Designing for Comprehension: Accessible Personalisation of Audiovisual Media for People with Aphasia

Alexandre Nevsky

Supervised by:

Dr Timothy Neate

Professor Elena Simperl

King's College London

Department of Informatics

A dissertation is submitted for the degree of

Doctor of Philosophy

February, 2026

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 100,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Begin dedication.

Abstract

This thesis treats audiovisual media accessibility for people with complex communication needs as the ability to maintain comprehension during viewing, working with people with aphasia as an exemplar community. It develops a comprehension-first framing in which access is assessed through followability, confidence, attention, and perceived effort during continuous playback, rather than on the presence or accuracy of accessibility features. The thesis focuses on accessible personalisation, where audiovisual media elements, such as audio mix, timing, or structure, are adjustable at playback. Adaptations are designed to be optional and reversible to fit domestic and shared viewing routines.

The work uses an exploratory research-through-design approach to investigate how comprehension support can be implemented in practice. It begins with a systematic review of 181 papers on technical accessibility interventions for digital audiovisual media to characterise how current research addresses disability groups and access problems. It then maps comprehension breakdowns through an online survey with 41 respondents with aphasia and two focus groups with 10 participants. These studies identify recurring access barriers linked to information pace, dense phrasing, multi-speaker scenes, rapid edits, competing on-screen elements, limited time to read on-screen text, and audio mixes that reduce speech clarity.

Participatory design activities with people with aphasia generate candidate interventions that adapt audiovisual media to meet their access needs. A design-for-one stage develops bespoke probes with 4 participants to examine acceptable degrees of content change, control preferences, and interaction constraints. The probes are then refined and examined in an in-home deployment with 9 participants over 2 weeks, including solo viewing and co-viewing, using a custom player that supports

simplified subtitles, speaker spotlighting, playback speed control, and multi-track volume control. Across these stages, the thesis documents how comprehension-first adaptations are selected, adjusted, and used during everyday viewing, including cases where features are avoided due to interaction effort or co-viewing constraints.

Acknowledgments

Begin acknowledgments.

List of publications

[1] Bircanin, F., Nevsky, A., Cruice, M., Markovic, O., and Neate, T. A Sound Understanding: An In-Situ Deployment of an Accessible Audio-Media Player with People Living with Aphasia. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2026), CHI '26, Association for Computing Machinery, pp. 1–21.

[2] Bircanin, F., Nevsky, A., Cruice, M., and Neate, T. "You meet the audience at where they're at, not where you want them to be": Rethinking Media Accessibility and Design with Media Practitioners, 2026. Awaiting peer-review.

[3] Bircanin, F., Nevsky, A., Perera, H., Agarwal, V., Song, E., Cruice, M., and Neate, T. Sounds Accessible: Envisioning Accessible Audio Media Futures with People with Aphasia. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2025), CHI '25, Association for Computing Machinery, pp. 1–22.

[4] Nevsky, A. Object-Based Access: Enhancing Accessibility with Data-Driven Media. In *ACM International Conference on Interactive Media Experiences* (New York, NY, USA, 2023), IMX '23, Association for Computing Machinery, pp. 1–5.

[5] Nevsky, A., Bircanin, F., Cruice, M., Simperl, E., and Neate, T. To Each Their Own: Exploring Highly Personalised Audiovisual Media Accessibility Interventions with People with Aphasia. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference* (New York, NY, USA, 2025), DIS '25, Association for Computing Machinery, pp. 1–18.

[6] Nevsky, A., Bircanin, F., Cruice, M., Simperl, E., and Neate, T. Tuning in to Personalised Accessible Video: A Home Deployment Study with People with Aphasia, 2026. Awaiting peer-review.

[7] Nevsky, A., Bircanin, F., Cruice, M., Wilson, S., Simperl, E., and Neate, T. "I Wish You Could Make the Camera Stand Still": Envisioning Media Accessibility Interventions with People with Aphasia. In *ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA, 2024), ASSETS '24, Association for Computing Machinery, pp. 1–17.

[8] Nevsky, A., Neate, T., Simperl, E., and Cruice, M. Lights, Camera, Access: A Closeup on Audiovisual Media Accessibility and Aphasia. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), CHI '24, Association for Computing Machinery, pp. 1–17.

[9] Nevsky, A., Neate, T., Vatavu, R.-D., and Simperl, E. Accessibility Research in Digital Audiovisual Media: What Has Been Achieved and What Should Be Done Next? In *ACM International Conference on Interactive Media Experiences* (New York, NY, USA, 2023), IMX '23, Association for Computing Machinery, pp. 1–21.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With the emergence of television in the middle of the 20th century, audiovisual media (e.g. TV, video) has steadily increased its presence in everyday life, starting as a luxury enjoyed on special occasions, entering homes with affordable television sets, and now being within reach at all times. As distribution moved onto the internet, video became embedded in platform ecosystems where people browse, search, and share across services rather than within a single broadcast schedule. Short-form and on-demand video are now part of how many people keep up with events, discuss culture, and coordinate daily routines, with platform-centred viewing being normalised on phones and tablets [149]. Audiovisual media is ubiquitous and shapes interpersonal interaction and learning about the world. As Newcomb and Hirsch explain [204], television replaced other functionally similar media, such as radio and newspapers, as a central motor for “public thought”, fulfilling a “bardic function” as a central cultural medium of contemporary societies [90]. Through this role, people express and communicate cultural identities, use television and online video as forums for debate, encounter different points of view, and interpret the viewing experience through personal, social, and ideological lenses [204]. In parallel, creator practices on user-generated platforms influence what viewers encounter and how accessibility is handled, extending discussion beyond institutional broadcasters [166, 187].

As audiovisual media has become a routine way to follow events, learn, and take part in social life, unequal access can translate into exclusion from everyday participation, as experienced by many people with disabilities. Accessibility work in this area, however, has largely been organised around sensory access (e.g. sight or hearing). These approaches do not fully address situations where the main barrier is comprehension during viewing, such as for people living with complex communication needs (CCNs). Therefore, in this thesis we try to understand how people living with complex communication needs experience audiovisual media by treating audiovisual accessibility as the ability to maintain comprehension during viewing. We present a ground-up exploration of the accessibility barriers experienced by people with CCNs and, through a research-through-design (RtD) approach [286], we develop and evaluate technological probes to examine how personalisation can improve accessibility. We step back from feature checklists to frame access around understanding during viewing rather than mere perceptibility. We treat comprehension-first personalisation as the organising idea, aiming to reduce effort while preserving agency and shared viewing. To do this, we use probes that adapt audiovisual media elements directly, rather than interventions that provide viewing support through addition. Our goal is to make everyday viewing fit the viewer by turning recurring barriers into practical adjustments to the video content, adapting it into a new, accessible form.

1.1 Access to Audiovisual Media

Audiovisual media has long been used as a cultural technology, influencing knowledge, values, and participation in public life. Television, in particular, has been described as a cultural forum that reflects and structures debate, enabling negotiation of meanings in and around programme content [90, 204]. The turn to multi-screen and the rise of internet video extended these forums, allowing viewers to watch live or on-demand and to discuss, annotate, and remix across social platforms and messaging channels [96, 324]. Indeed, television and online video remain major sources of news and political information [149]. When audiovisual access fails at scale, it limits entertainment and reduces opportunities for participation in cultural discourse and democratic decision-making [234].

Over the last several decades, accessibility policy and practice have evolved in response. Legal and technical frameworks codified rights to access digital content, driving provision of subtitles and audio description across broadcast, streaming, and web platforms [153, 301]. These frameworks produced measurable improvements for blind or low-vision and d/Deaf or hard-of-hearing audiences and advanced tooling and production processes [175, 222]. However, the field's evolution has at times been uneven. Historical emphasis on availability and coverage sometimes overshadowed the extent to which these accommodations support understanding rather than perceptibility, especially for people whose primary access barriers involve language processing rather than sensory input.

This gap between formal commitments to accessibility and their implementation emerges through distinct patterns in production priorities and platform contexts. First, authoring and delivery pipelines prioritised efficiency, throughput, and compliance, with many broadcasters and large-scale productions limiting accessibility efforts to legally mandated accommodations, with little investment in alternative forms of support [79]. Second, as video consumption patterns have shifted away from regulated broadcast to online platforms and user-generated content, accessibility practices have grown increasingly varied [187]. While regulations establish baselines for public service and commercial broadcast, platform norms for user-generated content remain limited, with creators developing ad hoc subtitling conventions without robust guidance [166]. When comprehension breaks for an individual or a community, exclusion from informational and cultural participation follows [234].

The resulting gap is both conceptual and practical. Conceptually, we lack a widely adopted framing that treats *comprehension* as something that can be deliberately designed into audiovisual media, rather than assumed to follow from sensory accommodations. By comprehension we refer to the viewer's ability to make sense of what is happening by using all available cues, including spoken words, gesture, facial expression, prosody, and the visual scene. Comprehension is grounded in situated functional communication, because understanding depends on combining these multimodal cues with context and prior knowledge to construct meaning rather than relying on word decoding alone [75]. Practically, we lack generalisable methods that apply comprehension-first design and provide evidence of when, for whom, and in what viewing contexts playback adaptations influence followability, confidence, attention, and interaction effort during everyday viewing. This thesis addresses that gap by focusing on a population for whom language processing is primary to the access problem – people living with CCNs, working with people with *aphasia* as an exemplar population. Our focus is on player-level personalisation grounded in object-based and flexible-media principles, ensuring that changes remain optional and flexible [17, 130]. While we propose design ideas, we do not develop production pipelines for real-world contexts, nor do we address clinical rehabilitation.

Aphasia is an acquired language impairment that can affect understanding, speaking, reading, and writing, and it varies considerably in nature and severity across individuals [29, 195]. Critically, aphasia affects language while leaving a person's intelligence, reasoning, and judgment intact; it is well suited as a lens through which to study comprehension-first access dissociated from sensory loss [304]. As stroke is a common cause of aphasia, co-occurring motor and fatigue-related factors can shape media use and interaction cost, placing additional demands on viewers [80, 268]. Taken together, this context motivates a re-framing of audiovisual accessibility in which *understanding* is treated as central to design, production, and evaluation. The studies examine viewing in domestic and co-viewing contexts where relevant, reflecting routine use. The remainder of this introduction develops that re-framing, defines the core concept that organises the thesis, and outlines the objectives and methodological approaches, returning to specific research questions and contributions in Section 1.4.

1.2 Access for Complex Communication Needs

Despite material progress in policy, production, and tooling, accessibility for audiences with complex communication needs remains limited in audiovisual media. A literature survey of accessibility research in human-computer interaction (HCI) by Mack et al. [181] shows a sustained focus on specific communities, technologies, and modalities. In aphasia specifically, decades of evidence-based supports exist for reading and language processing (e.g. chunking, plain language, and pace management) [154, 155], yet these principles have not been applied to video presentation or player-level features. The gap is not the absence of accommodation but the absence of a systematic orientation toward *understanding* as the goal.

The problem is amplified by changes in video production, distribution, and consumption. On major platforms, creator-generated content often lacks accessibility features such as subtitles, and when present, these vary in accuracy, timing, typography, and placement, with standards still evolving [166]. On short-form and creator-first video platforms, such as TikTok or Instagram, users develop subtitling styles that can be expressive and inclusive, but these practices are inconsistent and context-dependent, indicating a need for guidance [187]. In parallel, regulatory frameworks continue to focus on sensory accommodations while offering limited guidance for comprehension, creating a bias toward perceptibility rather than understanding [153, 301].

To motivate a design response, we draw on prior work to highlight recurring points at which comprehension breaks for audiences with CCNs, even when sensory access is nominally provided. Across studies of aphasia, media accessibility, and video interaction, difficulties often arise from high information pace and density (e.g. rapid speech, fast subtitle turnover, dense phrasing), visually and structurally complex presentation (e.g. rapid edits, camera motion, concurrent on-screen elements), and limited control over timing and structure for recovering lost context. Common access features such as subtitles and audio description primarily address perceptibility, leaving fewer supports for managing cognitive load and narrative understanding. Rather than implying a single fix, these pressures motivate a space of design strategies that can be adapted to viewing context and user needs.

We define the core concept that organises this thesis as *comprehension-first accessible personalisation for audiovisual media*. By *comprehension-first*, we mean design and evaluation explicitly tied to comprehension (e.g. followability, confidence, attention, and cognitive effort) rather than to the presence or absence of standard accessibility elements. By *accessible personalisation*, we mean system-driven tailoring that reduces cognitive effort and supports agency, grounded in user access needs and viewing contexts, and distinct from purely manual "user-tweaking" [145, 289]. Crucially, personalisation here is implemented at the level of the media object and player, where audiovisual elements such as pace, structure, layout, and mix are treated as parametric at playback through Object-Based Media (OBM) or flexible-media approaches [17, 130]. In practice, this positions domestic viewing and co-viewing as the primary settings for inquiry, with adaptations remaining optional and reversible at the point of use.

This concept is differentiated from several related ideas. It does not replace sensory accommodations, but complements them with accessibility interventions that reduce language processing barriers for audiences with CCNs. It does not rely on static universal guidelines alone, instead using flexible-media principles to adapt presentation to viewing context, aligning with work on context-specific access [138, 285]. It is not limited to technical implementation, involving negotiation among creators, producers, platforms, and viewers about artistic intent, authorship, and social acceptability. Finally, it is not clinical rehabilitation, drawing transferable approaches for audiovisual media comprehension from aphasia without making therapeutic claims.

Three themes make this concept practically relevant. First, enabling technologies now allow audiovisual media to be treated as addressable objects with metadata and rules, assembled at runtime to meet contextual and personal requirements [17, 130]. Second, related modalities show precedented success, with object-based audio personalisation demonstrating that individual listener controls can improve narrative clarity and intelligibility, supported by authoring practices [316, 317]. Third, there is demand for context-specific adaptation in real platforms and viewing contexts, especially beyond broadcast, as shown by studies of creator practices and user preferences on user-generated content platforms [166, 187]. These themes position comprehension-first accessible personalisation as a practical frame for research and design.

We work with aphasia as an exemplar within CCNs because its primary accessibility barriers are language-processing demands during continuous audiovisual playback. This lens lets us isolate where comprehension breaks under time pressure (e.g. rapid speech, dense on-screen text, busy mixes) without conflating results with sensory loss. We work with patterns and adjustable parameters, not fixed settings – we identify which adaptations are useful, when they are useful, and the effort they require in domestic viewing. This scope complements sensory access and focuses on player-level interventions for general-purpose video. We do not pursue clinical outcomes or design full production pipelines; instead we surface design commitments, parameter ranges, and interaction constraints that can inform platform and creator practice.

At a high level, this thesis examines video comprehension for CCNs as something that can be deliberately designed into audiovisual media objects rather than treated as a by-product of sensory features. It provides principles, prototypes, and evaluations that complement existing standards with comprehension-first approaches, arguing for contextual personalised support. It further identifies implications for content platforms where accessibility quality is inconsistent, showing how to move from reactive solutions toward systematic, context-aware standards.

1.3 Focus on Accessible Personalisation

This thesis centres on the viewer with aphasia and treats adaptation during viewing as a practical way to help keep up with what is happening on screen. The thesis follows an RtD approach, using iterative intervention design and study to develop knowledge about comprehension-first accessible personalisation during viewing. This keeps evaluation anchored in comprehension outcomes during viewing, since feature provision alone does not determine whether viewers can follow the content. We organise the research in four stages that answer connected research questions. See Figure 1.1 for an overview of the thesis chapters.

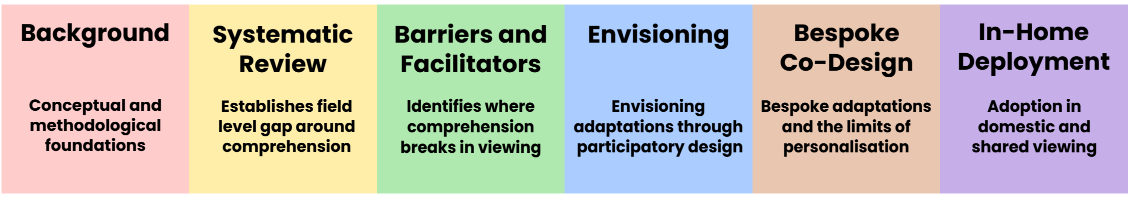

Figure 1.1: Map of thesis chapters.

In the first stage we examine where, how, and why comprehension breaks for people with aphasia during everyday viewing. To answer this research question, we identify accessibility barriers that go beyond single errors, and we relate them to language processing demands during continuous viewing. For instance, we identify that breaks in understanding often follow fast speech or dense phrasing, when edits come in quick succession, or when several visual or textual elements compete on screen. We focus on lived experiences by asking what viewers lose at those points, what they try to do to recover, and where current controls or features do not offer a workable solution.

In the second stage we work directly with people with aphasia through participatory design to envision a broad range of possible approaches to supporting everyday viewing. Here we are asking what adaptations to the audiovisual media people with aphasia want to have, in order to improve their comprehension. Since many design practices assume fluent language, which can exclude people with complex communication needs, we run workshops at aphasia charities, adapting materials and tasks so that people with aphasia can lead decision making. In these workshops we identify support tactics that participants already use, as well as what new accessibility features should be, and how they want the interaction with that feature to work. We focus on adaptations that remain understandable and meet the specific needs of our participants.

Building on the diversity of access needs and approaches seen in the second stage, the third stage focuses on a 'design-for-one' approach [118] by working independently with four people with aphasia to design their ideal, bespoke accessibility feature. In doing so, we identify possible futures for our comprehension-first personalisation approach, and try to understand the underlying tension between maintaining cultural or aesthetic values with supporting comprehension. We also document tensions that arise when personalised viewing meets current production and distribution practices, such as questions about authorial intent or acceptable degrees of change.

In the fourth stage we place a refined set of accessibility features into domestic viewing context through an in-home deployment, which includes people with aphasia who watch alone and those who watch with others, such as partners or family members. Here we focus on how adaptation choices are made and how these choices fit with viewing routines in the home. We consider how visible adaptations to the audiovisual media exist in shared viewing and whether their presence is acceptable when others are present, both for the viewer with aphasia and their co-viewers. We also investigate the adoption of comprehension-first personalisation features into viewing routines, as well as reasons for the lack of feature adoption.

Across these stages, the kinds of adaptations we explore act on how content is delivered to the viewer. The aim is not to add more to the screen, but to focus on the flow and presentation so that key information stays within reach, context is not dropped, and language that would otherwise overload a moment remains manageable. In practice, this can include easing the pace of demanding segments so that processing time matches what is needed, or it can include short cues that re-establish place in the story when attention has shifted. It can also include ways of presenting speech and on-screen text that lower reading or listening effort. These are examples rather than prescribed fixed solutions. In many cases, a small number of adaptations used together provide a workable path in which the presentation can change to fit the viewer without changing what the video is about.

Our settings and methods reflect this focus on practical use. Sessions at aphasia charity centres provide a safe environment for envisioning and trying ideas with immediate feedback. We adjust materials and instructions to match language abilities, and we keep tasks short to limit memory load. The domestic deployment then shows which features support viewing and how they are negotiated with co-viewers. We keep interactions simple and keep feedback about what changed clear so that viewers can undo actions quickly.

We collect evidence mainly through qualitative methods, identifying what viewers report losing and which changes they find helpful in the moment. We use simple Likert questions to capture perceived usefulness and preference of the different features we present. In the home deployment, we measure the real-time perceptions of our participants through ecological momentary assessments. This perception differs during co-viewing, since the use of accessibility supports becomes a social matter, with some viewers preferring adjustments that minimally interfere with the content, while others insisting on their understanding. Overall, we maintain a focus on understanding support for comprehension, rather than on sensory access, clinical therapy, or possible future production workflows.

The contributions follow from this scope, as we propose high-level facilitators for comprehension that show how content can be adjusted without assuming a single solution. We explore these facilitators by developing comprehension-first personalisation features through co-designing interventions so that preferences, trade-offs, and limits are recorded in the language of participants. We examine outcomes that shape adoption, including co-viewing in the home and underlying tensions with current production methods, user interactions, and artistic intent.

1.4 Thesis Structure and Proposed Contribution

This section gives an overview of what the thesis contributes and how the chapters fit together. The thesis examines comprehension-first personalisation for people with aphasia across four connected research stages, with each stage addressing a specific question and feeding into the next. The chapters move from evidence of a field-level gap, to a map of barriers in everyday viewing, to envisioned adaptations, to bespoke design, and then to an in-home deployment that tests use in practice.

Chapter 3 reports a systematic review of accessibility interventions for digital audiovisual media, identifying which challenges are addressed, and which interventions dominate, showing the limited attention given to comprehension and to cognitive or language-related access. This sets a baseline and motivates the need to study comprehension during viewing rather than only sensory access, as well as giving a reference set that later chapters use when positioning the work.

Chapter 4 corresponds to the first research stage, examining where, how, and why comprehension breaks during everyday viewing for people with aphasia. The chapter combines an online survey with a focus group and reports a taxonomy of accessibility barriers and facilitators. Breaks are associated with fast speech and dense phrasing, with rapid editing, and with competing on-screen elements that raise effort and reduce followability. The outputs are concrete targets for adaptation and a small set of simple measures for perceived effort, followability, and recovery that appear again in later chapters.

Chapter 5 corresponds to the second research stage, using participatory design to envision a broad range of accessibility features with people with aphasia to support their viewing. Workshops at aphasia charities and a probe-based activity run at home elicit which adaptations matter, producing mid-fidelity prototypes and recording criteria for perceived usefulness. These results narrow the space of ideas and define a small group of candidate features that can be carried forward.

Chapter 6 corresponds to the third stage and applies a 'design-for-one' approach. The chapter works individually with four participants to co-design their ideal, bespoke accessibility features. It records where content changes raise concerns about tone or intent and where alternatives are seen as legitimate ways to preserve meaning. These results sharpen the choices for deployment by identifying which features are stable enough to try at home and which require further design work.

Finally, Chapter 7 corresponds to the fourth stage and examines everyday use through an in-home deployment. The deployment includes solo viewing and co-viewing with family members or partners, using in-the-moment ecological momentary assessments during viewing to measure how easy the overall viewing experience was. The chapter reports patterns of use, how viewers negotiate visible changes in shared viewing, and why some features are adopted while others are avoided. It also records constraints that matter for uptake, such as interaction effort, predictability, and how visible changes match with household viewing routines.

Taken together, the chapters contribute an initial mapping of accessibility barriers in video for people with aphasia, an initial set of high-level facilitators for comprehension, and a co-designed exploration of interventions that express those facilitators in practice. The work also examines outcomes around the viewer that shape adoption, including the dynamics of co-viewing in the home. Across chapters, the results are expressed as families of adaptations rather than fixed features, with parameter ranges and control patterns that can be further refined. The studies provide an evidence model that is primarily qualitative, supported by small complementary measures that track perceived usefulness. The overall contribution is a coherent path that begins with the points where understanding breaks, moves through ideas and bespoke designs, and concludes with observations of real use in domestic viewing.

