.

- **Confirmed Usability Improvement**: The general increase in all the parameters of the UEQ (all now $\bar{x} > 1.0$) confirms that the system has moved from a state of low usability to a state of above-average experience.
- **Effectiveness of Segmentation**: Returning participants (P1, P2) explicitly confirmed that the instructions were clearer and the reading phase was significantly easier due to the shorter audio segments. They found the shorter segments much easier to remember and recite compared to the previous, longer segments, thus enabling them to perform the task more effectively.
- **Impact of Control Features**: New participants (P3–P7) and returning participants alike highlighted the repetition button as a critical feature that contributed directly to successfully completing the task of reading aloud.

In summary, the design iterations successfully mitigated the most severe cognitive load issues documented in the pilot, establishing a baseline of practical usability. The remaining challenges now focus on refining the contextual clues and improving the emotional and narrative quality of the shared reading experience.

# 7 Discussion

The results of the user study, synthesized with the findings of the pilot test, confirm the successful mitigation of initial usability issues while revealing critical areas for future design research. This discussion explores the implications of our quantitative data (TLX and UEQ) on the system's learning curve, analyzes the specific breakdowns in the interactive mode, and examines the disconnect between system performance and the delivery of affective storytelling to the user.

## 7.1 Mitigation of Extraneous Load and the System's Learning Curve

The quantitative results validate the success of our iterative design process, particularly the effective mitigation of the intense workload documented in the pilot study. The significant reduction in the NASA TLX Temporal Demand scores across all participants confirms that segmenting the book audio into natural sentence parts successfully eliminated the perception of being rushed. However, the continued variability in Mental Demand and Effort scores reveals a clear short-term learning curve. Participants new to the system reported a higher cognitive effort, while those with prior exposure (P1, P2) reported considerably lower scores.

This finding suggests that the core mechanics of WordBond are manageable, but require initial familiarization. This phenomenon aligns with *Cognitive Load Theory (CLT)* [32], suggesting that the initial, unfamiliar interaction patterns impose a significant extraneous cognitive load. Until the user masters the specific gestural commands and audio feedback mechanisms of the system, a substantial portion of their working memory is dedicated to managing this extrinsic burden rather than the intrinsic task of storytelling. Future work should address this initial hurdle by integrating a mandatory, example-based tutorial or guided setup phase. Although our goal was to test usage in common practice (without prior rehearsal), the data indicates that a brief, dedicated onboarding phase would significantly flatten the learning curve, providing a more accurate assessment of the steady-state cognitive load for the long-term user.

## 7.2 Bridging Novelty and Usability: Joint Attention and Self-Efficacy

The results of the User Experience Questionnaire (UEQ) highlight a fundamental design conflict: high hedonic quality versus moderate pragmatic quality. The participants enthusiastically embraced the novelty ($\bar{x}$ = 1.46) and stimulation ($\bar{x}$ = 1.89) of the concept, suggesting that the potential for emotional support is high. However, the moderate pragmatic score ($\bar{x}$ = 1.30), particularly the low Perspicuity ($\bar{x}$ = 1.21), confirms that the clarity and ease-of-learning of the system still need improvement.

The moderate scores in the UEQ's Pragmatic Quality, especially in Perspicuity, are largely attributable to the breakdown observed within the Interactive Mode. Although the feature—where a child points to a word, triggering an audio cue—was conceptually praised for fostering joint attention, its execution failed to establish a clear affordance. The ambiguous "ding!" notification lacked situational awareness, causing users to mistake it for a generic device alert and consequently disrupting the flow of shared reading. This failure goes beyond simple usability; it introduces a point of friction that directly threatens the parent's Parenting Self-Efficacy. In a critical moment designed for teaching and bonding, the parent is left confused and unable to respond effectively, undermining their confidence and control. The appreciation for the feature only emerged after the researcher's intervention, suggesting that the system must learn to mediate the social context more explicitly.

Future design must replace the generic chime with a rich, contextual auditory cue (e.g., "Child is pointing") that clearly establishes the social significance of the interruption, thereby empowering the parent to confidently and seamlessly assume the role of educator.