1.4.1 Author's Contribution

This thesis reports work undertaken within the CA11y project, from 2023 to 2026. All work presented in the thesis was led by the author and mostly includes work that has been published in peer-reviewed conferences, with the deployment chapter awaiting reviews at the time of writing – see Appendix A for more detail. The list of publications associated with this PhD and the CA11y project (see list of publications) also includes papers on which the author was not the lead. For these papers, the author contributions included support with the planning of workshops and interviews, participation in the workshops and interviews, taking part in data analysis, and contributions to drafting and revising the published papers. As the author did not lead these papers, they are not included in the thesis body. Findings and discussions from this work are, however, drawn on in Chapter 8, where they address questions posed in the thesis. One additional paper has been submitted and was awaiting peer review at the time of writing. The thesis author defined the research direction for the work with guidance from supervisors, designed the studies, produced the study materials, conducted the data collection, managed the research data, carried out the analysis, implemented the prototypes, and wrote the papers that underpin the thesis. Much of this work was done in close collaboration with the first supervisor and the post-doctoral researcher on the CA11y project.

The thesis makes a contribution to human-computer interaction research on audiovisual accessibility by developing a comprehension-first account of access for people with complex communication needs, working with people with aphasia as an exemplar population. The author developed the analytic framing used across the thesis work, including how comprehension breakdowns are described, how candidate adaptations are characterised as playback changes, how usefulness is assessed during viewing, how evidence is carried across studies without treating any single feature as a universal solution. The thesis contributes evidence about where comprehension breaks during everyday viewing for people with aphasia, describing recurrent sources of difficulty that include pace, density, rapid scene changes, competing on-screen elements, and limited opportunities to recover lost context, as well as recording the tactics viewers use to cope with these situations.

1.4.2 Positionality Statement

This research was conducted by the author as part of a doctoral programme at King's College London, within the Department of Informatics. The work was situated in an HCI-oriented and interdisciplinary research environment, with collaboration involving academics from speech and language therapy. As a PhD student, the author led the research reported in this thesis, including study design, data collection, analysis, and synthesis, with guidance and discussion from supervisors and collaborators at various stages – please see Appendix A for more detail.

The author's academic background is in computer science. This training shaped an approach that values clarity, structure, and systematic organisation of evidence. In the context of qualitative synthesis and participatory design research, this orientation influenced how studies were compared, how analytic categories were defined, and how design space was articulated. Across the thesis, RtD frames the intervention artefacts as a way to examine comprehension during viewing, with attention to where current accessibility practice does not support understanding in use.

The research engages with disability, and specifically aphasia, from the position of an outsider to the lived experience discussed. The author does not have aphasia and does not claim experiential authority over the phenomena described. At the same time, the work involved direct engagement with the aphasia community through participatory design activities and sustained contact outside formal research settings, including volunteering with an aphasia support charity. These engagements informed how published accounts were read and interpreted, and supported an approach grounded in listening to how disabled participants described their own needs, practices, and constraints.

Across the thesis, failures and limitations identified in technologies are treated as systemic and design-related, rather than as individuals' shortcomings. Interpretive decisions tended to align with this perspective, particularly in cases where the literature presented multiple plausible readings. The analysis reflects the author's position as a researcher learning from disabled participants and from prior work, rather than speaking on behalf of the community or claiming representational completeness. The findings and interpretations presented in this thesis should therefore be read as an informed and situated account shaped by the author's disciplinary background, institutional context, and mode of engagement with the research community. The work does not aim to provide a definitive characterisation of aphasia-related technology use, but to contribute an organised synthesis intended to support further design and research.

Chapter 2

Background

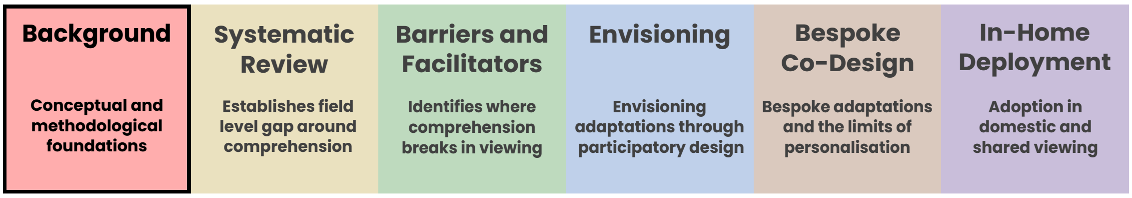

Figure 2.1: Map of thesis - background

This chapter provides a review of foundational concepts and relevant prior research to contextualise our work – see Figure 2.1. Our aim is to trace how accessibility practices have evolved, situate the everyday experiences of people living with aphasia in their interactions with audiovisual media, and introduce the conceptual and methodological frameworks that underpin our approach. Audiovisual media plays a central role in cultural, social, and informational participation, yet its multimodal and temporal complexity creates persistent accessibility challenges. While accessibility interventions, such as subtitles and audio description, have addressed sensory barriers, they often fail to support comprehension for people living with complex communication needs (CCNs). At the same time, technological shifts – such as immersive media, user-generated content, and flexible production pipelines – open new possibilities for personalised and participatory approaches to accessibility.

In Section 2.1, we survey accessibility in audiovisual media, tracing historical interventions, emerging modalities, and systemic gaps in current practice. Following this, in Section 2.2 we introduce personalisation as an accessibility strategy, defining its conceptual foundations, technical substrates, and design principles, and illustrating how it can adapt audiovisual content to diverse needs. Section 2.3 turns to inclusive participation, outlining why 'one-size-fits-all' approaches fail, and presenting participatory design (PD) and its adaptations for CCN communities, with aphasia as an exemplar. We address ethical and methodological considerations in Section 2.4, including privacy, editorial integrity, and social acceptability.

2.1 Accessibility in Audiovisual Media

Accessibility for digital audiovisual media (e.g. television, online video) developed gradually, beginning with pragmatic workarounds in broadcast, followed by legal and technical baselines, and later extending to web and streaming. These advances in the accessibility of audiovisual media aim to address the underlying complexity of the medium, which combines multiple information streams (e.g. visual, audio) that are integrated over time. Each stream of information can introduce accessibility barriers for audiences with disabilities, such as accessing audio information for people who are d/Deaf or hard-of-hearing (DHH) or visual information for those who are blind or low-vision (BLV).

2.1.1 Subtitles

Early accessibility efforts concentrated on subtitling for DHH viewers, which emerged from practices such as "surtitles" or "supertitles" in the silent films of the 1920s [219] – frames of text that were inserted between segments of video to let viewers read the conversation between the on-screen characters or to give commentary. Although these sur- or supertitles were not designed as accessibility tools, they enabled DHH audiences to follow the content. The rise of 'talkies', that is films that also included an audio element, introduced initial barriers for many DHH audiences as sur- and supertitles had become seen as obsolete. It is not until the 1970s that subtitles [1] –

[1]The terms 'subtitle' and 'captions' refer to similar but distinct interventions. Subtitles refer to the transcription of dialogue, while captions transcribe all meaningful audio, including sound effects or

on-screen text that transcribes audio conversations – were introduced on broadcast television through systems like Teletext and open-captioning standards [136].

Initial research focused on establishing workflows for subtitle presentation, including line length, on-screen placement, or timing heuristics. For instance, Braverman and Hertzog [39] examined how reading rates influenced comprehension and viewing performance, outlining the trade-off between the speed of the subtitles and their usability. Similarly, Kruger et al. [158] show how faster subtitle speeds increase re-reading and regressions, raising processing demands. These presentation conventions reflected an underlying challenge that subtitles introduced – subtitles enable access to dialogue but also introduce additional cognitive demand, which can reduce the overall viewing experience. For this reason, it was important to present subtitles in a way that both let the viewer read the dialogue and allowed them to look at the visual element. For example, Nugent [210] found that deaf and hearing students learned most effectively when visuals and subtitles were presented together.

With advancements in technical capabilities, focus has shifted towards ways to make subtitles more dynamic and work better with the content being watched. One way in which this can be achieved is by moving the subtitles away from their standard placement at the bottom of the screen to better integrate them in the viewing experience. For example, work by Brown et al. [40] and Crabb et al. [64] explore the use of dynamic subtitle positioning based on the underlying content, which can lead to a more immersive viewing experience. Eye-tracking technologies have also been used by Kurzhals et al. [160] to position subtitles near the viewers' focus while not overlapping important visual information, finding such an approach supported engagement. Subtitles have also been expanded to convey additional narrative information, such as the prosody of speech or the underlying emotional tone. Ohene-Djan and Shipsey [213] explore this by changing the text font to represent emotions, while Rashid et al. [239] animate the subtitles by manipulating elements such as the subtitle size, adding visual shake, or adjusting the opacity of the text. These concepts can be put together, such as in the work of de Lacerda Pataca et al. [73] which uses subtitles to represent both emotions and emphasis, improving understanding.

music. In this thesis, I will be using the term subtitle as that is the preferred term in British English [24]

The rise of novel viewing mediums, such as smartphones or immersive video using virtual or augmented reality (VR/AR) headsets, has also introduced new possibilities for how subtitles can be used. Work by Gerber-Moron et al. [101, 102] demonstrates the constraints small smartphone screens impose on subtitles, leading to poor readability and the need to rethink how line breaks are displayed. One way of addressing this constraint is introduced by Lee et al. [163], who develop an automated system to manipulate font size and line breaks using speech pause-driven subtitle chunking. When it comes to immersive environments, readability can be made more difficult by having to consider the placement, persistence and legibility of the subtitles under motion. Rothe et al. [256, 257] investigate the advantages of dynamic or 'in-world' subtitle positioning that appear next to speakers, which generally improved speaker identification, increased presence, and reduced eye strain, workload and motion sickness. Static subtitle positioning, on the other hand, was found to give viewers more freedom to explore the immersive experience. To improve navigation with dynamic subtitles, Orero et al. [218] and Hughes et al. [131] both explore ways to guide viewers towards the speakers, such as through arrows, radar, or auto-positioning. More recent work also describes creator-side tooling and machine learning methods that streamline subtitle authoring in user-generated ecosystems [278].

2.1.2 Audio Description

Similarly to subtitles, accessibility for BLV audiences developed through audio description (AD). Audio descriptions are verbal narrations inserted into audiovisual content to convey essential visual information, such as actions, settings, or facial expressions. Early practices date back to the onset of 'talkies' in the 1920s with simultaneous radio broadcasts, in which radio announcers would describe the visual information to the listeners [81]. Audio description as a structured practice emerged in the 1970s, with initial efforts to develop standardised techniques for describing visual media, which led to the first formal audio-described theatre performances and television programmes [26].

Over time, and with technological advances changing the way we consume audiovisual media, the approach to AD has shifted from an objective ‘fill the gaps’ narration to a richer and more stylistic approach. This evolution reflects a shift from purely factual descriptions toward approaches that incorporate narrative cohesion, intertextuality, and cinematic style, acknowledging that AD can function as an interpretive layer rather than a neutral commentary. As Elisa Perego puts it “*conveying film language through AD [...] not consisting only of characters, actions and setting [...] is seen as the only way to respect the aesthetic and narrative integrity of the original*” [228]. With this shift, researchers have explored the effect of more creative AD approaches. For example, work by Walczak and Fryer [312] finds that creative descriptions were more engaging and improved presence. Additionally, audio description has been enhanced by incorporating additional audio information, such as Lopez et al. [178] using sound design to include elements such as sound effects or binaural audio. These changes to standard AD practices, however, are context-dependent, with listeners preferring different types of descriptions for different types of content, for instance preferring having additional information in the audio track for informational content (e.g. news, learning materials) while having tonally appropriate sound effects for humorous content [138].

While there are efforts made in improving AD, there are significant barriers surrounding the creation process, leading to limited availability of AD content for television broadcasts or video-on-demand (VOD) platforms. For instance, in the UK, across all BBC broadcast channels, 35.0% of all air time was audio described, while 23.8% of BBC iPlayer programmes have AD [25]; both of which are above the 20.0% target set by the BBC [244]. The main challenges in creating these audio descriptions are cost and production bottlenecks, with many AD teams working disjointly from the rest of the production [314]. To address these production challenges, novel technologies and collaborative work have been employed to develop tools that help reduce costs and streamline the process. For instance, work by Stangl et al. [284] and Yuksel et al. [329] explore the use of artificial intelligence (AI) and human-in-the-loop methods to automate AD creation, finding such approaches result in AD that are viable while significantly reducing production efforts. Alternatively, Natalie et al. [199] developed a system that enabled collaborative authoring of AD, in which novice and experienced describers could author descriptions. Collaborative approaches have also been explored in the context of live video, with Killough and Pavel [150] developing a community-led AD authoring tool for the live-streaming platform Twitch.

2.1.3 Emerging Viewing Practices

With the rise of novel platforms and changing patterns of media interaction, new realities are emerging in how access services are delivered. Web-based platforms and user-generated content have reshaped media consumption, changing how video is produced, distributed, and accessed. This shift has prompted a rethinking of accessibility practices, as traditional broadcast norms give way to decentralised, participatory models. Smartphone proliferation has been central to this transformation, enabling viewers to watch content anytime and anywhere. Rigby et al. [241] show that mobile viewing introduces new experiential dynamics, where smaller screens reduce immersion compared to larger displays, yet convenience and portability drive widespread adoption. Similarly, AL-Zoubi [13] reports on how smartphones have become the dominant medium for media consumption, with 85% of users preferring them over traditional devices, reinforcing the need for mobile-friendly accessibility strategies.

However, this shift introduces new challenges. Accessibility practices are often inconsistently applied across platforms, particularly in user-generated ecosystems. Li et al. [166] finds that subtitling on YouTube varies widely in accuracy, timing, and presentation, directly shaping the experiences of DHH audiences and their understanding. McDonnell et al. [187] observes similar issues on the short-form video platforms TikTok, where emergent subtitling cultures lack stable standards for comprehension quality at scale. Even institutional channels are not immune, with work by Acosta et al. [10] highlighting systemic shortcomings in educational videos published by universities, revealing low compliance with accessibility guidelines despite formal obligations.

Beyond static video, new interaction paradigms complicate accessibility further. Unlike pre-recorded video content, live-streaming video unfolds in real time, making it difficult to provide accessibility support, such as accurate subtitles or audio descriptions without a delay. Features such as rapid interaction layers and dynamic overlays introduce barriers for viewers who depend on consistent placement of accessibility elements [116]. At the same time, evolving viewing practices such as second-screen use create attention competition. Rooksby et al. [252] documents how mobile devices interweave with television viewing, while Chambers [54] shows that multi-screen households generate new temporal experiences, challenging assumptions of linear engagement and raising questions about how accessibility can adapt to fragmented attention.

The viewing experience has also shifted away from simple and linear content towards more dynamic interactions. The use of dual screens, such as using a smartphone while watching television, introduces additional complexity and demands careful management of attention and information density to avoid cognitive overload [201]. Furthermore, advancements in technology have enabled novel social contexts, such as watching with others at a distance using VR headsets [190]. For instance, Alper [14] explored environmental personalisation for shared viewing through relaxed screenings for autistic audiences, which offers transferable viewing patterns for domestic settings.

Beyond these shifts in context, the accessibility of video content is strongly mediated by genre and format. Dementia-friendly television news designs by Funnell et al. [94] show that factors such as pacing, presenter cues, and technical aspects influence comprehension, and that incorporating visual and textual scaffolding helps reduce barriers and improve understanding. Similarly, learning-oriented genres, such as podcasts and lectures, demonstrate navigation and comprehension challenges, with Jones et al. [141] exploring persistent difficulties in podcast information access. Park et al. [227] explore how segmenting content with topic labels and visual cues significantly improves browsing and understanding. These findings suggest that analogous strategies (e.g. chaptering, summaries, and signposting) could simplify navigation and interaction with video-based media.

In contrast, live or rapid formats introduce unique challenges, such as real-time subtitles struggling with speed and placement. Interactive techniques (e.g. pausing or highlighting) help viewers maintain their place in transcripts, improving comprehension during fast-paced, multi-modal content [162]. Amin et al. [16] built on this by examining multi-speaker scenarios, finding that dynamically positioning subtitles and incorporating speaker-identifiers tailored to on-screen interactions can significantly improve understanding. These examples emphasise the need to evaluate accessibility practices not merely by their presence, but by their adaptability to immersive and context-specific viewing experiences [218].

Alongside these shifts in viewing practices, researchers have also explored accessibility interventions beyond 'standard' ones (i.e. subtitles and AD), developing different modalities to support viewers. Sign language interpretation – a display showing a person transcribing spoken speech into sign language – can provide an alternative to subtitles [318]. This requires, however, a sign language interpreter to transcribe the entire content, which, similarly to AD production, can be costly and time-consuming. To address this, researchers have explored automated approaches, such as using machine learning to generate skeletal representations of sign language interpreters that can animate avatars [261]. Additionally, work by Vinayagamoorthy et al. [307, 308] investigates how sign language interpreters can be presented in immersive video content, exploring different configurations in VR environments.

The use of haptic and tangible augmentations has also shown potential in supporting BLV audiences by rendering visual content in a tangible manner. For example, Viswanathan et al. [309] demonstrate the positive effect of adding haptic feedback to convey the positions of on-screen characters, including their direction and distance from the camera, using a haptic belt. Similarly, Guinness et al. [112] achieve this through small robots that visualise on-screen elements and provide haptic feedback to the viewer. Haptics have also been used by de Lacerda Pataca et al. [72] to augment subtitles with emotional cues. While such interventions help visualise elements in the video, they introduce additional complexity and interaction, expanding standard viewing practices in the process. These developments point to the broader trend that viewing is no longer a passive, linear experience, but is increasingly interactive.

2.1.4 Standards and Their Limitations

Accessibility interventions for audiovisual media typically centre on subtitles and AD, which broadcasters and platforms prioritise and researchers focus on – we expand on this claim in Chapter 3. Technical and structural constraints on broadcasters and platforms, as well as the legal and regulatory requirements have pushed for a standardisation of interventions which, currently, excludes alternative approaches to accessibility. This exclusion of alternative approaches often leads to the exclusion of various under-represented disabled communities, who receive less consideration than 'typical viewers'. For instance, the Web Content Accessibility Guidelines (WCAG) [153] sets baseline requirements for websites and their content, including any video content, which focus on allowing people to reach information. However, the guidelines only require 'standard' interventions (i.e. subtitles/transcripts, AD, or optionally sign-language interpretation), along with compatibility with assistive technologies and interaction modalities. They do not, however, require support for non-sensory disabilities, such as cognitive or language impairments. Similarly, regulatory instruments, such as the European Accessibility Act [301] or the US Section 508 [86], do not impose obligations for comprehension support. Moreover, the implementation of these guidelines and regulations often optimises for feature presence and coverage, rather than outcomes and usability [250, 292].

Critiques of this approach propose re-framing accessibility to be understood as part of a broader human rights framework for participation instead of a technical add-on. Standards often focus on measurable compliance without addressing why the features they require matter (i.e. preventing exclusion). Greco [106] suggests that accessibility should not be limited to providing technical access, as it must ensure that individuals can fully enjoy and participate in the content. Accessibility should be embedded as a condition for equitable participation in cultural, civic, and social life, not merely treated as technical compliance. Ellcessor [78] further argues that this positioning means moving beyond accommodation toward a proactive approach that anticipates human variation and integrates usability and universal design principles, ensuring that access is inclusive, not a regulatory minimum.

Moreover, accessibility features originally designed for disabled audiences often become widely adopted, demonstrating their broader social value. Subtitles, for instance, have evolved from a specialised accommodation into a widely used tool that benefits diverse audiences, including those in noisy environments or engaged in language learning. Morris et al. [194] report that subtitles improve comprehension and retention for students with and without disabilities, framing them as a principle of universal design. This pattern mirrors the 'curb cut effect' described by Hamraie [117], where ramps, initially intended for wheelchair users, also serve parents with prams, travellers with luggage, and delivery workers. These examples demonstrate that accessibility is not a marginal concern, but a foundation for inclusive design that benefits all users.

Building on these critiques, scholars have also pointed to emerging priorities for accessibility in new media contexts. Increasingly, the focus has shifted toward scalability through automation and templating, creating opportunities to integrate relevant metadata, including comprehension-oriented metadata, into audiovisual content workflows [130]. Realising this potential, however, requires creator-side tools and guidance to become commonplace in user-generated ecosystems. At present, such tools are either absent or highly inconsistent, leaving accessibility largely dependent on individual practices [166, 187].

Some, such as Orero et al. [218], have suggested an accessibility-first approach to new audiovisual media, arguing that fit-for-purpose interventions should be embedded by default rather than treated as optional additions. This principle extends beyond sensory accommodations to include strategies that support comprehension for all viewers. Yet current HCI research often falls short of this ambition. In their overview of accessibility research in HCI, Mack et al. [181] note that many disabled communities – such as people with complex communication needs – remain under-represented. This lack of inclusion perpetuates design assumptions that overlook diverse communicative realities.

Evidence from work with people with aphasia, for instance, illustrates what is possible when these communities are considered. Indeed, Knollman-Porter et al. [155] identify effective supports such as plain language, chunking, and pacing, yet observe that these strategies rarely translate into audiovisual media. Similarly, Cartwright and Elliott [51] show that interventions at the level of structure and pace can significantly improve television comprehension for individuals with language impairments. Including such approaches in mainstream audiovisual media design has been argued to not only enhance usability, but also advance communicative justice – organisation of communicative practices and media forms that permit meaningful participation by people with differing communicative capacities without requiring conformity to dominant norms of fluency [14, 104]. These critiques point to a mismatch between what standards measure and what viewers need in order to follow a programme. Therefore, this thesis assesses accessibility through comprehension during viewing.

2.2 Personalisation as an Accessibility Strategy

Audiovisual media integrate multiple streams of information, including speech, music, ambient sound, visual action, and on-screen text, presented over time. Even before disability-specific barriers are considered, this integration establishes a baseline of cognitive and perceptual load. Viewers must coordinate attention across different modalities under time pressure, yet this coordination is rarely addressed as a variable within design practices – it is commonly treated as a fixed property of the medium. By interacting with underlying elements of audiovisual media (e.g. timing, structure, and emphasis), viewers can influence how the audiovisual media are rendered, thus changing how they are comprehended by diverse audiences. These mechanisms enable personalised audiovisual media, functioning as a means of participation because platform and interaction choices influence who can follow the content. For instance, van Dijck et al. [303] describe how platform infrastructures encode hierarchies and dependencies that privilege certain forms of engagement, framing access in terms of citizen well-being instead of narrow consumer welfare. Personalisation can be a way to manage complexity and improve comprehension for diverse audiences.

The distinction between *adaptive* interfaces that change themselves and *adaptable* interfaces that users configure emerged in HCI from efforts to balance efficiency with control. Early work on split menus showed clear benefits in speed when the system adapts to frequent actions, yet it also revealed a cost when changes are not predictable or clear to the user [88]. In accessibility research this trade-off is often more pronounced. Changes that aim to ease access can be helpful, but they also introduce questions about transparency and control when the user does not know what has happened or how to restore a preferred configuration. Similar questions about automation and control appear in later work on media personalisation, making these early models relevant beyond interface design.

We use the term *personalisation* to describe changes to a media experience that adapt the content or its presentation to a viewer and their context, creating a new adaptation rather than simply restyling the original. This personalisation adapts content to reflect an individual's interests, history, and relationships [220]. In accessibility terms, this focuses on people's lived experience, recognising that individuals in different contexts require different interventions to follow what is happening. A useful distinction comes from communication research, which separates *customisation* – adjustments driven by the user – from *personalisation* – adaptations that may be driven by the system using profile or contextual data. Sundar and Marathe [289] suggest that agency with how content is assembled matters, allowing users to improve their experiences. This preference for more control was dependent on privacy expectations, revealing a tension between control and convenience. Kang and Sundar [145] extend this by showing that customisation can produce a "self-as-source" effect, where users perceive themselves as co-authors of the message, co-creating a new media experience that is accessible to them. If a viewer with language or attention difficulties can decide which speakers are subtitled, or when summaries appear, they are not just consuming access features, but are helping to produce an adaptation of the content that works for them in the moment. Such distinctions show that personalisation involves not only technical adjustments, but also changes in how viewers interpret the content.

Personalisation is often most useful when it fits the individual instead of aiming for a single ‘universal’ solution for everyone. For example, Hornof et al. [126] co-designed adaptive interactions with children that had Rett syndrome, showing that fine-grained adjustments to interaction complexity reduced cognitive load, demonstrating that adaptive behaviour can support participation when interaction demands are calibrated to individual capabilities. At the same time, systems that change on a user’s behalf can raise questions about what information is used to trigger a change and how that change can be inspected or reversed. In the context of personalisation for accessibility, these questions affect whether an adaptation is usable during routine viewing. If the effort of understanding or undoing a change is high, users may avoid the feature even when it could improve comprehension. Progressive disclosure addresses this by embedding short explanations and simple controls at the moment when the change occurs, offering deeper detail only when the user seeks it. Wolf and Ringland [325] examined explainable AI interfaces in accessibility contexts, proposing layered explanations that adapt to user needs. This interface incorporated progressive disclosure, which reduced initial complexity, enabling users to understand and override automated decisions when necessary. These studies illustrate how transparency has been approached in adaptive systems and why these concerns remain relevant for audiovisual media personalisation.

To reduce interaction effort while keeping user control, personalisation systems often use recommended settings and defaults. HCI nudging research shows that people accept defaults when they save effort and are easy to change, and respond better when the reason for a recommendation is clear. Caraban et al. [49] explored interface nudging, finding that defaults combined with context-aware explanations reduced resistance to automated adjustments. Springer and Whittaker [282] extended this by analysing adaptive system design patterns, showing that timely feedback with clear explanations for changes improved trust and willingness to engage with the automation. Kong et al. [156] find that situational reading support also points to the value of reducing micromanagement while keeping adaptations visible and readily overridable, which helps users maintain a sense of control. These findings indicate when adaptive behaviour tends to be accepted, informing later work on media interfaces.

In audiovisual media, these classic models have informed how researchers and industry groups think about personalisation within production pipelines. Armstrong et al. [17] describe object-based approaches in which programmes are represented as collections of addressable components linked by metadata and rules. Through such an approach, aspects of timing, structure, layout, and mix can be adjusted at the point of consumption without re-authoring a single fixed version. Surveys of practice report that such approaches require metadata that carries narrative and semantic information from authoring through delivery, and that tooling needs to align with existing production workflows so that describing adjustable regions does not add undue overhead [58, 130]. These observations position object-based approaches as a technical foundation for personalisation.

This re-framing is particularly important when personalisation goes beyond surface user interface (UI) tweaks to re-time, re-structure, or re-mix parts of the various audiovisual elements. Object-based media (OBM) provides the technical substrate for such adaptations. Armstrong et al. [17] describe OBM as packaging media as a set of audiovisual assets (e.g. video clips, audio tracks, subtitles, sound effects), along with accompanying metadata and rules. These assets can then be assembled at playback to create a version of the content that responds to the viewer's preferences, device capabilities, and environmental context. The British Office of Communications (Ofcom) similarly identify OBM as a near-term route to more responsive experiences, recommending industry tooling and interoperability standards to support its adoption, outlining the potential such interactions have in making audiovisual content more accessible [130]. OBM is considered here as an enabler rather than an end in itself, as it carries the controls for timing (e.g. adaptive pacing), structure (e.g. chapters, summaries, speaker cues), and emphasis (e.g. balancing dialogue and music) from authoring through to playback.

Building on these technical foundations, prior works have explored how OBM principles support adaptive features in practice, finding that preferences vary in real-time, with different viewers valuing different presentation choices. For example, work on metadata-driven adaptive subtitles by Gorman et al. [105] suggests that personalisation is not a single setting, but a group of adjustments that respond to the video content and the context in which they are viewed. Other projects have explored personalised narrative presentation, for example Concannon et al. [60] designed a personalised film that integrated open data about social care with geographically relevant information, adapting the narrative to the viewers. Their study shows how object-based techniques can be used to personalise storytelling, supporting engagement without disrupting the narrative structure.

Seen through a translation lens, personalisation can be read as a form of intra-language localisation – not just translating words, but adapting semiotic resources such as prosody, layout, montage, and sound design to the receiver and context. Media studies describe production-time localisation that shapes content for accessibility as domestication [55, 57]. Chaume [56] further argues that digital transformation has widened both the palette of techniques and the role of audiences, with localisation and "making translators visible" changing what counts as a faithful rendition. If we treat accessible personalisation as a kind of transcreation for comprehension, then a personalised cut or mix is not a deviation from authorial intent – it is a pragmatic rendering to preserve narrative meaning under varied constraints. translation and accessibility research further situates personalisation within broader adaptation practices. For instance, Chaume [56] note that audiovisual media is often translated, localised, or creatively adapted. Accessibility personalisation features can be understood within this broader group of practices that produce new renditions for different audiences and viewing contexts. Research in accessible filmmaking extends this idea into production contexts and argues that decisions about pacing, audio balance, and signposting influence reception for many audiences. Romero-Fresco [251] explains how, through such practices, accessibility and personalisation can be situated within standard editorial practice, rather than as simple post-hoc additions.

Extending these editorial considerations into technical practice requires infrastructure that supports personalisation across production and delivery. Supporting adaptable media requires production pipelines that carry semantic and narrative metadata from authoring through to playback. Armstrong et al. [17] describe object-based workflows that make timing, structure, and mix adjustable without re-authoring fixed versions, while OFCOM [130] identifies metadata continuity as essential for interoperability. Ursu et al. [302] and Cox et al. [63] note that toolchains need to make it feasible for creators to specify adaptable regions without prohibitive overhead. At the delivery end, platform players require APIs that expose controls safely and consistently, ensuring that personalisation features remain accessible to end users and compatible with assistive technologies.

2.2.1 Audiovisual Personalisation

Personalisation in audiovisual media has been explored through adaptive techniques that modify the underlying content, adjusting the presentation of content so that it better supports the needs of diverse viewers. One area of work focuses on pacing, with Gorman et al. [105] examining adaptive playback strategies that allow viewers to slow down or re-time segments, together with pacing cues that respond to reading speed and preference. Beyond timing, structural adaptations can help viewers maintain orientation during complex narratives. Hughes et al. [132] describe responsive subtitle rendering that varies layout and typography, alongside chaptering and summary overlays that help viewers maintain a sense of progression. These interventions move beyond static placement and provide navigational scaffolds within the audiovisual experience.

Adjustments to audio prominence extend this principle to the audio mix, where competing elements can obscure dialogue. For example, Ward and Shirley [316], Ward et al. [317] investigate audio controls that allow users to increase the volume of narratively important audio objects (e.g. speech), while lowering the volume of background music or sound effects. Their findings indicate benefits for DHH viewers, suggesting that fine-grained control over auditory elements can improve intelligibility, while reducing interaction complexity.

Using personalisation for accessibility can also span across multiple modalities. Natalie et al. [198] investigate adjustable audio description that allows viewers to control granularity, timing, and mix of descriptive content, finding that fine-grained control improved comprehension and engagement, particularly for viewers with cognitive or language impairments, who benefit from the ability to reduce descriptive density during fast-paced scenes and increase detail in complex sequences. Vinayagamoorthy et al. [308] investigate augmented reality delivery of sign language interpretation, using spatial placement to minimise occlusion. Other work considers haptic and tangible augmentations that render visual or narrative cues through touch, supporting diverse attentional and cognitive profiles. Environmental and social contexts also shape how viewers interact with accessible personalisation practices. Schirra et al. [262] describes second-screen interactions that enable branching choices and temporal agency, while social backchannels such as live-tweeting create shared experiences around broadcast content.

Not all adaptations require deep changes to the audiovisual content, however. Viewers can achieve satisfactory access by adjusting surface-level presentation variables (e.g. size and position of elements, simple audio balance). Such adjustments represent customisation, giving viewers control over presentation without changing the structure or meaning of the content. When understanding depends on altering the internal relationships, not merely its surface appearance, the adaptation moves from customisation to personalisation. Armstrong and Glancy [18] describe this distinction in the context of OBM, noting that some adaptations require access to the underlying representation of the content. For example, slowing down a single scene or inserting additional visual or audio cues for orientation cannot be achieved through simple interface controls. These changes require a system that can reassemble the media at playback.

This distinction depends on factors such as genre, the viewing context, and the profile of the viewer. A fast-paced drama may demand structural interventions for some audiences, while a lecture might only require adjustments to text size or audio balance. Tooling that supports both viewing patterns can accommodate this variability, enabling lightweight changes where they suffice and deeper adaptations where they are needed. These observations raise questions about usability, as personalisation depends on strategies for control and automation that keep adaptations transparent and reversible without introducing unnecessary complexity.

Control and automation can be combined to make personalisation usable in practice, with high controllability tending to support sustained adoption in sensitive tasks. For example, Roy et al. [258] show that reversible and lightweight edits help build trust, as users can experiment without fear of permanent change. Progressive disclosure has been proposed as a way to manage this complexity. Wolf and Ringland [325] describe how explanations and controls can be layered to avoid overwhelming users, a principle that parallels progressive disclosure in adaptive systems and frames complexity as something that should emerge gradually. This approach allows deeper detail to appear only when requested, reducing early cognitive load.

Alternatively, having reliable default settings also play an important role in shaping experience. Caraban et al. [49] note that defaults are most effective when they save effort and respond to context, yet they should not become 'sticky' in ways that limit agency. The ability to opt out remains essential for maintaining a sense of control. Pang et al. [225] observe that gentle scaffolding and immediate payoff encourage engagement, suggesting that systems benefit from clear feedback and low barriers to entry.

Beyond technical feasibility, adoption depends on how interventions are perceived in social settings, where audiovisual media is often shared. Social acceptability becomes critical in these contexts, as features that signal difference can influence willingness to use them. Features that signal difference can carry stigma, influencing whether viewers choose to use them in public or shared settings. Curtis et al. [69] show that visibility and aesthetics strongly affect willingness to use personalised supports, with designs perceived as medical often avoided outside private spaces. Building on this, they explored re-framing assistive functions as desirable, negotiating visibility and self-image through design choices that emphasise both style and integration [66]. Similarly, Pullin [235] argues for an aspirational approach to assistive technology, suggesting that framing devices as objects of design rather than clinical tools can reduce stigma and promote uptake.

2.3 Inclusive Participation

The previous section examined personalisation as a way to manage complexity in audiovisual media. This section considers how accessibility research has often treated access as a matter of perceiving content, leaving less attention to practices that support participation when understanding is at stake. Indeed, reviews of accessibility literature in HCI show an over-representation of sensory access [181, 305].

These gaps matter because challenges in comprehension depend on features of audiovisual media such as pacing, segmentation, and cueing. Ellcessor [79] argues that accessibility cannot be reduced to technical compliance as it is embedded in socio-technical systems and cultural norms that privilege some users while marginalising others, thus calling for approaches that centre lived experience. Kaur and Saukko [146] extend this critique to digital media practices, showing that online platforms can both mitigate and reinforce social exclusion. Their concept of social access emphasises that digital affordances interact with offline inequalities, enabling participation in some contexts while leaving structural barriers intact.

Additionally, a broader view recognises that comprehension challenges are not confined to disability categories. Older adults, linguistic minorities, and migrants encounter similar barriers when media practices or contexts change, such as shifts in platform norms or genre conventions [106]. This perspective situates accessibility within a social framework, where interventions anticipate variation in communication capacities rather than treating it as an exception.

The social model of disability builds on this structural orientation, with Oliver [216] framing disability as a product of social and environmental barriers rather than an individual deficit, shifting responsibility from personal adaptation to systemic change. Shakespeare [266] supports this emphasis on rights and participation but argues for an interactional view that acknowledges both impairment and context. This relational framing informs contemporary accessibility debates, where solutions are understood as co-produced.

While the social model shifted attention to structural barriers, its focus on removing obstacles can imply a one-directional process of accommodation. Recent work extends this by framing access as something that is negotiated. For example, Bennett et al. [28] describe this through an interdependence perspective, emphasising that access emerges through collaboration among people, technologies, and their environments. Such an approach contrasts with individualised fixes and foregrounds collective adjustments that make participation possible. Designing for this reality cannot rely on assumptions of uniformity, with Mankoff et al. [183] arguing that these assumptions obscure the diversity of people's needs. Spiel et al. [280] extend this point by showing that inclusive design requires deliberate engagement with this diversity from the outset, treating difference as a design parameter.

2.3.1 Participatory Design

Participatory Design (PD) offers a methodological response to these concerns by shifting from designing *for* users toward designing *with* them. Rather than treating end-users as passive informants, PD positions them as active contributors throughout ideation, prototyping, and evaluation. Sanders and Stappers [260] describes PD as a set of practices that redistribute decision-making power, enabling people with lived experience to shape technologies that affect their lives. This view is not only methodological but political, as PD emerged from a commitment to empowerment and equity, rooted in Scandinavian workplace democracy movements of the 1970s [245]. Ehn [77] traces these origins to efforts that gave workers influence over technologies shaping their jobs, framing PD as a rights-based approach and not as a neutral design method.

While Human-Centered Design and User-Centered Design have historically prioritised usability and efficiency [208], PD foregrounds collaboration and situated expertise. Over time, PD has expanded beyond industrial contexts into public services, education, and health, and more recently into HCI research, where it is used to surface diverse needs and co-create solutions [137, 206]. Prior work reports PD adaptations with deaf-blind travellers, autistic children, and people with dementia, using tangible and empathetic methods to reduce reliance on language [20, 93, 170]. Co-design, often used interchangeably with PD, describes iterative cycles of ideation and prototyping that involve people with lived experience alongside designers and technical experts. Participatory design workshops provide a structured setting for collaborative exploration and decision making, allowing participants and designers to share perspectives, negotiate priorities, and iteratively shape design directions.

Within these workshops, prototyping plays a central role in making abstract ideas concrete. Prototyping can vary significantly, from pen and paper mock-ups, to storyboards, or interactive simulations. These prototypes help participants envision and critique design concepts early in the process [260], creating a shared reference point which supports collaborative decision-making. For example, Page and Heiss [223] describe how tactile tools and low-fidelity prototypes can make abstract concepts concrete, supporting joint sense-making and iterative refinement. Such artefacts are defined by how they support joint exploration and decision-making, enabling participants to influence outcomes, operating as generative tools that redistribute agency in design and create conditions for meaningful participation.

Cultural probes complement these practices by eliciting situated insights that inform early design stages. Gaver et al. [99] introduced cultural probes as collections of evocative materials (e.g. maps, diaries, disposable cameras) intended to prompt reflection. Subsequent work has adapted probes for diverse contexts, using them to surface everyday practices that might otherwise remain unstated. For instance, Hendriks et al. [119] employed cultural probes in design research with older adults, capturing daily routines and preferences in domestic settings. This approach revealed tacit practices and emotional dimensions of everyday life, informing design directions that would not have emerged through conventional methods.

Despite these principles of inclusion, PD remains under-represented in accessibility research, with Mack et al. [181] reporting that only a small fraction of accessibility papers involving user studies employed PD, with many limiting participation to single-session engagements that dilute its intent. Bannon et al. [21] note that such practices risk reducing PD to tokenistic involvement rather than genuine co-creation, arguing for adapted PD approaches that sustain multi-session engagement and address communication barriers. Inclusive PD approaches have been shown to improve relevance for communities historically excluded from mainstream design, where designers often lack lived experience of the conditions they aim to address [92, 122]. PD thus functions as both a methodological and ethical framework, affirming the right of individuals to shape technologies that affect their lives while producing outcomes that reflect situated expertise.

However, conventional PD assumes fluent communication and cognitive flexibility, which can exclude many disabled communities. PD methods have been adapted to ensure valid participation of such communities, requiring deliberate adjustments so that people with disabilities can contribute meaningfully. Dalemans et al. [70] describe strategies for adapting research processes to include participants with severe language impairments, such as simplifying materials, using pictorial supports, and structuring consent procedures to maintain autonomy. Similarly, Kagan [143] emphasises supported conversation techniques that enable individuals with communication difficulties to express preferences and make decisions during collaborative activities. These adaptations reflect an ethical commitment to inclusion, where participation involves being heard and understood. Ethical inclusion also involves respecting participants' autonomy, ensuring informed consent through adapted materials, and creating conditions where participants can shape outcomes [70, 143].

Inclusive PD also requires attention to method bias. Wilson et al. [320] argue that standard co-design techniques often presume verbal fluency, which can marginalise participants with aphasia or other communication disorders. They propose tangible design languages that allow ideas to be expressed through physical artefacts, reducing reliance on speech. Similarly, Galliers et al. [95] note that multimodal supports and simplified interaction pacing are essential for sustaining engagement. Bircanin et al. [31] demonstrate similar principles in work with adults with severe intellectual disabilities, using graded assistance and multimodal scaffolds to support agency in shaping design outcomes. These practices reveal that inclusive PD is not a single technique but a set of adjustments that respond to diverse communicative and cognitive profiles.

Different disabled communities require distinct considerations for involvement in PD, such as studies with autistic children being adapted to support attention. Hughes et al. [133] report that visual and auditory cueing does not uniformly improve tracking performance and can increase cognitive demands, suggesting implications for reducing clutter in audiovisual interfaces. Similarly, Hijab et al. [121] introduced playful activities to explore interactions with collaborative play tools in participatory sessions with autistic children. Their findings informed recommendations for toolkits that support co-design, illustrating how empirical insights can guide inclusive methods.

Some proponents of inclusive participatory approaches have pointed out additional limitations in current PD approaches. For instance, Spiel et al. [281] argue that participation must also be extended past ideation and prototyping to evaluation. They suggest working with participants to determine how the resulting designs should be evaluated, defining success criteria rather than leaving the assessment solely to researchers. Similarly, standard ethical guidelines are often insufficient, necessitating reliance on multiple moral frames of reference and situated ethical judgments (i.e. micro-ethics) when following participatory approaches [279].

2.3.2 Participatory Design in Aphasia

Building on the discussion of inclusive design, we turn to aphasia, an exemplar community on whom we focus in this thesis for whom standard accessibility features are insufficient because of their heavy reliance on language. Aphasia is an acquired language impairment that can affect a person's speaking, understanding, reading, or writing, most often following stroke [80]. Aphasia often presents as pan-modal language impairment, with expressive and receptive difficulties across speech, reading, and writing [22]. While these changes affect communication, they do not diminish intelligence, reasoning or recall. Varley et al. [304] show that individuals with aphasia that had severe challenges with grammar could still perform complex mathematical operations, including recursive calculations, indicating that core cognitive functions remain intact. This distinction matters because it frames aphasia not as a global cognitive deficit, but as a barrier to language-based interaction.

These barriers extend beyond everyday conversation into research and design activities, where methods often presume fluent speech and reading. Supported strategies are therefore essential to enable participation. Kagan [143] introduced supported conversation techniques that help people with aphasia express preferences and make decisions, using verification and multimodal cues to maintain agency. Dalemans et al. [70] describe adaptations for research materials, such as one question per page, pictorial supports, and removal of negatives, which reduce processing demands and support informed consent. These principles have informed aphasia-friendly guidelines for documents [120], which recommend clear layout and visual signposting to aid navigation.

Speech and language therapy literature offers further practices that can be applied when preparing materials or structuring interaction. For instance, Knollman-Porter et al. [155] explored strategies such as using plain language and breaking text into smaller chunks, finding that these can help maintain engagement with written materials. Galliers et al. [95] also find that multimodal scaffolds, such as visual and tactile aids, were necessary to ensure continued engagement. Technology-mediated supports also play a role, with Knollman-Porter et al. [154] reporting that text-to-speech with customised settings for voice and rate can reduce processing time without compromising comprehension, making longer texts more manageable.

Participatory design with people with aphasia, therefore, requires moving beyond purely verbal methods. Wilson et al. [320] propose tangible design languages that allow ideas to be expressed through physical artefacts, reducing reliance on speech. Galliers et al. [95] show how picture cards and simplified tasks help maintain involvement during co-design sessions. Roper et al. [253] adapt usability testing by using short, direct tasks and visual prompts, enabling participants to contribute meaningfully despite language barriers. Recent work also explores think-aloud protocols tailored for aphasia, using structured prompts and breaks to capture reasoning without imposing continuous narration [66].

Additionally, comprehension should not be assumed during these activities. Kruger et al. [158] demonstrate how pace interacts with opportunities for re-reading, suggesting that evaluation of materials must consider whether participants can recover when attention shifts. Interactive techniques such as pausing and highlighting have been shown to improve understanding in time-sensitive contexts [42]. These findings point to practical checks during co-design, that if participatory methods involve timed elements, controls for re-timing or local review should be available so that comprehension can be assessed collaboratively. Methods that invite participants to reason beyond their own immediate situation can reduce self-presentation pressure. For example, Neate et al. [200] explore creating co-created personas with aphasia, informing exercises where participants with aphasia can articulate preferences indirectly.

2.4 Ethical and Methodological Considerations

This work considers ethical constraints that arise when comprehension-first personalisation adapts audiovisual media during playback. Personalisation can rely on sensitive viewer data, including preferences or inferred profiles, which creates a risk of unintended disclosure through storage, transmission, or model behaviour. Ethical practice therefore prioritises on-device processing, explicit boundaries on what data are collected, and a transparent description of what is stored or shared [197]. Explicit consent is required for adaptive features, especially where adaptations materially change the delivered content, since these changes affect what the viewer experiences as the programme itself. In practice, this implies an opt-in approach for adaptation, together with a clear opt-out that can be used during viewing. The opt-out should include a direct way to return to the unadapted version.