## 7.3 Recitation vs. Storytelling Mindset: The Affective and Narrative Gap

The most complex challenge revealed by the study lies in the quality of emotional delivery. The observed flat, monotonous reading tone, coupled with participants' self-reports of focusing on technical accuracy (repetition) over affective expression (intonation), indicates a severe cognitive bottleneck. Users were spending too much mental effort reconstructing the narrative flow of the segmented audio and compensating for the LLM's occasional unnatural sentence breaks. This high cognitive burden effectively starved the resources needed for affective labor—the ability to modulate tone, rhythm, and emotion crucial for engaging a child. This inability to expressively read directly impedes emotional synchrony between the parent and the child, which is the psychological bedrock of the desired bond.

The participants' suggestion to integrate an image recognition model to describe illustrations is a profound insight into this psychological mechanism. Providing this affective context would allow the parent to instantly visualize the story, enabling them to naturally generate the corresponding expressive tone and shifting the task from blind repetition to mediated co-creation of a vivid narrative.

## 7.4 Ethical Boundaries and the Future of Relational Assistive Technology

Although this study successfully validated the prototype's core technical feasibility and established a robust usability baseline, its primary limitation remains the reliance on simulated users. The findings, though essential for refining the system, cannot yet confirm the ultimate impact of WordBond on the parent-child relationship, including metrics such as increased parental pride or genuine child engagement. Testing with real visually impaired parents and their children is the definitive next step, requiring a transition across the current ethical boundary from simulation to longitudinal, ecologically valid deployment. This future research must go beyond usability scores to investigate the role of the system as a Mediated Relational Tool—how it affects family routines, reinforces the VIP's identity as an active parent, and facilitates sustained emotional connection. WordBond's value lies not only in its function as a compensatory technology, but in its potential to serve as a catalyst for identity empowerment, establishing a design precedent for future assistive technologies that prioritize affective communication and relational well-being.

# 8 Limitations and Future Work

The current WordBond prototype, while validating key usability improvements, presents several notable limitations that define the crucial roadmap for future work. A significant constraint was the testing environment itself; conducting sessions in a public setting created social discomfort that prevented participants from fully immersing themselves in the expressive storytelling required for affective reading, thus limiting the ecological validity of the qualitative data. This limitation strongly necessitates the transition to in-situ field studies involving the actual target population of visually impaired parents (VIPs) within the intimacy of their homes. Engaging VIPs will not only eliminate environmental bias, but will also accurately account for any unique compensatory skills they possess, which may lead to a vastly different and potentially lower long-term cognitive load profile than observed with our simulated users.

Further limiting the fidelity of the current interaction was the technical architecture, where machine learning models were offloaded to a laptop while participants interacted with the front-end on a mobile device. The next stage of development must achieve full on-device autonomy to enable truly seamless interactions, allowing for the intended use of the mobile camera and the custom holder aid. Tying into this, the compromise made to stabilize unreliable OCR output with an LLM, while improving immediate usability, introduced a trade-off in fidelity to the original text. Future work must resolve this by developing a robust multi-modal stability pipeline that minimizes text generation, ensuring that the parent always reads a faithful representation of the source material.

The most critical area for future research lies in refining the system's affective and relational model. The qualitative finding regarding the monotonous reading tone indicates a cognitive bottleneck where resources are currently directed toward technical repetition rather than emotional expression. To liberate the parent's working memory, we plan to prioritize the integration of Image Description Integration, as suggested by participants, to provide the necessary affective context for expressive reading. This will be coupled with the implementation of an optional rehearsal phase during book preparation, shifting the focus from simply reciting the text to internalizing the emotional arc of the story.

Finally, to truly validate the purpose of WordBond as a relational tool, future research must explicitly incorporate the child's feedback and engagement metrics. The current study is adult-centric; a complete assessment requires measuring the child's learning outcomes and emotional responses. Therefore, the ultimate future step is a longitudinal deployment study with a larger sample of VIP families, which will fully validate the long-term impact of the system on parental self-efficacy and the affective quality of the shared parent-child bond.

## 9 Conclusion

WordBond illustrates the potential of leveraging AI technologies to enhance inclusive practices for visually impaired parents and their children. By focusing on accessibility, interaction, and the emotional connection intrinsic to shared reading, the application provides a meaningful tool for fostering intimacy and supporting early childhood development. Iterative user testing highlighted both the system's strengths and areas for improvement, leading to refinements that address usability and cognitive load challenges.

Future work could focus on deploying WordBond in real-world scenarios involving visually impaired parents to gather authentic feedback and refine its capabilities further. Integrating additional features, such as machine vision for image descriptions and expanded support for diverse languages, will enhance its utility and inclusivity. By bridging technological innovation with human-centered design, WordBond paves the way for assistive tools that not only are functional but also create space for an intimate bond between VIPs and their sighted child.