Editorial integrity is a separate constraint because playback adaptation can change how narrative meaning is conveyed. Adaptations should respect creative intent by constraining what can change, when it can change, and how far a change can go. Narrative-importance metadata provides one way to express these constraints, allowing adaptation rules to preserve key elements while still enabling adjustments that reduce comprehension effort [317]. Ethical framing for this work treats adaptation limits as part of the system design, since the boundary between access support and distortion depends on the permitted transformation set.

Social context affects whether adaptations are used, since viewing often takes place with others. Visible adaptations can stigmatise by marking the viewer as different or by framing accessibility controls as special-case tools. Interface choices can reduce this risk by using neutral language, placing controls alongside mainstream playback controls, and avoiding presentation patterns that single out accessibility features [69]. Shared viewing also introduces negotiation cost when adaptations affect what co-viewers see or hear. Ethical design aims to minimise disruption and support rapid reversal during routine viewing [69].

Methodologically, the studies treat participatory design as a power-sharing commitment, where participants influence design decisions [260] and RtD provides the overarching structure [330]. This commitment shapes the choice of activities and the interpretation of results, since comprehension-first personalisation involves trade-offs between understanding support, interaction effort, social acceptability, and perceived legitimacy of content change. Research bias can arise when design and evaluation rely only on verbal or written communication, which can exclude people with complex communication needs from contributing to decisions about what is built and how it is judged [46].

2.5 Conclusion

This chapter reviewed established approaches to audiovisual accessibility and located their limits for people with complex communication needs. It traced how subtitles and audio description developed as dominant interventions, then described how platform shifts and emerging viewing practices complicate delivery and consistency. It introduced personalisation as an accessibility strategy, framing playback adaptation as changes to timing, structure, layout, and mix that aim to reduce effort while preserving viewer agency. It then discussed inclusive participation through participatory design, with attention to adaptations needed for aphasia and other complex communication needs. It also set out ethical and methodological considerations for comprehension-first personalisation, focusing on privacy, consent, creative intent, and social acceptability in shared viewing. Taken together, these strands motivate a shift from feature presence toward whether viewers can follow what unfolds over time with manageable effort. The next chapter, Chapter 3, establishes a field-level baseline by reviewing technical interventions for digital audiovisual media accessibility and by characterising which communities and challenges current research addresses. This baseline then supports the empirical chapters that follow, which examine where comprehension breaks for people with aphasia during everyday viewing and how playback adaptation can be explored in practice.

Chapter 3

Systematic Review

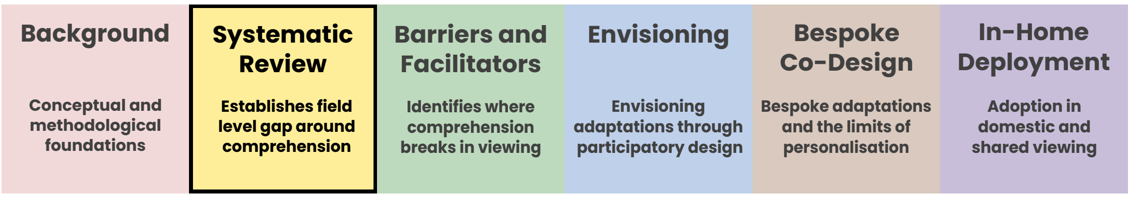

Figure 3.1: Map of thesis - systematic review

In this chapter we address the claim that accessibility research in audiovisual media has largely focused on sensory barriers, leaving comprehension barriers largely overlooked. We do this through a systematic literature review of research on technical accessibility interventions for digital audiovisual media. We identify a significant gap in prior research, which we aim to bridge in the following chapters – see Figure 3.1.

3.1 Introduction

Audiovisual media is present in daily life through television broadcasts, films, and streaming services. Consumption of audiovisual media enables social engagement through shared viewing, civic engagement through access to news and current affairs, and cultural engagement through film and television series that convey human values and traditions. Access to media is integral to participation in modern society, “*vital to informed political knowledge*” [234] and necessary for maintaining awareness of events beyond local surroundings.

As discussed in Chapter 2, audiovisual media are often inaccessible for people with disabilities. Indeed, these media have inherent complexities in that they are multimodal, containing visual, audio and narrative information, as well as being time-sensitive. The context in which audiovisual media is consumed can also introduce accessibility barriers, such as small screens on smartphones [290], or the additional cognitive demands of virtual, augmented or mixed reality (VR/AR/MR) environments [190, 196, 224].

To address these accessibility barriers, researchers have explored different technical solutions, from subtitles to audio descriptions, standardising practices into guidelines (see Section 2.1). Evaluation is often centred on feature provision and perceptual accuracy rather than comprehension during viewing. To understand what these solutions are and who they are designed for, we conducted a systematic literature review (SLR) of accessibility research in digital audiovisual media, focusing on the technical interventions developed. We identify 181 papers spanning 27 years that fit our criteria. This SLR is modelled on two prior reviews of accessibility research done by Mack et al. [181] and Curtis et al. [68], as well as building on a review of accessibility research at the ‘Interactive Media Experiences’ (IMX) conference by Vatavu [305]. Vatavu audits accessibility papers in IMX proceedings to quantify representation and participant involvement, but does not run a cross-venue review of audiovisual media accessibility interventions more broadly. This SLR focuses on what *disabled communities* are represented in research, the *accessibility challenges* experienced by disabled viewers, and the proposed *accessibility interventions* across literature. Through this focus, we aim to answer the following three research questions:

RQ1: What are the main accessibility challenges faced by people with disabilities when accessing digital audiovisual media?

RQ2: Who does accessibility research in audiovisual media focus on?

RQ3: What interventions are used to help support different accessibility needs?

3.2 Method

We reviewed scientific literature using the PRISMA method [240], applying additional guidelines from Silva and Francila Weidt Neiva [275] and Siddaway et al. [274]. First, we outlined the scope of the SLR by specifying the requirements for the papers. We then outlined the steps for creating the dataset. Lastly, we described the qualitative and quantitative analyses performed.

3.2.1 Scope

Research on individual accessibility interventions for digital audiovisual media is extensive, but there has been no comprehensive review that provides an overall view of the field and identifies gaps in knowledge about accessibility challenges and interventions. Targeted reviews have examined specific disabled communities, such as Brule et al. [43] or Bhowmick and Hazarika [30] exploring assistive technologies for people who are BLV, or work by Alsalamah [15] on assistive technologies used by DHH students in higher education. Other work has focused on particular technical interventions, including collaborative technologies for children with special needs [23], or studies of interactions with audiovisual media technologies, such as eye tracking research among the DHH community [12]. Our work, however, examines literature on accessibility interventions across a range of communities and technologies. This differs from previous reviews by concentrating on a subset of accessibility technologies and providing detailed analysis of this area. We have therefore limited the scope to research that focuses specifically on accessibility interventions that aim to improve the access of digital audiovisual media for people with disabilities. Below we have defined three requirements for papers in our SLR:

1. **Digital audiovisual media** – papers must explore issues in digital audiovisual media. This excludes, for example, live performances (e.g. [299]) and media that does not have both audio and visual components, such as radio broadcasts (e.g. [319]).

2. **Technical accessibility interventions** – papers must address a technical intervention that aims to make digital audiovisual media more accessible. This excludes, for example, papers on social support (e.g. [51]).

3. **Audiovisual media content** – papers must address the accessibility of a piece of audiovisual content. This excludes the platform on which the content is consumed (e.g. [283]) or any interactions with said platform (e.g. [221]).

3.2.2 Dataset Creation

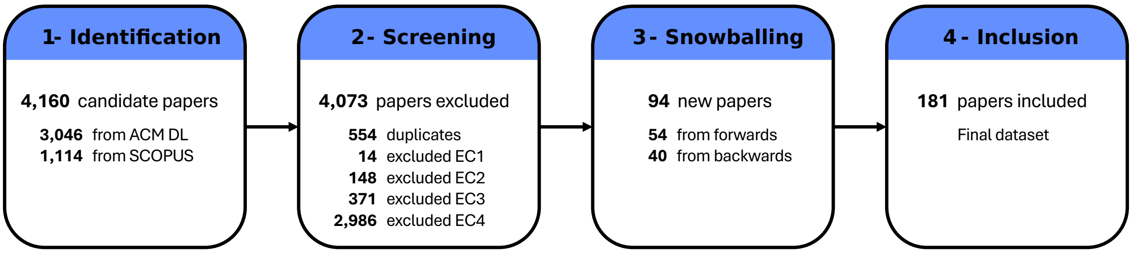

Figure 3.2: Flowchart summarising the creation of our dataset.

Following the requirements outlined above, we followed a three-stage process to create our dataset – see Figure 3.2 for details. We start with the *identification* of potentially relevant papers, followed by a *screening* stage in which irrelevant papers were excluded. Finally, we followed guidelines from Wohlin [322] and conducted a reference and citation *snowballing* stage.

Identification

To identify literature on HCI and digital audiovisual media accessibility, we conducted the search using three major electronic databases commonly used for Computer Science and HCI research: the ACM Digital Library, SCOPUS, and IEEE Xplore. These databases reference most relevant publication venues for researching intersecting computer science and accessibility, including the ACM Conference on Human Factors in Computing Systems (CHI) and ACM Conference on Computers and Accessibility (ASSETS) [181].

We set out to create a search string for these databases, searching across paper titles, abstracts, and keywords to reduce the number of false positives from full-text search. To identify the specific search string used for the query, we followed an iterative approach, resulting in a decision to use generic truncated keywords, such as ‘*access**’ and ‘*impair**’. Using specific terms, such as ‘*deaf*’, could have biased the search due to our own terminology and keyword inclusion choices. We also included relevant keywords for audiovisual media, such as “*video*”, “*television*”, and “*audiovisual*”. We added exclusion keywords as initial searches returned a high number of papers on topics that were not relevant to us, such as video games. We consider video games to be a distinct category of entertainment applications compared to the digital audiovisual media examined in this thesis, due to differences in consumption patterns, user engagement, and content interactivity. The final search query for the ACM Digital Library resulted in 3, 046 papers and can be seen below:

Title:((video OR tv OR television OR broadcast OR audiovisual OR "audio visual" OR "audio-visual") AND (access* OR disab* OR impair*) AND NOT (game OR games OR videogame OR videogames OR "video game" OR "video games")) OR Abstract:((video OR tv OR television OR broadcast* OR audiovisual OR "audio visual" OR "audiovisual") AND (access* OR disab* OR impair*) AND NOT (game OR games OR videogame OR videogames OR "video game" OR "video games")) OR Keyword:((video OR tv OR television OR broadcast* OR audiovisual OR "audio visual" OR "audiovisual") AND (access* OR disab* OR impair*) AND NOT (game OR games OR videogame OR videogames OR "video game" OR "video games"))*

The use of common keywords (e.g. ‘*video*’) and truncated keywords (e.g. ‘*access**’) resulted in this same search query returning 75, 000 results on SCOPUS and 17, 000 results on IEEE Xplore, with most papers being fully irrelevant to this review. To address this high number of results, we manually checked all venues with at least 25 candidate papers. The manual check involved using the built-in relevancy feature, looking at the 25 most relevant papers for each of these venues and excluding all venues if 3 or fewer of the 25 papers were deemed relevant (10%). Conducting this manual check on the results of IEEE Xplore resulted in no relevant venues. Therefore, the database was not used from then on. The same manual check brought the total number of candidate papers from SCOPUS down to 1, 114 papers. The adjusted search string for SCOPUS, limiting results to relevant venues, can be seen below:

```
TITLE-ABS-KEY((video OR tv OR television OR broadcast*
OR audiovisual OR "audio visual" OR "audio-visual") AND
(access* OR disab* OR impair*) AND NOT (game OR games OR
videogame OR videogames OR "video game" OR "video games"))
AND (LIMIT-TO(EXACTSRCTITLE, "ACM International Conference
Proceeding Series") OR LIMIT-TO(EXACTSRCTITLE, "Communications
In Computer And Information Science") OR LIMIT-TO(EXACTSRCTITLE,
"Conference On Human Factors In Computing Systems Proceedings")
OR LIMIT-TO(EXACTSRCTITLE, "Procedia Computer Science")
OR LIMIT-TO (EXACTSRCTITLE, "Conference on Computers and
Accessibility") OR LIMIT-TO(EXACTSRCTITLE, "Interactive
Experiences for TV and Online Video") OR LIMIT-TO (EXACTSRCTITLE,
"Interactive Media Experiences") OR LIMIT-TO(EXACTSRCTITLE,
"European Conference on Interactive TV and Video") OR
LIMIT-TO(EXACTSRCTITLE, "IEEE Transactions On Multimedia")
OR LIMIT-TO(EXACTSRCTITLE, "Conference On Computer-Supported
Cooperative Work And Social Computing"))
```

The final search was conducted on 2 September 2022, resulting in a total of 4, 160 candidate papers.

Screening

Following the PRISMA 2021 guidelines [240], we created eligibility criteria for the screening stage based on our scope and requirements, resulting in four eligibility criteria (EC1 – EC4):

EC1: **Availability of text** – the full text of the paper must be available in English.

EC2: **Peer-reviewed research** – the paper must be a piece of peer-reviewed research (e.g. conference papers, journal articles, posters, etc.) and excludes other research (e.g. preprints, theses and dissertations, book chapters, datasets, etc.).

EC3: **Digital audiovisual media** – the paper must focus on digital audiovisual media.

EC4: **Accessibility intervention** – the paper has to explore a technical intervention that aims to improve the accessibility of the digital audiovisual media.

The screening process began by removing 554 duplicate papers, which represented 13.3% of the initial set of 4, 160 papers. After duplicates were removed, the remaining papers were assessed against the eligibility criteria EC1–EC4 (see Table 3.1). The first stage involved checking titles and abstracts to exclude papers that were clearly irrelevant. When there was uncertainty, papers were retained for further review to avoid discarding potentially relevant work. Next, the full text of each paper was examined to confirm eligibility. To manage this process, we created a spreadsheet where each paper was marked as relevant or irrelevant. For papers deemed irrelevant, the specific eligibility criterion that led to exclusion was recorded.

Eligibility Criteria	**Number of papers**
Duplicate papers	554 (13.3%)
EC1 - *Availability of text*	14 (0.3%)
EC2 - *Peer-reviewed research*	148 (3.6%)
EC3 - *Digital audiovisual media*	371 (8.9%)
EC4 - *Accessibility intervention*	2,986 (71.8%)
Total relevant papers	87 (2.1%)

Table 3.1: The number of candidate papers excluded because of duplicates and by the Eligibility Criteria, as well as the number of resulting relevant papers.

A large proportion of exclusions occurred under EC4 (71.8%) because the papers were clearly unrelated to accessibility, many of which focused on technical topics such as video processing (e.g. [238]) or video transcoding (e.g. [167]). Other papers addressed audiovisual media and accessibility, but did not study interventions to improve accessibility. Instead, they used media content in medical contexts (e.g. [34]) or explored aspects of disability without focusing on accessibility interventions (e.g. [246]). Although some of these papers could have been excluded under EC3 or EC4, the prevalence of technical papers unrelated to accessibility made EC4 an effective criterion for initial filtering. After completing all screening and eligibility checks, $N = 87$ papers were identified as relevant.

Snowballing

After completing the screening and eligibility review, we applied a snowballing procedure as described by Wohlin [322]. This process identified $N = 94$ additional papers. Snowballing was carried out in four iterations of forward and backward searches. The iterations continued until no new candidate papers were found, following the same screening and eligibility checks described in the *Screening* stage.

Backward snowballing began by examining paper titles, author names, and publication venues. Forward snowballing was performed using Google Scholar. When the information available on Google Scholar was insufficient to make a decision, the full text of the citing paper was reviewed.

From this process, 54 papers were identified through forward snowballing and 40 through backward snowballing. This brought the total number of papers in the dataset to 181. Of these, 87 papers (48.1%) were identified during the initial identification and screening steps, and 94 papers (51.9%) were obtained through snowballing. This proportion aligns with previously reported expectations for snowballing [107].

3.2.3 Data Analysis

We analysed the dataset of $N = 181$ papers using both qualitative and quantitative approaches. The qualitative analysis began with the development of a codebook by three researchers, following an iterative process in which the researchers discussed codes over several rounds, each time coding a sample of ten papers to refine the categories. The final codebook, shown in Table 3.2, contains 13 categories with 2–11 sub-codes each. For most categories, multiple codes could apply to a single paper, except for binary categories. This means that percentages reported in Section 3.3 do not always sum to 100%.

Category	**Codes**	**Multiple codes**	**Pairwise agreement**	**IRR**
Type of content	TV broadcast; on-demand video; web video; live video; other	Yes	72.2%	0.477 (Moderate)
Viewing device	Television; desktop; smartphone; tablet; big screen; other	Yes	88.9%	0.911 (Almost perfect)
Viewing context	General viewing; education; commercial; other	Yes	94.4%	0.660 (Substantial)
Type of challenge	Viewing video; hearing audio; reading subtitles; understanding speech; following narrative; image clarity; on-screen clutter; other	Yes	72.2%	0.493 (Moderate)
Community of focus	BLV; DHH; motor/physical impairment; autism; IDD; other cognitive impairment; older adults; general disability; other	Yes	83.3%	0.851 (Almost perfect)
Participant groups	No user study; people with disabilities; people without disabilities; older adults; specialists; caregivers	Yes	88.9%	0.781 (Substantial)
Use of proxies	Yes; no	No	94.4%	0.444 (Moderate)
Ability based comparison	Yes; no	No	94.4%	0.444 (Moderate)
User study method	Controlled experiment; survey; usability testing; interviews; case study; focus group; field study; workshop/design; other; none	Yes	66.7%	0.774 (Substantial)
User study location	No user study; near/at researcher's lab; participant's home, residence, or school; neutral location; online/remote; other	Yes	72.2%	0.775 (Substantial)
Participatory design	Yes; no	No	88.9%	0.654 (Substantial)
Contribution type	Empirical; artifact; methodological; theoretical; dataset; survey; opinion	Yes	77.8%	0.910 (Almost perfect)
Type of intervention	Subtitles; audio description; tangible device; sign language; audio or video manipulation; personalisation; customisation; in-person assistance; second screen; other	Yes	77.8%	0.837 (Almost perfect)

Table 3.2: Codebook with the calculated pairwise agreement, IRR scores and interpretation for each category.

Of the 13 categories, five were created by the researchers to capture trends in audiovisual media accessibility interventions, including the context of media consumption (device, environment, and content type), user challenges, and the interventions studied. Eight categories were adapted from Mack et al. [181] to enable comparison with broader accessibility research and to examine user study methods and target communities. Once the codebook was finalised, two researchers coded all papers, with the second researcher coding a random sample of 10% ($N = 18$) to calculate inter-rater reliability (IRR) using Fleiss' Kappa [91].

Quantitative analysis focused on paper and participant counts across the 27-year period, as well as keyword frequencies. We computed descriptive statistics for participant counts, including mean, median, interquartile range, and standard deviation, examining these values by community focus, participant groups, and the use of proxies or ability-based comparisons. For keyword analysis, we extracted 660 unique keywords, of which 209 (31.7%, 209/660) appeared at least twice. A researcher manually grouped similar keywords into higher-order categories, such as combining "*blind*", "*blindness*", and "*blind people*" under "*BLV*". Sixteen such groupings were created, including three that addressed British and American spelling differences (e.g. "*personalisation*" and "*personalization*"). Of the 181 papers, 13 (7.2%) did not include any keywords.

The results of our qualitative coding were compared with findings from Mack et al. [181], which analysed $N = 506$ papers on HCI accessibility with a broader scope. Other reviews (e.g. [23, 43, 305]) were considered less suitable for comparison due to their narrower focus on specific communities, technologies, or venues.

3.3 Results

We now present the results of the analysis, starting with the accessibility challenges addressed, followed by the communities and participants involved, the research methods used, and finally the accessibility interventions investigated. We then look at trends in the 27-year period and report results of the qualitative analysis.

3.3.1 Accessibility challenges addressed

To examine the accessibility challenges faced by people with disabilities when engaging with digital audiovisual media, we analysed the context emphasised in the research. First, we look at the nature of the content, distinguishing between different formats and delivery models. Second, we examine the devices through which this content is watched, as device characteristics influence both interaction and accessibility requirements. Third, we consider the viewing context, which reflects the purpose and setting in which the media is consumed. Finally, we identify the specific accessibility challenges addressed by researchers.

Type of content

Type of content	Papers	This code only
TV broadcast	58 (32.0%)	50 (27.6%)
VOD	56 (30.9%)	43 (23.8%)
Web video	45 (24.9%)	36 (19.9%)
Live video	1 (0.6%)	1 (0.6%)
Other	39 (21.5%)	37 (20.4%)

Table 3.3: The frequency of applied codes for *type of content*.

As shown in Table 3.3, research on audiovisual media accessibility spans several content types. A substantial proportion of papers focus on television broadcasts (32.0%, $N = 58$), including work on subtitle positioning and user experience in broadcast environments (e.g. [40, 64]). Video-on-demand services account for 30.9% of papers ($N = 56$), with studies examining responsive subtitle design for web-based platforms (e.g. [132]) and accessibility evaluations of streaming services (e.g. [147]). Web video represents 24.9% of papers ($N = 45$), where research often addresses subtitling practices and challenges on platforms such as YouTube and TikTok (e.g. [166, 187]). Only one paper [162] investigated live video (0.6%), focusing on enabling DHH students to pause and highlight subtitles during real-time educational streams.

The category labelled as 'other' comprises 21.5% of papers ($N = 39$). Many of these do not specify the content type, while others explore emerging formats. Examples include studies on subtitle readability in VR (e.g. [218]), dynamic subtitle placement in cinematic VR (e.g. [256]), and accessibility considerations for subtitles in 360-degree video experiences (e.g. [41]). This group also includes interactive audiovisual experiences designed for sensory substitution, such as systems that allow users to perceive colours through sound or hear visual elements (e.g. [53]). A smaller subset of papers (7.7%, $N = 14$) was coded under multiple content types. These typically examined accessibility features across web-based VOD platforms (42.9%, 6/14) or addressed television features applicable to both broadcast and on-demand contexts (35.7%, 5/14).

Viewing device

Viewing device	Papers	This code only
Television	83 (45.9%)	66 (36.5%)
Computer	59 (32.6%)	48 (26.5%)
Smartphone	18 (9.9%)	1 (0.6%)
Tablet	11 (6.1%)	0 (0.0%)
Big screen	7 (3.9%)	1 (0.6%)
Other	48 (26.5%)	38 (21.0%)

Table 3.4: The frequency of applied codes for *viewing device*.