## Acknowledgments

The authors gratefully acknowledge the assistance provided by large language models, specifically ChatGPT and Google Gemini, during the development phase of this project. These tools were utilized to assist with writing and refactoring specific code portions, debugging sections of the implementation, and providing valuable suggestions for streamlining efficiency. Their contributions were limited to assisting the implementation process under the direct supervision of the authors.

## References

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 249–258. doi:10.1145/3234695.3236344

[2] Ali Abdolrahmani, Kevin M. Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M. Branham. 2020. Blind Leading the Sighted: Drawing Design Insights from Blind Users towards More Productivity-oriented Voice Interfaces. *ACM Trans. Access. Comput.* 12, 4, Article 18 (Jan. 2020), 35 pages. doi:10.1145/3368426

[3] Mauro Avila, Katrin Wolf, Anke Brock, and Niels Henze. 2016. Remote Assistance for Blind Users in Daily Life: A Survey about Be My Eyes. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Island, Greece) *(PETRA '16)*. Association for Computing Machinery, New York, NY, USA, Article 85, 2 pages. doi:10.1145/2910674.2935839

[4] Syed Khuram Shahzad Madeeha Gul Natash Ali Mian Muhammad Raza Naqvi Babar Ayub Khan Azeem Shera, Muhammad Waseem Iqbal. 2021. Blind and Visually Impaired User Interface to Solve Accessibility Problems. *Intelligent Automation & Soft Computing* 30, 1 (2021), 285–301. doi:10.32604/iasc.2021.018009

[5] Berenika Balcer. 2019. A Rhetoric of WeWalk-ing. Smart Canes and "feeling of freedom" in the urban environment. Webpage. https://mastersofmedia.hum.uva.nl/2019/09/smart-canes-and-urban-environment/ Accessed: [13-Nov-2025].

[6] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 161–173. doi:10.1145/3234695.3236348

[7] Stacy M. Branham, Steve H. Harrison, and Tad Hirsch. 2012. Expanding the design space for intimacy: supporting mutual reflection for local partners. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 220–223. doi:10.1145/2317956.2317990

[8] Stacy M. Branham and Shaun K. Kane. 2015. Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2373–2382. doi:10.1145/2702123.2702511

[9] Cameron Tyler Cassidy, Isabela Figueira, Sohyeon Park, Jin Seo Kim, Emory James Edwards, and Stacy Marie Branham. 2024. Cuddling Up With a Print-Braille Book: How Intimacy and Access Shape Parents' Reading Practices with Children. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 125, 15 pages. doi:10.1145/3613904.3642763

[10] Drew Cingel and Anne Marie Piper. 2017. How Parents Engage Children in Tablet-Based Reading Experiences: An Exploration of Haptic Feedback. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 505–510. doi:10.1145/2998181.2998240

[11] Dot Corporation. [n. d.]. Dot Corporation: Making the World Accessible, Dot by Dot. Webpage. https://www.dotincorp.com/ [n. d.]. Accessed: 2025-11-13.

[12] Microsoft Corporation. 2017. Seeing AI: An App for Visually Impaired People That Narrates the World Around You. Webpage. https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/ Accessed: 2025-11-13.

[13] NASA Human Systems Integration Division. [n.d.]. Task Load Index (TLX) - Human Systems Integration Division. https://humansystems.arc.nasa.gov/groups/tlx/. Accessed: 2025-11-13.

[14] Judy Dixon. 2019. Supercharge Your Braille Reading for Pleasure and Productivity. Webpage, American Foundation for the Blind (AFB). https://www.afb.org/aw/22/7/17623 Accessed: 2025-11-13.

[15] Google AI Edge. 2023. MediaPipe: Cross-platform, customizable ML solutions for live and streaming media. https://github.com/google-ai-edge/mediapipe. Accessed: [2025-11-13].

[16] Christina Granquist, Susan Y. Sun, Sandra R. Montezuma, Tu M. Tran, Rachel Gage, and Gordon E. Legge. 2021. Evaluation and Comparison of Artificial Intelligence Vision Aids: Orcam MyEye 1 and Seeing AI. *Journal of Visual Impairment & Blindness* 115, 4 (2021), 277–285. arXiv:https://doi.org/10.1177/0145482X211027492 doi:10.1177/0145482X211027492

[17] Samuel Hoffstaetter. 2023. pytesseract: Python Wrapper for Google's Tesseract-OCR Engine. PyPI Package. https://pypi.org/project/pytesseract/ Accessed: 2025-11-13.

[18] Fraide A. Ganotice Jr., Kevin Downing, Teresa Mak, Barbara Chan, and Wai Yip Lee. 2017. Enhancing parent-child relationship through dialogic reading. *Educational Studies* 43, 1 (2017), 51–66. arXiv:https://doi.org/10.1080/03055698.2016.1238340 doi:10.1080/03055698.2016.1238340

[19] Rakpong Kittinaradorn and Jaided AI Team. 2024. EasyOCR: End-to-End Multi-Lingual Optical Character Recognition (OCR) Solution. GitHub Repository. https://github.com/JaidedAI/EasyOCR Accessed: 2025-11-13.

[20] Barbara Leporini and Marina Buzzi. 2023. Visually-Impaired People Studying via eBook: Investigating Current Use and Potential for Improvement. In *Proceedings of the 2022 6th International Conference on Education and E-Learning* (Yamanashi, Japan) *(ICEEL '22)*. Association for Computing Machinery, New York, NY, USA, 288–295. doi:10.1145/3578837.3578879

[21] LetsEnvision. [n. d.]. Envision App: AI-Powered Accessibility App for Visual Information Recognition. Webpage. https://www.letsenvision.com/app/ [n. d.]. Accessed: 2025-11-13.

[22] LetsEnvision. [n. d.]. Envision Glasses: AI-Powered Smart Glasses for Visual Accessibility. Webpage. https://www.letsenvision.com/glasses/home [n. d.]. Accessed: 2025-11-13.

[23] Rachael Levy, Melanie Hall, and Jenny Preece. 2018. Examining the Links between Parents' Relationships with Reading and Shared Reading with their Pre-School Children. *International Journal of Educational Psychology* 7, 2 (Jun. 2018), 123–150. doi:10.17583/ijep.2018.3480

[24] Liang Li and Marilyn Fleer. 2015. Family pedagogy: parent–child interaction in shared book reading. *Early Child Development and Care* 185, 11-12 (2015), 1944–1960. arXiv:https://doi.org/10.1080/03004430.2015.1028398 doi:10.1080/03004430.2015.1028398

[25] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[26] Margaret Kristin Merga and Susan Ledger. 2018. Parents' views on reading aloud to their children: Beyond the early years. *The Australian journal of language and literacy* 41, 3 (2018), 177–189. doi:10.1007/BF03652018

[27] Braille Authority of North America (BANA). 2023. BANA in Brief. Webpage. https://www.brailleauthority.org/bana-brief Accessed: 2025-11-13.

[28] Sohyeon Park, Cameron Tyler Cassidy, and Stacy M. Branham. 2023. "It's All About the Pictures:" Understanding How Parents/Guardians With Visual Impairments Co-Read With Their Child(ren). In *Proceedings of the 25th International*

*ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 88, 4 pages. doi:10.1145/3597638.3614488

[29] Binod Kumar Pattanayak, Anil Kumar Biswal, Suprava Ranjan Laha, Saumendra Pattnaik, Bibhuti Bhusan Dash, and Sudhansu Shekhar Patra. 2023. A Novel Technique for Handwritten Text Recognition Using Easy OCR. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*. 1115–1119. doi:10.1109/ICSSAS57918.2023.10331704

[30] Martin Schrepp, Andreas Hinderks, and Jorg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40 – 44. doi:10.25968/opus-3397

[31] Kevin M. Storer and Stacy M. Branham. 2019. "That's the Way Sighted People Do It": What Blind Parents Can Teach Technology Designers About Co-Reading with Children. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 385–398. doi:10.1145/3322276.3322374

[32] John Sweller. 2011. CHAPTER TWO - Cognitive Load Theory. Psychology of Learning and Motivation, Vol. 55. Academic Press, 37–76. doi:10.1016/B978-0-12-387691-1.00002-8

[33] Gareth Terry, Nikki Hayfield, Victoria Clarke, Virginia Braun, et al. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2, 17-37 (2017), 25.