When considering the devices used to access audiovisual media, nearly half of the papers examined television (45.9%, $N = 83$) – see Table 3.4. These include studies on adaptive subtitle positioning for broadcast environments (e.g. [64]) and dynamic subtitle systems designed to improve user experience (e.g. [40]). Desktop and laptop computers were the second most common category (32.6%, $N = 59$), often appearing in work on web-based accessibility tools and subtitling systems, such as research on subtitle visualisation for Arabic script across multiple platforms (e.g. [263]).

Mobile devices were addressed less frequently. Smartphones appeared in 9.9% of papers ($N = 18$), and tablets in 6.1% ($N = 11$). These devices were often used together, particularly in second-screen scenarios where a phone or tablet complements the main display. For example, some systems allow subtitles to be displayed on a smartphone during cinema viewing (e.g. [61, 97]). The only paper focused exclusively on tablets explored tactile interaction using small robots that move across the screen to convey information [112].

A smaller group of papers investigated immersive or alternative devices. Examples include research on subtitle readability in virtual reality environments [218] and dynamic subtitle placement in cinematic VR experiences [256]. Other work examined haptic feedback systems for conveying emotional or visual cues, such as tactile robots or wearable devices [112].

Overall, 27 papers (14.9%) addressed more than one device type. The most frequent combination was smartphone and tablet (37.0%, 10/27), representing over half of all smartphone-related papers and nearly all tablet-related papers. Smartphones and tablets were also commonly paired with television, either for synchronised content or accessibility enhancements. Desktop viewing appeared alongside smartphones (29.6%, 8/27), tablets (25.9%, 7/27), and television (18.5%, 5/27), typically in studies of cross-platform systems.

The 'other' category (26.5%, $N = 48$) included papers where the device was unspecified or where the focus was on algorithmic or software development rather than hardware, such as work on emotion-based video annotation (e.g. [259]). It also covered hybrid approaches, such as augmented reality sign language interpreters for TV viewing using head-mounted displays (e.g. [308]), and systems that communicate emotional states through environmental lighting and mobile applications (e.g. [11]).

Viewing context

The distribution of viewing contexts in the dataset is shown in Table 3.5. Most papers addressed general viewing scenarios (81.8%, $N = 148$), reflecting a strong emphasis on making audiovisual media accessible in everyday settings such as home environments. Examples include work on adaptive subtitle systems and personalised accessibility features for television and streaming platforms (e.g. [11, 98, 189]).

Educational contexts accounted for 7.7% of papers ($N = 14$). These studies often

Viewing context	**Papers**	**This code only**
General viewing	148 (81.8%)	137 (75.7%)
Educational	14 (7.7%)	12 (6.6%)
Commercial setting	7 (3.9%)	2 (1.1%)
Other	23 (12.7%)	19 (10.5%)

Table 3.5: The frequency of applied codes for *viewing context*.

explored methods for improving or automating captioning and audio description to support learning. For instance, research has examined real-time captioning tools for classroom use (e.g. [162]) and automated approaches for generating descriptions in instructional videos (e.g. [142]). One paper [277] investigated accessibility in immersive video environments designed to help people with intellectual and developmental disabilities (IDD) acquire new skills. Commercial contexts were less common, representing 3.9% ($N = 7$). Most of these papers focused on cinema accessibility through second-screen solutions, such as systems that deliver personal subtitles or audio description via smartphones (e.g. [61, 311]).

The remaining papers were grouped under 'other' (12.7%, $N = 23$). This category includes work that did not specify a particular context or addressed unique scenarios. Examples include studies on captioning practices and challenges faced by DHH content creators on platforms like YouTube (e.g. [166]), as well as research on making security surveillance video accessible to BLV users (e.g. [48]).

Accessibility challenges addressed

As shown in Table 3.6, the most frequently addressed challenges were viewing video (43.1%, $N = 78$) and reading subtitles (42.5%, $N = 77$). Many papers in the first category implemented AD to make visual content accessible for BLV viewers. For instance, Wang et al. [314] developed a system that automatically generates AD by analysing audiovisual content and predicting insertion points. Research on subtitle accessibility included work on improving presentation and usability. For instance, Kurzhals et al. [160] examined adaptive subtitle positioning to reduce visual distraction during viewing.

Challenges addressed	**Papers**	**This code only**
Viewing video	78 (43.1%)	37 (20.4%)
Hearing audio	22 (12.2%)	3 (1.7%)
Reading subtitles	77 (42.5%)	49 (27.1%)
Understanding speech	17 (9.4%)	6 (3.3%)
Following narrative	36 (19.9%)	3 (1.7%)
Issues with image	10 (5.5%)	2 (1.1%)
Screen clutter	0 (0.0%)	0 (0.0%)
Other	37 (20.4%)	8 (4.4%)

Table 3.6: The frequency of applied codes for *challenges addressed*.

Other challenges coded in the dataset were following narrative (19.9%, $N = 36$), hearing audio (12.2%, $N = 22$), and understanding speech (9.4%, $N = 17$). The narrative related papers often combined subtitles and AD to support comprehension of plot and dialogue. A code for screen clutter was included based on broader accessibility literature, such as dementia-friendly television design that identified excessive on-screen elements as a barrier (e.g. [94]). However, none of the papers in this dataset proposed interventions for this issue.

The 'other' category accounted for 20.4% of papers ($N = 37$) and included work on emotional accessibility and social viewing contexts. For example, McGill et al. [189] explored the use of acoustically transparent headsets to allow BLV viewers to access AD without isolating them from group conversations.

Multiple codes were common, with 73 papers (40.3%) assigned to more than one challenge. The most frequent pairing was viewing video and other (26.0%, 19/73). Papers addressing both viewing video and reading subtitles (12.3%, 9/73) typically implemented dual interventions, such as systems providing synchronised AD and subtitles for cinema audiences (e.g. [97]).

3.3.2 Disabled communities involved

The previous section examined the accessibility challenges addressed in the literature. Here, we turn to the people and communities represented in this research and how participation is structured.

This part of the analysis considers three dimensions. First, we look at the communities that the research targets, identifying which disability groups are the focus of accessibility interventions. Second, we examine who participates in user studies, including whether participants belong to the communities being studied or come from other groups. Third, we review the use of proxies and ability-based comparisons.

To provide context beyond audiovisual media accessibility, we compare the findings from our systematic review with those reported by Mack et al. [181]. This comparison helps situate the scope of participation and representation in our dataset relative to the wider field.

Communities of focus

Community of Focus	Papers	Mack et al.	This code only	Mack et al.
BLV	83 (45.9%)	43.5%	70 (38.7%)	41.1%
DHH	97 (53.6%)	11.3%	77 (42.5%)	8.5%
Motor impairment	3 (1.7%)	14.2%	0 (0.0%)	11.7%
Autism	0 (0%)	6.1%	0 (0.0%)	4.2%
IDD	1 (0.6%)	2.8%	0 (0.0%)	1.6%
Other cognitive	6 (3.3%)	9.1%	0 (0.0%)	5.7%
Older adults	8 (4.4%)	8.9%	2 (1.1%)	5.7%
General disability	16 (8.8%)	9.1%	8 (4.4%)	6.1%
Other	5 (2.8%)	9.1%	0 (0.0%)	4.0%

Table 3.7: The frequency of applied codes for *community of focus*, along with data reported by Mack et al. [181].

As shown in Table 3.7, most papers in the dataset focused on the DHH community (53.6%, $N = 97$) and the BLV community (45.9%, $N = 83$). Together, these two groups accounted for 93.9% ($N = 170$) of all papers. Other communities appeared far less frequently. Papers coded as addressing general disability represented 8.8% ($N = 16$) and included such work as improving accessibility in video-based learning environments [264] or on reducing barriers in web video players [306]. Research on immersive media personalisation for broader audiences was also included under this label (e.g. [193]). A smaller proportion of papers focused on older adults (4.4%, $N = 8$), cognitive impairments (3.3%, $N = 6$), motor impairments (1.7%, $N = 3$), and IDD (0.6%, $N = 1$). No paper in our dataset addressed autism, although Mack et al. [181] reported 6.1% of papers in their broader accessibility review as focusing on this community.

The 'other' label (2.8%, $N = 5$) was applied to papers that targeted communities not covered by predefined codes. Examples include developing a device that reads subtitles aloud for BLV users and people with dyslexia [207], or a similar system for individuals with reading difficulties [177].

A total of 33 papers (18.2%, $N = 33$) considered more than one community. Most papers focusing on BLV (84.3%, 70/83) and DHH (79.4%, 77/97) addressed these communities exclusively. In contrast, half of the papers coded as general disability (50.0%, 8/16) and two papers on older adults (25.0%, 2/8) focused solely on those groups, while all other communities appeared only in combination with others. Among papers that included multiple communities, the most frequent pairing was DHH and BLV (30.3%, 10/33). These papers often evaluated systems offering both subtitles and audio description (e.g. [97, 306]). Other combinations involved DHH with general disability (24.2%, 8/33), cognitive impairments and older adults (both 15.2%, 5/33), motor impairments (9.1%, 3/33), and the 'other' category (6.1%, 2/33).

When compared to Mack et al. [181], the DHH community is proportionally over-represented in our dataset (53.6% vs. 11.3%). The BLV community is also slightly higher (45.9% vs. 43.5%). All other communities are proportionally under-represented, with motor impairment showing the largest difference (1.7% vs. 14.2%) and autism absent entirely.

Participant Group	Papers	Mack et al.	This code only	Mack et al.
People w/ disabilities	104 (81.9%)	84.7%	66 (52.0%)	44.9%
People w/o disabilities	51 (40.2%)	32.1%	20 (15.7%)	1.0%
Older adults	7 (5.5%)	8.4%	1 (0.8%)	3.1%
Specialists	8 (6.3%)	17.0%	1 (0.8%)	1.9%
Caregivers	1 (0.8%)	9.4%	0 (0.0%)	0.8%

Table 3.8: The frequency of applied codes for *participant groups*, along with data reported by Mack et al. [181].

Participant groups

Table 3.8 shows the distribution of participant groups in user studies. The percentages are based on the 127 papers that included some form of user study (70.2%, $N = 127$). Participants with disabilities appeared in most user studies (81.9%, 104/127). People without disabilities were included in 40.2% of user studies ($N = 51$), specialists in 6.3% ($N = 8$), older adults in 5.5% ($N = 7$), and caregivers in one study (0.8%, $N = 1$).

Studies involving more than one group were common, with the most frequent pairing was participants with and without disabilities (24.4%, $N = 31$). Pairings of participants with disabilities and specialists occurred in 4.7% ($N = 6$), with older adults in 3.9% ($N = 5$), and people without disabilities and older adults in 3.1% ($N = 4$). Other combinations appeared once or not at all.

When compared to findings reported by Mack et al. [181], most participant groups were proportionally under-represented. People without disabilities were the most over-represented group (40.2% vs. 32.1%). Specialists (6.3% vs. 17.0%), caregivers (0.8% vs. 9.4%), older adults (5.5% vs. 8.4%), and participants with disabilities (81.9% vs. 84.7%) were all under-represented in our sample.

For studies with a single participant group, people with disabilities accounted for 52.0% compared to 44.9% in Mack et al. [181], and people without disabilities accounted for 15.7% compared to 1.0%. Caregivers (0.0% vs. 0.8%), specialists (0.8% vs. 1.9%), and older adults (0.8% vs. 3.1%) were also slightly under-represented.

Proxies and ability-based comparison

Proxies appeared in 10.2% (13/127) of papers that included a user study. This proportion is slightly higher than the value reported for the broader HCI accessibility community (10.2% vs. 8.0%). Proxies were generally used as substitutes for the intended community when providing feedback on prototypes, often due to convenience sampling (e.g. [11, 189, 306]).

Ability-based comparisons were identified in 9.4% ($N = 12$) of user studies, which is lower than the proportion reported by Mack et al. [181] (9.4% vs. 13.6%). These comparisons involved participants without reported disabilities serving as a control group against which performance or preferences of participants with disabilities were measured (e.g. [87]). Some studies also separated participants into distinct groups for task performance comparisons (e.g. [113]).

3.3.3 Research methods

In this section, we examine how user studies were conducted. We look at the methods applied in the user studies, the locations where they took place, the extent to which participatory design was used, and the reported sample sizes across different contexts.

User study methods

Table 3.9 shows the distribution of methods used in papers that included user studies. Questionnaires were the most common (75.6%, 96/127), followed by usability testing (57.5%, 73/127). Controlled experiments appeared in 40.2% of studies ($N = 51$), and interviews in 37.8% ($N = 48$). Workshops and design sessions were less frequent (11.8%, $N = 15$), as were focus groups (10.2%, $N = 13$). Case studies and field work were included as codes, but did not occur in the dataset.

Of the 127 papers with a user study, 110 (86.6%) used more than one method. The most frequent combination was usability testing and questionnaire (21.3%, $N = 27$).

Study method	Papers	Mack et al.	This code only	Mack et al.
Controlled experiment	51 (40.2%)	11.5%	8 (6.3%)	11.5%
Survey/questionnaire	96 (75.6%)	1.3%	1 (0.8%)	1.3%
Usability testing	73 (57.5%)	41.7%	5 (3.9%)	9.6%
Interviews	48 (37.8%)	42.1%	1 (0.8%)	5.7%
Focus groups	13 (10.2%)	5.9%	2 (1.6%)	0.8%
Case study	0 (0.0%)	4.0%	0 (0.0%)	0.2%
Field study	0 (0.0%)	17.8%	0 (0.0%)	4.6%
Workshop/design	15 (11.8%)	18.4%	0 (0.0%)	3.1%
Other	0 (0.0%)	16.1%	0 (0.0%)	0.8%

Table 3.9: The frequency of applied codes for *study method*, along with data reported by Mack et al. [181].

A subset of these also included interviews (12.6%, $N = 16$). Controlled experiments and questionnaires appeared together in 19.7% ($N = 25$), with seven of these also adding interviews (5.5%). Controlled experiments used alone accounted for 6.3% ($N = 8$), which represents 47.1% (8/17) of all papers that relied on a single method. All other combinations occurred in fewer than 5% of papers.

User study locations

Study location	Papers	Mack et al.	This code only	Mack et al.
Near/at research lab	29 (22.8%)	27.3%	20 (15.7%)	19.5%
Home/residence/school	12 (9.4%)	28.9%	7 (5.5%)	17.8%
Neutral location	6 (4.7%)	6.7%	1 (0.8%)	3.1%
Online/remote	21 (16.5%)	20.5%	16 (12.6%)	10.1%
Other	73 (57.5%)	41.1%	71 (55.9%)	28.1%

Table 3.10: The frequency of applied codes for *study location*, along with data reported by Mack et al. [181].

Looking at the distribution of user study locations, authors mostly did not specify where studies were conducted (see Table 3.10). The label 'other' was applied to 56.7% (72/127) of papers, primarily because the location was unclear. For example, Aydin et al. [19] describe participant recruitment, screening procedures, compensation, and equipment used, but do not explicitly state where the study took place.

Among papers that reported a location, 23.6% ($N = 30$) took place near or at a research lab, with some explicitly referring to a 'laboratory study' in the abstract (e.g. [159]). Online or remote studies were the second most common, accounted for 16.5% ($N = 21$). Studies conducted at home, in a residence, or at a school represented 9.4% ($N = 12$), and neutral locations accounted for 4.7% ($N = 6$). Twelve papers (9.4%) reported using more than one location, with the most frequent combination was research lab and online (3.9%, $N = 5$), followed by neutral location and online (3.1%, $N = 4$).

When compared to findings reported by Mack et al. [181], the 'other' label was more common in this dataset than in ours (57.5% vs. 41.1%). All other locations were proportionally lower, with home, residence, or school showing the largest difference (9.4% vs. 28.9%), followed by online or remote (16.5% vs. 20.5%), near or at a research lab (22.8% vs. 27.3%), and neutral location (4.7% vs. 6.7%).

Participatory design

Participatory design (PD) involves end users in the research process directly, see Section 2.3.1 for more detail. Within our dataset, 17.3% user study papers (22/127) adopted PD methods, which is higher than the proportion reported by Mack et al. [181] (17.3% vs. 10.3%). Studies that included caregivers ($N = 1$), specialists ($N = 6$), and older adults ($N = 4$) tended to use PD more frequently.

PD always involved either people with disabilities (81.8%, 18/22), older adults (9.1%, 2/22), or both (9.1%, 2/22). Studies focusing on IDD ($N = 1$), motor impairments ($N = 2$), and other cognitive impairments ($N = 3$) also used PD more often than studies targeting other communities. Research involving BLV participants ($N = 11$) and DHH participants ($N = 9$), as well as papers coded under general disability ($N = 2$), rarely adopted PD methods.

Sample sizes

Group	N	Median	Mean	IQR	Range
Overall	124	30.0	52.0	34.3	2-602
BLV	58	20.0	46.4	27.8	2-602
DHH	66	34.0	59.8	35.8	9-314
Motor	3	25.0	25.0	5.0	20-30
IDD	1	18.0	18.0	0.0	18-18
Other cognitive	5	25.0	62.4	10.0	18-219
Older adults	6	39.0	64.7	11.8	16-219
General disability	4	17.5	43.0	34.5	4-133
Other	2	115.5	115.5	103.5	12-219
People w/ disabilities	102	30.0	48.9	37.0	2-602
People w/o disabilities	50	34.0	73.8	34.5	12-602
Older adults	7	30.0	30.4	17.0	16-45
Specialists	8	25.5	35.8	27.5	4-133
Caregivers	1	18.0	18.0	0.0	18-18

Table 3.11: User study participant count for the 124 papers that clearly reported a user study and a sample size (3 user study papers did not report sample sizes), broken down by *community of focus* and *participant group*.

We analysed sample sizes for the 124 user study papers that clearly reported participant numbers, the results of which are shown in Table 3.11. Three papers mentioned running a study but did not provide a total count, such as Konstantinidis et al. [157] who report a user study and its results, but no participant count.

The median number of participants was 30 ($M = 52.0$, $IQR = 34.3$, $SD = 73.5$), which is higher than the median reported for the broader HCI accessibility community [181] (30 vs. 18). Studies that included people with disabilities or older adults had a median of 29.5 ($Mean = 47.7$, $IQR = 39.0$, $SD = 68.5$). Median sample sizes varied across communities of focus, ranging from 18 for those including people with

IDD, to 115.5 for user studies that were coded as 'other'. For participant groups, medians ranged from 18 for caregivers to 34 for people without disabilities.

Most communities of focus and participant groups had only a small number of papers reporting user studies. Papers involving BLV participants ($N = 58$) and DHH participants ($N = 66$), as well as studies including people with disabilities ($N = 102$) and without disabilities ($N = 50$), were the only cases with double-digit counts. Low counts for other groups make it difficult to determine whether reported sample sizes are representative.

3.3.4 Accessibility interventions

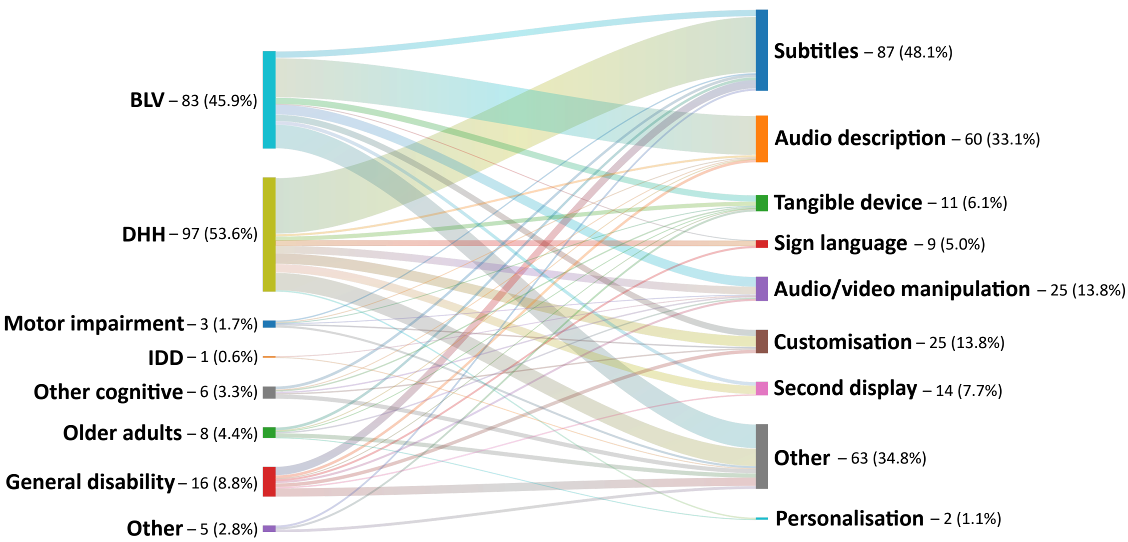

Figure 3.3: Sankey diagram representing the *type of intervention* explored with different *community of focus*.

We now look at the accessibility interventions explored in the literature and their role in accessibility research for audiovisual media. As shown in Table 3.12, subtitles (48.1%, $N = 87$) and AD (33.1%, $N = 60$) were the most frequently studied interventions, together accounting for 77.3% (130/181) of all papers in our dataset. When grouped by community of focus, 82.5% of papers ($N = 80$) addressing the DHH community implemented subtitles, while 67.5% ($N = 56$) of papers focusing on BLV participants used AD. Subtitles were also applied in research targeting general disability ($N = 12$), other cognitive impairments ($N = 4$), motor impairments ($N = 2$),

Type of intervention	Papers	This code only
Subtitles	87 (48.1%)	48 (26.5%)
Audio description	60 (33.1%)	35 (19.3%)
Tangible device	11 (6.1%)	1 (0.6%)
Sign language	9 (5.0%)	2 (1.1%)
Audio/video manipulation	25 (13.8%)	9 (5.0%)
Personalisation	2 (1.1%)	0 (0.0%)
Customisation	25 (13.8%)	0 (0.0%)
In-person assistance	0 (0.0%)	0 (0.0%)
Second display	14 (7.7%)	1 (0.6%)
Voice commands	0 (0.0%)	0 (0.0%)
Other	63 (34.8%)	11 (6.1%)

Table 3.12: The frequency of applied codes for *type of intervention*.

and older adults ($N = 4$). In contrast, AD was largely confined to BLV research, with only four papers reporting its use outside this community. Figure 3.3 illustrates how interventions were distributed across communities.

Beyond subtitles and AD, other interventions appeared less frequently. Audio or video manipulation accounted for 13.8% ($N = 25$) of papers, including work on zoom magnification to assist people with central vision loss [62]. Customisation was reported in 13.8% ($N = 25$) of papers, such as systems enabling viewers to adjust aspects of ASL interpreting [161]. Second displays (7.7%, $N = 14$), tangible devices (6.1%, $N = 11$), sign language interpreters (5.0%, $N = 9$), and personalisation (1.1%, $N = 2$) were rarely explored. Papers coded as ‘other’ ($N = 63$) included interventions outside predefined categories, for example interactive video bookmarking that allows viewers to revisit sections they found difficult to understand [306].

Multiple interventions were reported in 40.9% ($N = 74$) of papers. The most frequent combinations involved subtitles and AD alongside interventions coded as ‘other’. For instance, Matousek et al. [185] proposed a system that uses subtitles to generate highly intelligible text-to-speech dubbing for DHH viewers, improv-

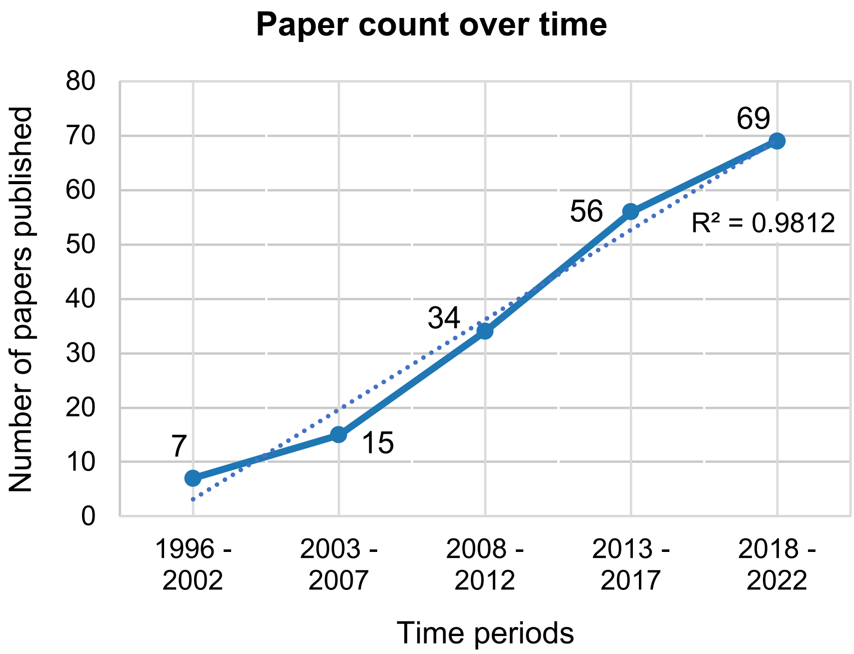

Figure 3.4: Number of papers on audiovisual media accessibility published over time, showing the number of papers increases with time. The R^2 statistic represents the trend between the number of papers published and time period.

ing speech clarity against background sounds. Most combinations (77.0%, 57/74) paired subtitles and/or AD with additional interventions. Considering contribution types [321], most papers were classified as artefact (56.4%, $N = 102$), empirical (44.8%, $N = 81$), or theoretical (30.9%, $N = 56$).

3.3.5 Trends and keywords

The previous sections examined the overall state of accessibility research in audiovisual media. In this section, we focus on how this research has evolved over the 27-year period covered by our dataset. We look at shifts in the communities that research focuses on, the accessibility challenges addressed and the technical interventions explored, and the kinds of user studies that have been conducted. We then examine how the keywords used by authors have evolved, as these terms reflect the language and priorities of the field.

To reduce year-on-year variability and provide a clearer view of long-term patterns, papers were grouped into five time periods: 1996–2002, 2003–2007, 2008–2012, 2013–2017, and 2018–2022. It should be noted that the review was completed before the end of 2022, so papers published late in that year, as well as earlier papers that had not yet been indexed or cited, are not included. This may limit the presence of more recent interventions.

Figure 3.4 shows the overall growth in publications on accessibility interventions for audiovisual media. The number of papers increased steadily across the five periods, from 7 papers between 1996 and 2002 to 69 papers between 2018 and 2022. Alongside this analysis of publication trends, we also report programmatic analysis of keyword usage to examine how terminology and research themes have shifted over time.

Trends in community of focus

Across all time periods, papers primarily focused on the DHH and BLV communities. Together, these groups appeared in more than 90% of papers in every period, which reflects the emphasis on addressing sensory barriers in audiovisual media.

Research labelled as targeting general disability began to appear in the 2008–2012 period. This coincided with a decline in papers coded as 'other', which decreased from 14.3% in 1996–2002 to 2.9% in 2003–2007. Other communities did not show consistent trends and appeared intermittently. For example, research on interventions for older adults was absent in the 2003–2007 and 2018–2022 periods, but present at modest levels in 1996–2002 (14.3%), 2008–2012 (8.8%), and 2013–2017 (7.1%).

Looking at participants in user studies, people with disabilities and people without disabilities accounted for most participation across all periods, with a combined presence of at least 81.4% of user study papers. A relative decline in these two groups during 2013–2017 coincided with an increase in participation by older adults (10.2%) and specialists (8.5%).

Accessibility challenges and interventions

Across all time periods, two accessibility challenges dominated the research focus, that of viewing video and reading subtitles. Together, these accounted for more than half of all papers and generally increased over time, rising from 57.1% in the first period to 91.3% in the most recent. Other challenges were addressed at a relatively stable rate, with notable presence between 1996 and 2007 and again during 2013–2017. The five-year period from 2013 to 2017 saw a marked increase in the number of papers tackling additional challenges. Hearing audio rose from 5 papers across all previous periods to 13 papers in this period. Papers addressing following narrative increased from 12 to 13, understanding speech from 5 to 8, and challenges coded as 'other' from 7 to 16. These changes explain the relative decrease in the proportion of papers on viewing video and reading subtitles, which reflects diversification rather than reduced interest.

Patterns in accessibility interventions show a similar concentration. Subtitles and AD remained the most common interventions throughout, but their combined prevalence declined over time, from 77.8% in the earliest period to 46.5% in the most recent. This shift coincided with growing attention to other interventions. Those labelled as 'other' initially increased before later declining, while second displays and sign language interpretation gained more focus in recent years, reaching 10.1% and 7.2% respectively in the final period. Although no single alternative intervention accounted for a large share of papers, the overall number of distinct interventions studied expanded, indicating a gradual broadening of research beyond the dominant approaches.

Trends in user study methods

User studies were present in every time period, with a slight decline during 2003–2007 before increasing steadily in later years, as shown in Figure 3.5. The distribution of study methods remained largely consistent. Questionnaires, usability testing, controlled experiments, and interviews were the most common approaches throughout the dataset. A smaller number of papers reported workshops or design sessions, and focus groups began to appear in the 2008–2012 period.

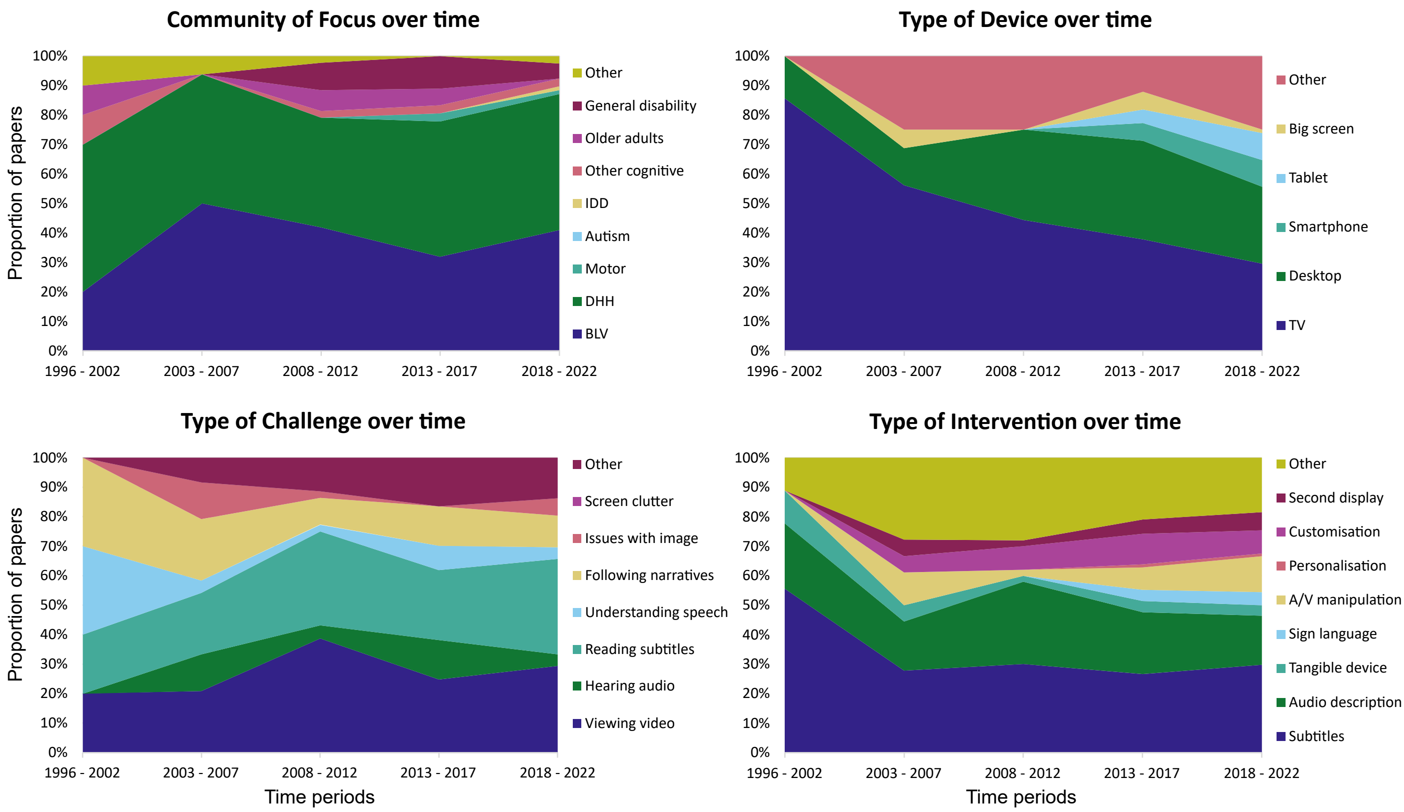

Figure 3.5: Proportion of papers published over time that addressed *community of focus*, *type of device*, *type of challenges*, and *type of intervention*.

Reporting of study locations was often unclear. Indeed, most papers were coded as ‘other’ because authors did not specify where studies took place. Excluding these, 55 papers explicitly stated the location of their user study. Within this smaller sample, some patterns emerge. Studies conducted in neutral locations and in homes, residences, or schools declined over time, while online and remote studies became more frequent, reaching 65.4% in the most recent period. Year-on-year data shows a sharp increase in remote studies after 2020, rising from 4 before 2020 to 17 afterwards, a change likely driven by the COVID-19 pandemic.

Device distribution also shifted over time, with early studies focused heavily on television, which accounted for 85.7% of papers in the first period, but this fell to 37.7% in the most recent period. Desktop computers gradually replaced television as the dominant device, followed by the emergence of smartphones and tablets starting in 2013–2017. Devices coded as ‘other’, which include cases where no device was specified, remained relatively common across all periods.

Trends in keywords

Patterns in keyword usage generally align with the qualitative coding reported earlier. Among the ten most frequent keywords, the only communities of focus represented were BLV (64 occurrences) and DHH (28), and the only interventions were subtitles (73) and AD (67). This reflects the dominance of these topics in the literature.

Several shifts in terminology and emphasis are evident over time. Keywords related to television declined from 13.3% in the earliest period to 6.3% in the most recent, and film dropped from 5.1% to 0.6%. In contrast, AD rose sharply to become the most frequent keyword during 2008–2012 (12.1%) before falling to 6.5% in the latest period. Subtitles maintained steady use across all periods, reaching 11.3% between 2018 and 2022.

New technical terms began to appear in later years, such as artificial intelligence (AI) and machine learning (ML), which emerged in the 2013–2017 period and increased significantly in the following period, from 0.3% to 3.9%, often in the context of automating interventions such as subtitles and AD (e.g. [314, 329]). Keywords related to the web and online platforms rose to 6.1% in 2008–2012 but declined to 1.5% more recently. The term 'accessibility' showed a consistent upward trend, increasing from 2.6% in 2003–2007 to 11.0% in the most recent period, which aligns with broader findings that accessibility is gaining prominence as a research area [181].

3.4 Discussion

This systematic review has outlined the current state of research on accessibility interventions for digital audiovisual media and identified areas that receive disproportionate attention compared to the broader HCI accessibility field. In this section, we interpret these findings in relation to our research questions. We consider the challenges faced by people with disabilities when accessing audiovisual media, the interventions proposed to address these challenges, and the communities that research has focused on. Finally, we discuss gaps in the literature and how we address these in future chapters.

3.4.1 Challenges faced by people with disabilities

We examined the challenges reported in the literature to address RQ1. This research question concerns the main accessibility challenges faced by people with disabilities when accessing digital audiovisual media. Our examination included the type of content, viewing device, and viewing context, finding that research has concentrated on conventional viewing scenarios. Most papers focused on general viewing contexts (81.8%), with television (45.9%) and desktop computers (32.6%) as the primary devices, and a preference for TV broadcasts (32.0%) and video-on-demand services (30.9%). Examples include work by Thorn and Thorn [293] on subtitle presentation rate or Wolffsohn et al. [326] on real-time edge detection and image enhancement for TV broadcasts.

While improving accessibility for common viewing patterns is important, this emphasis has left notable gaps. Few papers addressed challenges such as following narrative, hearing audio, or understanding speech, and most of these relied on subtitles as the primary solution. Alternative approaches exist, such as object-based audio technologies (e.g. [273, 316]), but these remain under-explored. This lack of diversity in interventions aligns with our findings that subtitles and AD dominate the field, together accounting for 77.3% of all interventions. These patterns suggest that accessibility research has prioritised sensory barriers over comprehension barriers, leaving issues such as narrative understanding and cognitive load insufficiently addressed.

Emerging media formats introduce further gaps. Immersive environments such as AR and VR are gaining popularity, yet research rarely considers the accessibility challenges they pose. Only isolated studies examined subtitle readability in VR or dynamic placement in 360-degree video, despite the growing adoption of these technologies. Similarly, mobile devices have become central to media consumption (e.g. [76, 241]), but accessibility research has not kept pace. Existing work largely applies traditional interventions to mobile contexts, such as second-screen subtitles (e.g. [61, 192, 247]), rather than developing solutions tailored to mobile viewing patterns. This gap is particularly evident for short-form social media content, such as TikTok (e.g. [315]), which is now a dominant mode of consumption. The absence of research on accessibility for social media video is striking given its rapid growth and unique interaction patterns.

Live streaming is another area with limited attention, despite its rapid increase in viewership [115, 323]. Real-time features could enable novel interventions, such as caption highlighting or pausing to support DHH viewers [162], but these approaches remain uncommon. Given the increasing prevalence of synchronous media experiences, future research should explore interventions that leverage real-time interactivity to improve accessibility. Addressing these gaps will require moving beyond established solutions and considering how accessibility challenges intersect with evolving technologies and viewing behaviours.

3.4.2 Communities of focus for accessibility research

To address RQ2, which concerns the communities that accessibility research in audiovisual media focuses on, our analysis shows that an overwhelming majority of papers target the DHH and BLV communities. Given the nature of audiovisual media, which combines audio and visual information, it is unsurprising that interventions for hearing and visual impairments dominate the literature. However, this emphasis has resulted in significant under-representation of other communities. Autism did not appear at all in our dataset (0.0%), while IDD accounted for only 0.6% and other cognitive impairments for 3.3%. These groups face distinct challenges when accessing audiovisual content, requiring approaches beyond those designed for sensory impairments [243]. The absence of autism and the low representation of cognitive impairments contrast sharply with their prevalence in broader HCI accessibility research, as reported by Mack et al. [181].

Older adults were also under-represented, along with conditions not explicitly coded in our review, such as aphasia or dementia. Both occur more frequently in ageing populations and are likely to become more prominent as demographic trends shift [202, 300]. Language impairments such as aphasia demand specialised interventions, as common solutions like subtitles may not fully address comprehension barriers [109, 249]. This gap is notable given that comprehension-related challenges were rarely prioritised in the dataset, reinforcing the need for research that goes beyond sensory accessibility. The under-representation of these communities, combined with their prevalence in the wider accessibility landscape, suggests an important area for future work. Research should explore interventions tailored to cognitive and language-related disabilities, as well as strategies for older adults who may experience multiple overlapping impairments.

Participation practices also warrant attention. As Bannon et al. [21] note, PD is often reduced to minimal involvement, such as a single feedback session – see Section 2.3.1 for more details. Many papers in our dataset followed this pattern, limiting opportunities for meaningful engagement. PD can help researchers understand user needs and design interventions that reflect real-world contexts (e.g. [191]). Iterative approaches, starting with low-fidelity prototypes, allow designs to evolve in collaboration with users (e.g. [147, 310]). We encourage future studies to adopt richer PD methods and involve a broader range of stakeholders, including specialists and caregivers, who can provide valuable insights or support participants during research [50]. Given that only 17.3% of user study papers in our dataset employed PD, and that specialists and caregivers were rarely included, this represents a clear methodological gap.

3.4.3 Interventions for addressing accessibility

To address RQ3, which concerns the interventions used to support different accessibility needs, we examined both the types of interventions explored and the methods applied in user studies. Subtitles (48.1%) and AD (33.1%) were by far the most common interventions. This dominance is expected, as these techniques are the established standards for making audiovisual media accessible to DHH and BLV communities, which were the most represented groups in our dataset, and shows the importance of continuing work to iteratively refine these interventions. However, this concentration also reflects a narrow scope, leaving many potential research directions under-explored.

The evolution of technology offers opportunities to move beyond conventional solutions. Immersive media formats, such as AR and VR, introduce new accessibility challenges that require innovative approaches. Our dataset shows only isolated work on subtitle readability in VR and dynamic placement in 360-degree video, despite the growing adoption of these technologies. Similarly, the rise of personalised and adaptive content delivery suggests the need for interventions that can be tailored to individual requirements rather than relying on fixed formats.

Object-Based Media (OBM) exemplifies this potential. By decomposing content into constituent elements and rendering them dynamically, OBM enables personalised experiences [130] – see Section 2.2 for more details. Yet current applications of OBM largely replicate existing interventions, such as subtitles and AD (e.g. [192]). Only limited work explores its capacity for highly configurable accessibility features, such as adapting incidental audio channels for DHH viewers (e.g. [317]). This gap indicates an opportunity to design interventions that leverage OBM for cognitive and comprehension-related challenges, which remain under-addressed.

Recent advances in machine learning (ML) and artificial intelligence (AI) also open new possibilities. These technologies have been applied to automate the generation of subtitles (e.g. [184]) and AD (e.g. [314, 329]), reducing production costs and improving scalability. Beyond automation, AI could enable dynamic transformation of content to match user needs, such as adjusting complexity or visual emphasis, or even generating alternative representations using text-to-image models. Given the increasing presence of AI-related keywords in recent years, this trend is likely to accelerate, but research should prioritise approaches that go beyond replicating existing interventions.

3.4.4 Limitations

We identify several limitations with this systematic review, as is common in reviews of this type. Our dataset does not capture all research on accessibility interventions for audiovisual media. This is despite following PRISMA guidelines [240], additional guidance from Silva and Francila Weidt Neiva [275] and Siddaway et al. [274], and applying a snowballing procedure as described by Wohlin [322]. As outlined in

Section 3.2, the initial search used three databases (ACM DL, SCOPUS, IEEE Xplore), but IEEE Xplore was excluded after returning a high proportion of false positives. Other databases not included in the search (e.g. SAGE Journals, Elsevier, Springer, Routledge) may contain relevant papers that meet our criteria.

The search query and identification method may also have constrained the dataset. For example, the SCOPUS query limited venues to reduce the large search space and high number of irrelevant results, which could have excluded relevant work. While snowballing added papers from sources such as Springer and Elsevier, this introduced challenges in data cleaning [30]. For future consideration, reviews could consider automated deduplication or metadata harmonisation to reduce manual effort and improve consistency. Additionally, since this review has been conducted, advances in generative AI tools designed for conducting systematic review (e.g. Rayyan [1]) may help reduce certain challenges, however may also introduce others.

The creation of the codebook and the manual coding process introduced potential bias. We attempted to mitigate this by having a second researcher code 10% of the papers, discussing disagreements, and calculating inter-rater reliability. However, subjective interpretation of categories cannot be fully eliminated. It is important to note that all researchers involved in developing the codebook and analysing the dataset identified as white, cisgender male or female of European background, and none identified as disabled. Two researchers had prior experience in accessibility research. These demographic characteristics may have influenced coding decisions, which is a recognised limitation in qualitative synthesis.

Language variability posed additional challenges when applying community-of-focus labels, similar to issues reported by Brule et al. [43]. Terms such as ‘blind’, ‘visually impaired’, and ‘low vision’ were used inconsistently across papers. Comparable difficulties arose for neurodiversity and cognitive or learning disabilities, as noted in the review by Mack et al. [181]. Terminology also shifted over time, with meanings of community descriptors, technologies, and study methods evolving during the 27-year period, which we did not analyse. A longitudinal semantic anal-

[1] rayyan.ai

ysis could help future work understand how these changes affect keyword-based searches and coding. Finally, our comparison with the dataset from Mack et al. [181] is limited because their review focused exclusively on CHI and ASSETS. This restricts the generalisability of cross-review comparisons and suggests the need for broader benchmarking across multiple venues.

3.5 Conclusion

This chapter reviewed technical interventions for digital audiovisual media across 27 years and $N = 181$ papers to establish a field-level baseline. Across this literature, accessibility is treated primarily as sensory access, with subtitles and audio description having disproportionate focus, while support for cognitive or language processing being rare. The result is a body of work that improves perceptibility for many viewers, yet leaves gaps in terms of comprehension.

The distribution of communities and interventions reflects this emphasis. Most papers focus on d/Deaf or hard-of-hearing viewers and blind or low-vision viewers, with subtitles appearing in nearly half of the studies and audio description in about one third. Cognitive and language-related access is addressed only occasionally, with outcomes that speak directly to comprehension, such as following dialogue or narrative, being uncommon. Accessibility is therefore framed less in terms of whether viewers can keep up with what unfolds on-screen, and more in terms of whether they can perceive audiovisual elements. Interventions have also failed to keep pace with where and how people watch, adapting interventions designed for one context to another.

Participation practices also limit what can be inferred about real use. Although user studies are common, participatory design is rare and proxies are still employed. Study settings are often unspecified, with relatively few conducted in homes or other routine environments. This matters because comprehension is shaped by the viewing context (e.g. co-viewing, interruptions, and habitual practices) that laboratory sessions cannot capture. Limited attention to domestic contexts also obscures the negotiation that occurs when viewers balance their own access needs with those of others who share the screen.

Over time, the literature diversifies without shifting its centre of gravity. Paper counts rise and the relative share of subtitles and audio description falls as new ideas appear, yet few are evaluated in terms of whether viewers actually understood. Recent algorithmic work on automation and assistive authoring largely reproduces established interventions. Feature availability continues to stand in for access, with little agreement on how to assess understanding during ongoing playback.