[34] Paraskevi Theodorou, Kleomenis Tsiligkos, and Apostolos Meliones. 2023. Multi-Sensor Data Fusion Solutions for Blind and Visually Impaired: Research and Commercial Navigation Applications for Indoor and Outdoor Spaces. *Sensors* 23, 12 (2023). doi:10.3390/s23125411

[35] Technische Universität München (TUM). [n.d.]. User Experience Questionnaire (UEQ) - Official Website. https://www.ueq-online.org/. Accessed: 2025-11-13.

[36] Yvonne Vezzoli, Sara Kalantari, Natalia Kucirkova, and Asimina Vasalou. 2020. Exploring the Design Space for Parent-Child Reading. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376696

[37] Ethan Waisberg, Joshua Ong, Mouayad Masalkhi, Nasif Zaman, Prithul Sarker, Andrew G Lee, and Alireza Tavakkoli. 2024. Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. *Eye* 38, 6 (2024), 1036–1038.

[38] Ting Wang, Rupert Grobler, and Eric Monacelli. 2020. EVAL Cane: An IoT based Smart Cane for the Evaluation of Walking Gait and Environment. In *2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. 1–4. doi:10.1109/BMSB49480.2020.9379468

# A Detailed User Interaction Flow of WordBond

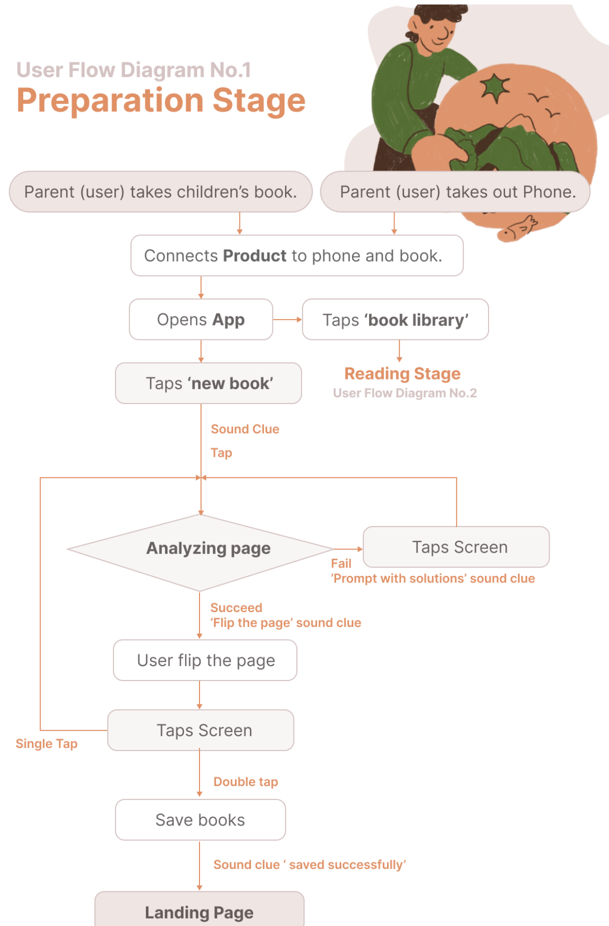


**Figure 10: User Flow Diagram 1: Preparation Stage.**

# B Semi-structured Interview Questions

### Introduction

Thank you for taking the time to participate in our evaluation. We'll now ask you a few questions about your experience with the prototype. This will be a semi-structured interview, so while we have some prepared questions, we encourage you to share any additional thoughts or opinions that come to mind. Your input is highly valuable to us.


**Figure 11: User Flow Diagram 2: Reading Stage.**

**Part 1: General Impression:**

- What did you think about the experience of being blindfolded and interacting with the application?
- What were your first thoughts or impressions of the prototype when you started using it?
- What aspects of the prototype did you find most engaging or effective?

**Part 2: Experience when using the Application:**

- Were you able to easily navigate through the application? Were there any challenges?
- How clear were the instructions or prompts within the application?
- Did you experience any technical issues, such as delays or errors, while using the app?
- How did the application enhance or hinder your experience with the book?
- If you could suggest improvements for the application, what would it be?

**Part 3: Accessibility and Thoughtfulness of the system:**

- Were there any features that made it easier or harder for you to interact with the system?
- Do you feel the system was designed with thoughtful consideration of its target audience's needs and preferences?
- How well do you think the prototype fulfills its intended purpose?

Note: Additional questions may be introduced by the researchers after reviewing the observation notes prior to the interview. The questions listed on this document serve as the foundation for the interview.