Against this backdrop, this thesis treats comprehension as the outcome of interest. For viewers facing language-processing barriers, perceptibility alone is insufficient – elements of the audiovisual media can demand significant cognitive load. The reviewed literature rarely targets these factors directly, and it offers limited accounts of methods that reduce effort while remaining in the flow of viewing.

This chapter therefore motivates a shift from the presence of features to the management of effort. Rather than asking whether subtitles or descriptions exist, the question becomes whether a viewer can follow events with confidence, track speakers, and stay oriented as a programme unfolds. This re-framing supports a route to accessible personalisation – optional changes that can reduce cognitive load while preserving agency. These findings also shape the kinds of evidence required, with heavy reliance on short, controlled tasks obscures how access is achieved across extended viewing or in the presence of others. Measures of perceived effort, followability, and recovery are needed alongside qualitative accounts of when meaning is lost and how viewers attempt to regain it. The review points to a shortage of such approaches and to the value of designs that accommodate ordinary routines. If the field prioritises perceptibility while leaving comprehension under-specified, the next step is to locate where understanding breaks during everyday viewing for people with aphasia and to identify points in the audiovisual media where changes reduce effort. The chapters that follow take up that task by working outward from lived experience, maintaining the same organising thread: access should centre on comprehension. The remainder of the thesis uses RtD intervention artefacts to examine that thread in practice, moving from reported breakdowns to prototype changes that can be tried during viewing.

Chapter 4

Accessibility Barriers and Facilitators

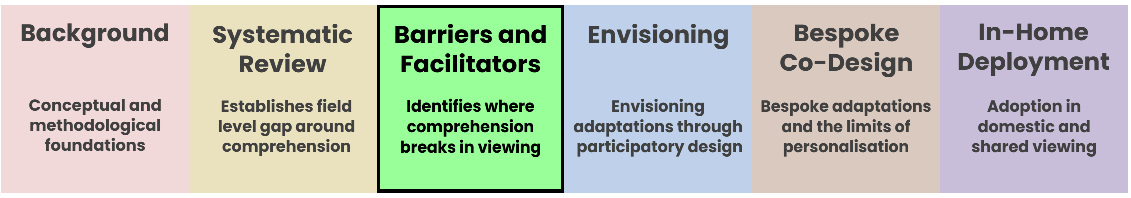

Figure 4.1: Map of thesis - barriers and facilitators

In this chapter we investigate the accessibility barriers and facilitators that people with aphasia experience when watching digital audiovisual media. We do this by first conducting an online survey, followed by an in-person focus group. Through this, we identify specific access barriers and facilitators which inform our later work – see Figure 4.1.

4.1 Introduction

As outlined in previous chapters, research on audiovisual media accessibility has largely under-represented people with aphasia. One consequence of this under-representation has been that the accessibility challenges faced by people with aphasia have not been mapped, since prior work had leaned on assumptions drawn from the communication difficulties experienced by people with aphasia, rather than direct evidence. A second consequence is that the claim that aphasia negatively affects the viewing experience itself has not been examined directly by the literature.

In this chapter we ask two questions. Does aphasia significantly affect the viewing experience, and which aspects of digital audiovisual media disrupt comprehension during viewing for people with aphasia? The chapter focuses on comprehension during viewing, including situations where access services are present.

We use two studies to answer these questions. We ran an online survey with 41 respondents with aphasia to examine how they engage with audiovisual media and to test whether aphasia is associated with a poorer viewing experience. Informed by the survey results, we ran a focus group with 10 participants with aphasia to identify specific barriers in digital audiovisual media and to surface potential facilitators. Speech and language therapists supported the sessions so that participants could contribute using verbal, non-verbal, and tangible methods.

4.2 Online Survey

4.2.1 Method

We followed mitigation strategies to improve accessibility for people with aphasia. In line with nationally recommended guidelines on preparing materials for people with aphasia [120], and informed by feedback from SLTs and researchers experienced in aphasia, we divided the survey into thematic blocks of no more than six related questions. Questions were accompanied by representative images, such as a picture of a person reading a book for the item on reading ability. We piloted the survey with SLTs and people with aphasia to confirm clarity and reading load. We used skip logic to hide inapplicable items and reduce reading effort, including a progress indicator and optional breaks to manage fatigue. We hosted the survey on Qualtrics [1], and received approval from the King's College London ethics board. The survey was anonymous, requesting only age, gender identity, and experience with audiovisual media.

The survey began with informed consent. Participants first read information

[1]https://qualtrics.com

about the research and then answered true or false questions about the study to check attention. We included screening questions to confirm eligibility – adults with aphasia for at least six months. Participants then answered demographic items, including age and gender identity, followed by self-evaluation of reading, writing, speaking, and understanding on a five-point Likert scale from ‘very hard’ to ‘very easy’. Most answers used multiple choice to reduce text-entry demands, with images accompanying each option. Optional open-text fields allowed elaboration, accepting dictation or caregiver assistance where preferred, noting when assistance was used. We designed all items to be readable on smartphones and tablets with large touch targets.

The next section asked about the viewing context: devices used, interaction methods with those devices, amount of audiovisual media consumed, types of media consumed, and concurrent activities while viewing. Finally, participants described their viewing experience prior to aphasia and with aphasia. This included overall viewing experience and specific aspects such as understanding dialogue, following narrative, reading on-screen text, and tracking who the speaker was. These questions used a five-point Likert scale ranging from ‘I found/find it very hard’ to ‘I found/find it very easy’ and included smiley face icons. An optional open-text field captured additional details or examples.

4.2.2 Participants

We shared the survey with four aphasia support groups and charities in the UK, the US, and Australia through their social media pages (e.g. Facebook) and other online resources (e.g. blogs). This route aimed to reach people with aphasia who already took part in online communities and were comfortable with technology, as well as aligning with how these organisations typically communicate with members. We received responses from 41 people with aphasia, 21 who identified as female and 20 as male. Ages ranged from 30 to 90, with a mean of 59.44 (SD = 14.62). Respondents could be assisted by another person, such as a friend or family member; 4 reported such support. Participants evaluated their language abilities on a five-point Likert scale, where 1 represented 'very hard' and 5 represented 'very easy'. Mean score for reading was 2.78 (SD = 1.26), writing 2.10 (SD = 1.00), speaking 2.20 (SD = 1.12), and understanding speech 3.41 (SD = 1.20). These self-reported ratings informed later analysis of viewing experience. We did not classify participants as having receptive or expressive aphasia, since doing so would have required extensive assessment by an SLT and would have introduced clinical procedures beyond the scope of the survey. To thank respondents for their time, they could opt into a prize raffle for one $50 and two $25 Amazon vouchers.

4.2.3 Data Analysis

We analysed the data in two stages. First, we ran a statistical analysis of the viewing experience, covering the overall score and the individual viewing aspects. We compared paired scores reported for before aphasia and with aphasia using the Wilcoxon Paired Signed-Rank test [180], chosen to suit paired ordinal responses. We reported an estimated effect size following the approach recommended by Rosenthal et al. [255] to give a comparable magnitude for the observed differences, since this test does not yield a standardised effect size. We then examined how self-reported language abilities related to each viewing aspect by computing Pearson correlations [125] to assess linear associations between ability scores and aspect scores. Second, we analysed the qualitative answers describing barriers to accessing audiovisual media. We assigned each comment to one of two temporal categories, indicating whether it referred to viewing before aphasia or with aphasia, and we labelled it by the viewing aspect it concerned. This organisation allowed us to compare patterns across the qualitative and quantitative datasets and to describe barriers and experiences using a consistent set of viewing aspects.

4.2.4 Survey Results

Quantitative Results

We report results from the online survey (N = 41), which asked about aphasia, the viewing context, and experiences with consuming audiovisual media. Answer frequencies for context items, including devices used, interaction methods, content types, and concurrent activities while viewing, are shown in Figure 4.2. These items established the typical settings in which respondents watched, providing a basis for interpreting subsequent experience scores.

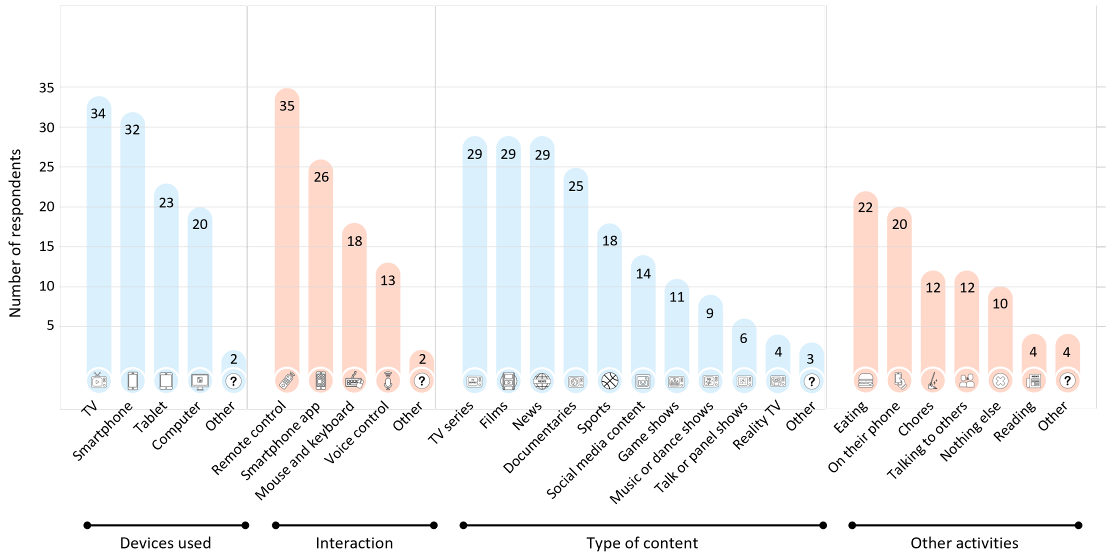

Figure 4.2: The viewing context of survey respondents showing the frequency of responses for the devices used, how they interact with those devices, the types of content watched, and what activities they do while watching.

Viewing experience is significantly worse with aphasia. The average score on a five-point Likert scale decreases from 4.55 (SD = 0.71) to 2.83 (SD = 0.90), a drop of 1.73 points out of 5 (34.6%). All measured viewing aspects show similar declines, as illustrated in Figure 4.3. Reading on-screen text shows the largest decrease (−2.00/5, 40.0%), indicating that time-pressured text remains a pronounced difficulty. A Wilcoxon Signed-Rank test comparing paired before-aphasia and with-aphasia scores confirms the difference for overall viewing experience ($Z = -4.897$, $p < .001$), with a strong estimated effect size ($r = -0.828$). Test results for individual aspects follow the same pattern and are listed in Table 4.1. Together, these paired comparisons indicate that respondents experienced broad deterioration across core tasks involved in watching digital audiovisual media.

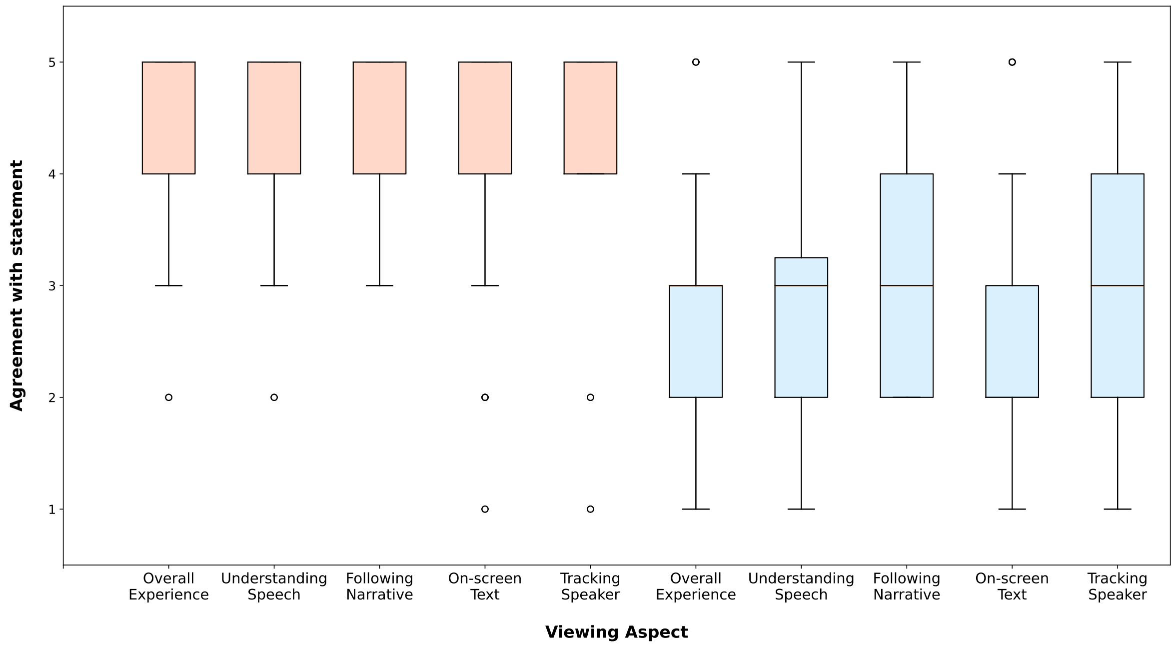

Figure 4.3: Agreement with the statement “*Did/do you struggle with [viewing aspect]*”, with 1 representing “I found/find it very hard” and 5 representing “I found/find it very easy”. The box plots in orange represent the respondent’s experience before aphasia and the blue ones represent their experience with aphasia.

We then examined associations between self-reported language abilities and viewing outcomes using Pearson correlations. Overall language ability correlates positively with overall viewing experience with aphasia ($r = 0.541$, $p < .001$), indicating that lower ability is associated with poorer viewing. Understanding ability shows a large positive correlation with overall viewing experience ($r = 0.533$, $p < .001$), while speaking ability shows a moderate positive correlation ($r = 0.420$, $p = .007$). In contrast, overall viewing experience does not correlate significantly with reading ($r = 0.164$, $p = .312$) or writing ($r = 0.245$, $p = .127$). These relationships are consistent with the demands of following spoken content, where comprehension of speech contributes more directly to maintaining orientation than written production skills.

	Overall experience	Understanding dialogue	Following narrative	On-screen text	Tracking speaker
Mean Difference	-1.73	-1.70	-1.58	-2.00	-1.55
Z-value	-4.952	-4.852	-5.064	-5.074	-5.131
Sample size (N)	34	30	33	33	34
Effect size (r)	-0.828	-0.873	-0.855	-0.872	-0.848

Table 4.1: Results of Wilcoxon Signed-Rank test on the different viewing aspects and their mean difference between scores prior to aphasia and with aphasia. The effect size is estimated using the formula $r = \frac{Z}{\sqrt{N}}$ [255].

Aspect-level analyses show that understanding and speaking abilities correlate positively and significantly with understanding dialogue, following narrative, reading on-screen text, and tracking the speaker after aphasia (see Table 4.2). Reading ability correlates positively with challenges related to on-screen text (r = 0.451, p = .003) and with understanding dialogue (r = 0.367, p = .020), reflecting the interdependence between reading demands and moment-to-moment comprehension during viewing. Across measures, the pattern supports the interpretation that language understanding capacities are closely tied to the quality of the viewing experience for people with aphasia.

	Overall experience	Understanding dialogue	Following narrative	On-screen text	Tracking speaker
Reading	0.164 (p=0.312)	0.367 (p=0.020)	0.215 (p=0.183)	0.451 (p=0.003)	0.099 (p=0.544)
Writing	0.245 (p=0.127)	0.114 (p=0.485)	0.233 (p=0.149)	0.357 (p=0.024)	0.103 (p=0.529)
Speaking	0.420 (p=0.007)	0.438 (p=0.005)	0.394 (p=0.012)	0.356 (p=0.024)	0.332 (p=0.036)
Understanding	0.533 (p<0.001)	0.550 (p<0.001)	0.685 (p<0.001)	0.527 (p<0.001)	0.544 (p<0.001)

Table 4.2: Correlation matrix between viewing aspects and language ability. Highlighted cells indicate statistical significance.

Qualitative Results

We collected qualitative responses to let respondents describe their viewing experiences in detail, with a focus on barriers to accessing audiovisual media. We recorded a total of 106 relevant comments from 26 respondents, excluding remarks that were irrelevant or contained no content, such as "*N/A*". Of these relevant comments, 34 concerned viewing before aphasia and 72 concerned viewing with aphasia. To organise the material, we first considered reflections on pre-aphasia viewing and then turned to experiences with aphasia, followed by comments specific to on-screen text.

Comments about viewing before aphasia tended to focus on on-screen text or subtitles. Several respondents noted that text appeared too briefly, reporting *text wasn't on screen long enough*", with one adding *It goes too quickly or I misinterpret it*". Text legibility was also raised, being described as *too small to read*", for example when reading *temperature weather forecast and sporting scores*". A smaller set of remarks addressed speech comprehension in everyday viewing, where strong accents were described as challenging, although "*very occasionally*", and some viewers said that rapid speech led to missed information. These pre-aphasia reflections framed the kinds of timing and clarity issues that later became more pronounced.

When discussing viewing with aphasia, respondents reported a broader set of barriers. Many described difficulty understanding speech because of its fast pace, rendering it *all to quick for me to comprend*". Being unable to keep up was confusing, led to misinterpretation, and was tiring. This was particularly pronounced in crowded scenes with multiple speakers, as one respondent explained: *I'm still focusing on the... what one person is saying so I miss what the next person is saying*". In contrast, situations with one person speaking at a time and clear separation between speakers were easier to follow. Loud background noise also distracted from speech or made it hard to understand, including cases where "*the background music is loud*". These accounts consistently linked pacing, turn-taking, and the sound mix with listeners' ability to maintain attention.

Narrative follow-up featured prominently. Several respondents said that missing or misunderstanding speech made it harder to follow the storyline, since *when you miss words, you cannot follow a story*". Control over playback helped some viewers recover, whether by pausing or rewinding unclear segments or by relying on support from another person: *Often my wife has to help me sort out the storyline*". This control was not always available, such as with live television, leading one respondent to *tape shows/series and go over them again*". Complex casts and shifting plots were also problematic, with multiple respondents citing difficulties recalling earlier events: *Depending on the story, some things can get quite complex and it's remembering everything that happened beforehand that's a must and having aphasia doesn't help with that*". These remarks indicated that comprehension challenges in speech often carried through to narrative understanding.

On-screen text was a recurring source of difficulty and aligned with the quantitative decrease observed for this aspect. Respondents frequently described text as too brief to read and process, stating *Text are not on screen long enough for me to read it*". This applied both to text present within the shot, as in *When there is a sign or a note... I have to pause to read it*", and to subtitles, as in *Text subtitles too fast*". Beyond timing, some viewers reported reduced enjoyment of reading after stroke: *Although my reading ability is fine, I don't enjoy reading nearly as much after my stroke, particularly more complex subjects. I suspect it's taking me a little bit more cognitive effort. This affects reading books, as well as reading text on a screen*". Others found complexity itself demanding, stating "*Too much information I prefer one or two words it's too hard for a lot of sentences*". Taken together, the qualitative responses described a pattern in which rapid speech, limited reading windows, and dense information made viewing effortful, with pace control and social support offering partial mitigation.

4.2.5 Reflections on the Survey Results

The survey examined whether aphasia affects audiovisual media consumption and how viewing contexts relate to experience. Results showed a significant negative impact on overall viewing, with declines across all measured aspects and the largest drop for on-screen text. A Wilcoxon Paired Signed-Rank test indicated a strong effect size for the overall difference. Associations between language abilities and viewing outcomes supported this pattern. Lower self-reported understanding and speaking abilities were linked to poorer viewing experience, indicating that speech comprehension played a central role in maintaining orientation during watching. Reading and writing abilities showed weaker relationships with overall experience, though on-screen text remained a frequent source of difficulty. These findings pointed to speech clarity, pace, and limited reading windows as consistent pressures on comprehension. While the survey had limitations (see Section 4.4.4), including a likely bias toward respondents with milder aphasia who participate in online communities, the evidence indicated that aphasia affected multiple aspects of viewing and that greater language difficulty was associated with more barriers. To enrich these quantitative results, we then ran focus group sessions to collect detailed accounts of barriers and facilitators in everyday viewing.

4.3 Focus Group

We ran two focus groups with people with aphasia to better understand the specific challenges they faced when accessing audiovisual media, to complement the survey results, and to triangulate responses more meaningfully. Each session combined an in-depth discussion with a structured video critiquing activity so that reflections referred to concrete, shared experiences.

4.3.1 Method

Both focus group sessions followed the same procedure. They were held a week apart and lasted about two and a half hours each. The sessions were video and audio recorded. To respect participants' privacy, they were given three options for how they would be presented in stored videos and stills, namely not to be shown at all, to be shown with their face blurred, or to be fully visible. The sessions were divided into three sections, namely informed consent and demographics, an open discussion of audiovisual viewing experiences, and a video viewing and critiquing activity, as can be seen in the bottom left image of Figure 4.4. A researcher introduced the session, explained the study aims, and guided participants through the information sheet and consent form. Communication support was provided throughout by SLTs and tangible communication aids [320]. These aids included paper-based visual materials, a tangible Likert scale that participants could point to, and pen and paper

Figure 4.4: Photos from the focus group sessions.

to enable non-verbal contributions which participants used to indicate ratings or to sketch when words were hard to retrieve. Following consent, participants completed a short questionnaire covering age, gender identity, self-perceived language abilities, and aspects of their viewing context such as the devices used and the types of content consumed. We seated participants at a single table to support turn-taking and ease of facilitation, and scheduled a short break at the midpoint to manage fatigue.

We opened with an icebreaker in which participants discussed their favourite films to establish comfort within the group. The open discussion then focused on barriers encountered when accessing audiovisual media, features that facilitated viewing, and methods used to overcome difficulties. SLTs supported the conversation by prompting quieter participants and checking understanding (e.g. recapping what had been said and inviting confirmation).

Participants then engaged in the video viewing and critiquing activity. We selected 11 video clips representing a wide range of broadcast formats such as films, documentaries, and news broadcasts, and a spread of genres, as listed in Table 4.3. Selection followed a complexity heuristic approach (see [201]) that considered intense speech, such as multiple people talking, and novelty, such as new actions or scene changes. The clips were short, ranging from 36 to 96 seconds, with an average of 63.4 seconds and a standard deviation of 18.2 seconds. These choices, while not exhaustive, were intended to elicit critical reflections on common viewing situations that participants faced and to keep cognitive demands manageable. Clips were projected on a large screen in front of all participants for the video viewing, as shown in Figure 4.4, and each was briefly introduced by a researcher before playback. Once a clip finished, ample time was given to discuss aspects that participants found challenging, as well as characteristics that facilitated viewing. SLTs assisted by asking probing questions such as whether pace, number of speakers, background sounds, or on-screen text contributed to difficulty, and by supporting responses. We encouraged participants to use the tangible Likert scale and other materials during the critique to mark perceived difficulty consistently across clips.

Name	**Description**	**Duration**
VC1	Extract from the BBC documentary series *Africa*, involving slow shots and narration.	0m44s
VC2	News story about Sri Lanka presented by a journalist, showing the journalist, relevant B-roll, and on-screen text.	1m16s
VC3	Scene from UK comedy TV show *Black Books*, involving visual comedy and unclear speech.	0m36s
VC4	Scene from the 2018 film *Bohemian Rhapsody*, involving a group discussion with many people.	1m22s
VC5	Extract from a documentary on the Volkswagen emission scandal, with a mix of slow and fast shots and narration.	1m00s
VC6	Extract from a cooking show presented by James Martin, with fast shots, narration, and background music.	0m52s
VC7	Extract from a 2020 Japanese documentary *The Seeds We Sow*, involving non-English narration (Japanese) and subtitles (in English).	0m44s
VC8	BBC news headlines presented by a host, with fast speech, on-screen text, and relevant footage.	1m00s
VC9	Scene from the 2004 comedy film *Shaun of the Dead*, involving comedy, fast-paced speech and on-screen action.	1m24s
VC11	Extract from the game show *The Weakest Link*, involving fast-paced speech.	1m36s

Table 4.3: List of video clips shown during the video critiquing activity.

4.3.2 Participants

Participants were recruited through Dyscover, an aphasia charity in the South East of England that offers support sessions and activities to people living with aphasia. We approached the charity to run the sessions and they recruited from among people who attend their support groups. Before signing up, prospective participants received information about the research and the plan for the sessions and they had the opportunity to ask questions so that consent decisions were informed.

We recruited 10 participants and ran two groups with 5 participants in each, see Table 4.4. Three SLTs with experience supporting people with aphasia took part, with at least two present in each session to provide continuous communication support. Ages ranged from 52 to 71, with a mean of 58.90 and a standard deviation of 5.28. Time living with aphasia ranged from 3 to 16 years, with a mean of 7.90 and a standard deviation of 3.69. All participants had been fluent in English before their stroke.

To support participation within groups, we stratified by the severity of expressive aphasia, which we understood as difficulty in expression with relative strength in understanding rather than receptive aphasia. Participants in Session 2 had more severe expressive aphasia. SLTs at Dyscover assessed expressive aphasia severity so that grouping reflected current communication needs. Participants received 20 GBP in the form of an Amazon voucher as compensation for their time.

The sessions took place a week apart at Dyscover, in line with recommendations to work in a familiar setting [181]. We followed an accessible consenting procedure with support from SLTs and tangible communication aids, and we allowed sufficient time for each activity, a break at the halfway point, and any pauses that participants required to manage fatigue. The room was separate from other support group activities. Participants sat together at a large table, see Figure 4.4, and a projector was available for the video viewing and critiquing activity. None of the participants used aided augmentative and alternative communication other than the tangible communication aids we provided.

4.3.3 Data Analysis

We transcribed the video recordings from both sessions in NVivo 14. The transcript captured verbal and non-verbal communication, since many people with aphasia could find verbal expression difficult [320], including P6 who relied on non-verbal communication. We therefore included instances where participants used pen

Name	Gender	Age	Language	Devices Used	Interactions	Content Consumed	Other Activities
P1	Female	54	R: 2; W: 1; S: 3; L: 3	TV; Tablet; Smartphone	Remote control; Apps; Fingers	TV series; News; Talk shows; Films; Reality TV; Social media	Chores; Using phone; Eating; Mobile games
P2	Female	61	R: 2; W: 1; S: 2; L: 2	TV; Tablet	Remote control; Fingers	TV series; News; Talk shows; Panel shows; Documentaries; Sports; Films	Nothing else
P3	Female	56	R: 2; W: 3; S: 3; L: 3	TV; Smartphone; Tablet; Kindle	Remote control; Voice control; Apps; Children help	TV series; News; Documentaries; Films; Reality TV; Facebook	Chores; Using phone; Eating
P4	Female	64	R: 2; W: 2; S: 4; L: 1	TV; Tablet; Smartphone	Remote control; Husband helps	TV series; Films; Documentaries	Using phone
P5	Female	64	R: 3; W: 3; S: 2; L: 4	TV; Tablet; Smartphone; Computer	Remote control; Mouse/keyboard; Apps	TV series; News; Talk shows; Panel shows; Documentaries; Sports; Music or dance shows; Game shows; Films; Reality TV; Social media	Eating
P6	Male	56	R: 4; W: 2; S: 1; L: 4	TV; Tablet; Smartphone	Remote control; Apps	TV series; News; Panel shows; Documentaries; Game shows; Sports; Films; Reality TV; Social media	Using phone; Reading news; Talking; Eating; Internet browsing
P7	Male	52	R: 1; W: 1; S: 3; L: 4	TV; Tablet; Smartphone; Computer; Radio	Remote control; Mouse/keyboard; Voice control; Apps	TV series; News; Social media; Documentaries; Sports; Films; Reality TV	Chores; Talking
P8	Male	58	R: 3; W: 2; S: 2; L: 3	TV; Tablet; Smartphone; Computer	Remote control; Mouse/keyboard; Voice control; Apps	TV series; News; Social media; Music or dance shows; Films; Reality TV; Documentaries	Using phone; Reading news; Talking; Eating
P9	Male	71	R: 2; W: 1; S: 2; L: 3	TV; Tablet; Smartphone; Computer;Cinema	Remote control; Mouse/keyboard; Voice control; Apps	TV series; News; Talk shows; Panel shows; Documentaries; Sports; Music or dance shows; Films; Social media	Chores; Using phone; Talking; Eating; Listening to radio
P10	Male	56	R: 2; W: 2; S: 2; L: 4	TV; Smartphone; Computer; Cinema	Remote control; Mouse/keyboard; Apps	TV series; News; Sports; Films; Documentaries; Social media	Chores; Using phone; Talking; Eating;

Table 4.4: List of participants in the focus group sessions, along with their self-reported reading, writing, speaking, and listening abilities. The self-reported language abilities (**R**eading, **W**riting, **S**peaking, **L**istening) were expressed on a five-point Likert scale, with 1 representing finding that aspect of language "very hard" and 5 representing finding it "very easy". Participants P1 to P5 attended the first session and participants P6 to P10 attended the second session.

and paper, as shown in the bottom right image in Figure 4.4, as well as physical communication, an example of which is shown in Figure 4.5. We also noted pointing to tangible materials and other observable gestures alongside spoken turns.

Figure 4.5: Series of images showing a participant using non-verbal communication cues to express that they found the video clip uninteresting.

We then analysed the transcript using thematic analysis as recommended by Braun and Clarke [37]. The aim was to identify and classify perspectives on the accessibility barriers and facilitators that participants reported when accessing audiovisual media. Coding proceeded inductively from the data and focused on barriers and facilitators that participants explicitly mentioned in their reflections and avoided inferring unspoken intent. Following the initial analysis, themes and sub-themes were discussed with the other researchers to check coverage.

In total, the transcript contained 185 relevant references that were categorised into 26 codes. We compiled the results into a taxonomy of barriers and facilitators based on key viewing aspects, broken down by causes and effects, as shown in Figure 4.6, which aligned with the survey viewing aspects to support triangulation.

4.3.4 Focus Group Results

We present the qualitative results from the focus groups, combining participants' accounts of barriers with comments from the SLTs and drawing together reflections from everyday viewing and the clip critique. The transcript contained 185 relevant references organised into 26 base codes. Thematic analysis identified four main themes, namely *Understanding Speech*, *Cognitive Load*, *On-Screen Text and Subtitles*, and *Following Narrative* (see Figure 4.6). We structure the results around these themes and indicate the causes and effects discussed within each.

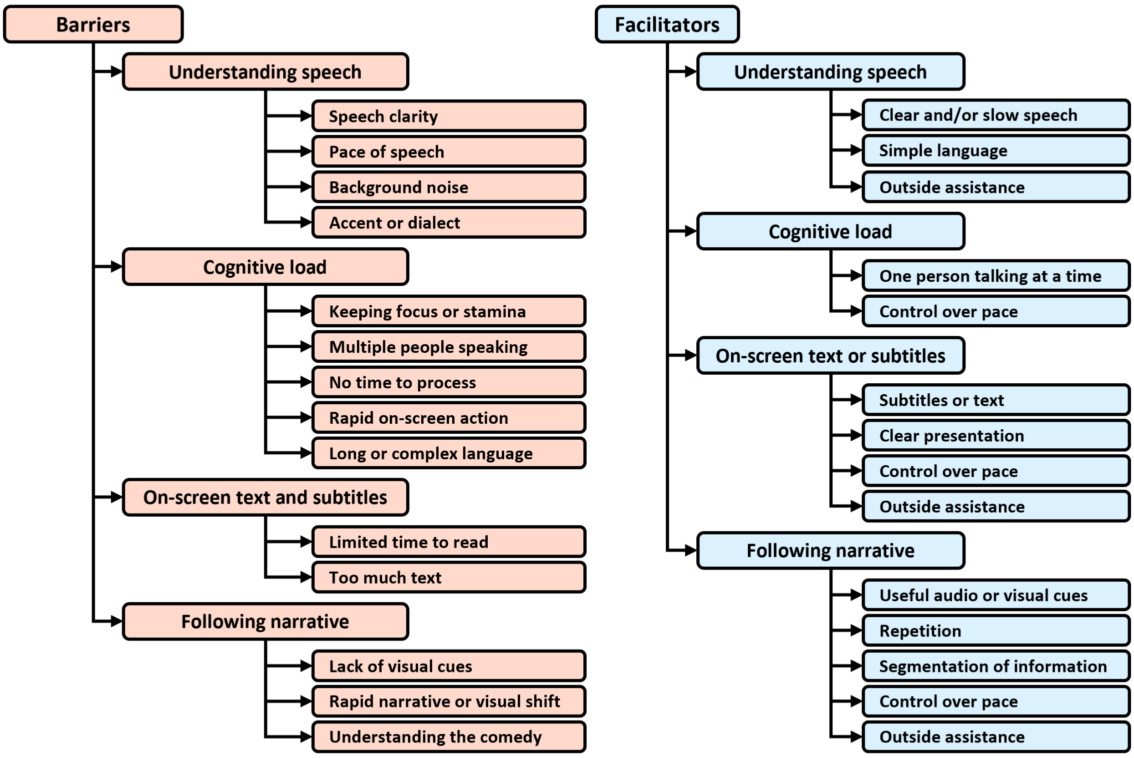

Figure 4.6: A taxonomy of the themes and sub-themes identified through the thematic analysis. These are divided into two sections: barriers and facilitators.

Understanding Speech

Participants most frequently discussed challenges with understanding speech, noting that characteristics of delivery could make spoken content less accessible for people with aphasia. Several participants reported difficulty when speakers did not use clear articulation. During the open discussion section, P9 described a recent viewing experience of a new TV show, stating "*I was watching the Silo uhh yeah and it's set in a futuristic environment and umm conspiracies are rife and umm that uhh you talk quite lowly umm whisper and I can't hear it*". Others conveyed similar concerns, with P6 proposing a volume increase to aid listening, saying "***P6**: You? [gesturing with a finger, starting from the table moving up, while making a noise that increased in pitch]*". P9 also mentioned using headphones as a personal strategy, though this was only feasible when watching alone, since social viewing introduced a further constraint, as captured in "***R2**: But then your wife wouldn't be able to hear? **P9**: [laughing] Yeah and even more importantly umm I can't hear my wife!*". Another participant found it hard to understand men speaking, explaining "***P4**: After my stroke, I could not* *understand men at all, because they not talking like that [gesturing with hand in front of her mouth, opening up from closed fist], they talking so so [mumbling, hand over mouth, head tilted downwards]*". Many participants likewise reported difficulty with strong accents or dialects, for example "***P9**: The Irish accent is... **P8**: Harder! **P9**: Hard, yeah... Yeah it is*".

Fast speech was another recurring barrier. Participants described losing words when the pace increased, as P1 put it "***P1**: Because it's too fast sometimes, so I don't... so I lose a lot of the words... okay?*". The speed of delivery also led to missed information in specific formats, such as quiz shows, for instance "***P1**: Yeah, there is quiz shows, the Chase? [...] They are so quick, I just glance at it and I thought phew [hand moves over her head]*" and "***P5**: It's quick, I didn't register the joke*". Participants noted that once they lost track of the spoken content it became difficult to continue watching and that keeping up was tiring.

Observations during the video viewing and critiquing activity reinforced these points. Some clips were easier to follow because of narration style and pacing. VC1 was cited as an accessible example, with the narration by David Attenborough (famous for his clear 'Received Pronunciation' English and slow speaking rate) described as clear and slow, which made it easier to follow, for instance "***P3**: I find that he's very good, how he talks, he's very umm proper, with every word he's saying*". The clear and simple language and pauses between sentences supported processing time. In contrast, VC11 was challenging because the host and participants spoke quickly within limited time, as expressed in "***P4**: I don't understand her [the host]. She's always going fast, isn't she?*". This meant many participants could not comfortably watch fast-paced quiz shows even if they enjoyed them, as in "***P5**: I like watching quiz programmes [...] It was uhh it was a challenge because it was so quick*".

Background audio was a further source of difficulty. Participants explained that distracting sounds and effects interfered with hearing speech clearly, for example "***P9**: Road noise, I was aware of the noise... Road noise... **P8**: Yes, yes. **R2**: Did it make it harder... what he was saying? **P9**: Yes it did, because you could hear the uhh rumble*". Loud background music also added barriers, as captured in "***P7**: No, no, not me but uhh for me, no because it is the uhh the music [...] yes, yes. Hard work though because* *music*". In clips that combined both distracting sound effects and loud music, some participants reported losing attention entirely, as in "***P8**: Noise, speaking, forget it [...] yes, the noise, pops, forget it*".

Cognitive Load

The second sub-theme concerned barriers with cognitive load and processing, a challenge explored in other contexts for people with CCNs, including people with aphasia during conversation [70] and people with dementia when playing games [295]. We use this framing to connect everyday communication demands with the multimodal demands of audiovisual media. Participants said that consuming audiovisual media was often tiring and that attention was difficult to sustain when content introduced certain barriers, with **P5** describing "*Every time I look at the umm video or watch the television, I have to work hard, much more hard than I used to*".

One recurring barrier involved scenes with multiple people speaking in a group setting. As **P4** put it, ***P4**: I'm understanding and then suddenly when there is more people on the TV then there is more people and... and I a bit lose it, I don't really understand it*". Clear delineation between speakers helped, with the group conversation in VC4 perceived as well structured. This was captured in ***P10**: Yes, yes, slow down fine [gesturing with his hand] **R2**: It was turn by turn, so it was... **P10**: Yes, good*". These remarks tie structure and pacing to reduced processing effort.

Temporal delimitation was important for individual speakers as well. Participants said they lost focus when speakers used long and complicated language without breaks between key points. Reflecting on VC11, **P1** noted *I think the- the questions are so wordy, and I lose... [shakes her head, waves her hand away]*". Similar concerns arose with continuous delivery without pauses, such as in VC2. The exchange read ***P10**: Voices, umm pausing or long sentences... **SLT1**: Did [the presenter] pause? **P8**: Oh no way! [...] **SLT1**: There was no break? **P8, P10**: [shake their heads]*". Additionally, long utterances without pauses increased processing demands and reduced recovery opportunities.

The cognitive load challenges reported by people with aphasia made it difficult to concentrate on continuous streams of information from different sources, often leading to confusion and loss of focus. Participants mentioned that fast, uninterrupted dialogue or narration was tiring to follow. Some said they would lose focus and stop paying attention entirely if they did not have enough time to process what was said before the next speaker started. As **P5** explained, ***P5****: I listen but then I have to comprehend... umm it takes a minute to comprehend, and then it's rushing on to the next one and it's [waves her hand away]*". **P2** described watching when no pauses were offered as *you can't carry on [waves her hand to go on] because they... you get... so behind*". These accounts link insufficient processing windows with loss of attention and increased fatigue.

Constant fast-paced on-screen action without visual breaks also introduced barriers. In VC9, **P9** commented "*The cars were moving too fast... umm I have problems umm concentrating on... the speed the cars are moving, yeah*" . Such rapid visual motion compounded auditory demands and taxed sustained attention.

Participants said that having control over pacing facilitated viewing. Pausing content when overwhelmed or rewinding to re-watch segments helped restore comprehension. Social viewing also supported recovery, allowing a co-viewer to fill in gaps. As **P4** noted, "***P4****: My husband has to stop and explain to me about this and this and this*". In general, participants repeatedly described pace as a constant barrier and stated that slowing down both dialogue and action would improve viewing substantially. These strategies point toward practical mitigation that aligns with the quantitative findings on pace-related difficulties.

On-Screen Text and Subtitles

On-screen text often introduced challenges, both when text was present within the shot (e.g. a book the actor was reading) and when text was superimposed on the shot (e.g. subtitles or a news ticker). A common barrier for both focus group participants and survey respondents was not having enough time to read the text before it disappeared, as in "***P2****: Yes but sometimes also the processing... you get it...*

and then you [moving hand as if reading from the screen]... three-quarters of the way and then changes, you know what I mean?". This time pressure coupled with divided attention made textual information difficult to use.

This frustration, together with the cognitive demand of attending to multiple information sources at once, meant that some participants did not attempt to read on-screen text. Throughout the focus group session, **P9** repeatedly said he did not try to read any of the text, both in the opening discussion, "*I cannot concentrate on the picture and the rolling text*", and during clip critique, "***R2**: You just still didn't bother [reading the text]* ***P9**: Uh... still concentrating on the speech [...]* ***SLT1**: So you're focusing on the auditory input* ***P9**: On the auditory input, yes*". Even when participants kept up with subtitles, they often missed other visual information, as **P6** illustrated, "***P6**: Umm... you [imitates reading the subtitles on the screen] Ah! [imitates looking up at the rest of the "image"] Ah, fucking hell! [looks around, then back down at the "subtitles"]* ***SLT1**: Yeah, you can't quite keep up with the text and the picture* ***P6**: Yeah [points at "subtitles", looks up at "image", looks surprised]*".

Large amounts of simultaneous on-screen text introduced a different barrier, where the volume of information itself was overwhelming. In VC2, textual information relevant to the news story appeared alongside the journalist, leading **P4** to comment, "*Well putting those words together... those words together I... I don't understand, see? I understand the words, I can say those words, but I can't put those words together [hand gesture and facial expression of confusion]*". High density text increased processing load even when individual words were familiar.

Participants described managing these issues by controlling pace, either pausing the content or rewinding to re-read. This was judged more helpful for on-screen text than for subtitles, since it gave ample time to process important information at a personal rate. As **P4** explained, "*So sometimes at the end of a movie they will put three pages of writing, and switching umm real films... you know what I mean, um True Love, so [her partner] has to stop, stop it, and then I read it [imitating reading with hand] and then I say go on, and then he goes on and I stop it again and read it little bit*". Such control was not always available, particularly in social viewing or on certain devices, as **P9** noted, "*In order to um for me to watch something with subtitles*

*I have to use my computer laptop and you can control the speed of the... **P10**: Fantastic! **SLT3**: Which you can't do on the TV? **P9**: No, no, no*".

Participants also remarked that presentation mattered. Clear segmentation and simple language improved comprehension. Textual cues, including subtitles, could serve as reinforcement when a viewer felt lost, even without reading every word, as in **P9**'s explanation, "*Sometimes umm it can be uh more accessible kind of in a reinforcement way, because you can scan the text and check to see if you are reading wrong **R2**: Ah, so if somebody is says something and you don't understand, then you look at the... **P9**: Yeah, yeah, yeah*". Similarly, having a second screen with additional textual information facilitated viewing, especially when combined with pace control, as **P2** said, "*But then I would go to the BBC news on my tablet, and then I can read it little bit by little bit by little bit*".

Following Narrative

The final sub-theme concerned barriers to following the narrative of audiovisual media. Participants described this as a frequent source of frustration that arose from multiple accessibility issues. They reported using additional visual and auditory information to piece together the storyline when they lost track. When such cues were absent, understanding what was happening became difficult, as reflected in "***P5**: You don't have the uhh actions and uhh visual stimulus, and uhh I couldn't follow it*". Visual cues could also appear too quickly to process, which limited their usefulness. Repetition helped restore understanding, as indicated by "***P8**: One time, okay... [rocks his hand side to side, indicating 'so-so'] Two times, fine [waves his hand down]*". Clearly segmented pieces of information likewise supported comprehension, giving time to process before the next unit of content. This was captured in this exchange: "***P7**: Yeah, it's uh short- short this uh [showing small distance with fingers] clip? **SLT1**: Short questions? **P7**: Yeah, but it's [holds up* 5 *fingers]* 5 *seconds and stop and* 5 *seconds... it's pretty good the whole lot, I can't **R1**: Like they had pauses, someone spoke then there was a pause... **P7**: Yeah, yeah, much better for me, the whole lot no I can't do it, but this... [nods]*". This points to the benefit of paced delivery with brief, discrete segments.

Sudden narrative or visual shifts also created barriers. Several participants said they struggled to maintain continuity when scenes or questions arrived without lead-in, as noted for VC11 in "***P1**: But, it's um out of the blue [gesturing, hand waving off], umm so umm the questions umm and then you answer it and I've [chuckles, makes a facial expression of confusion] yeah, I can't*". This difficulty increased when there were few cues to scaffold interpretation, for example, little redundancy across modalities or minimal recap.

Comedy presented a related challenge. Participants reported missing that a joke had occurred or failing to grasp humour because the effort needed to track dialogue left little capacity for interpretation, expressed in "***P1**: I don't have... don't have jokes, can't... follow, I don't understand it, if you have that problem, yeah? [...] Yeah, no just goes [gesturing her hand flying over her head]*". Audio or visual signals sometimes mitigated this problem. In VC3, a laugh track indicated that a humorous moment had occurred, as shown in the following exchange which illustrates how simple cues can restore narrative footing: "***SLT1**: What helped you to understand the joke? **P9**: The laughter **SLT1**: So the background laughter helped you? **P9**: Yes, yes*".

Complicated narrative structures added further difficulty. Temporal shifts and non-linear storytelling increased the cognitive effort needed to track events and relationships, summarised by "***P9**: Back in time you... some stories re-relay their story to... what happens yesterday or future, and you find it confusing*". This indicates that consistency in time and sequence supported comprehension, while jumps in chronology demanded additional processing.

4.4 Discussion

The following discussion examines how people with aphasia access audiovisual media, bringing together barriers and facilitators identified in our studies and explaining how the taxonomy can guide future research. The taxonomy is organised around points where comprehension breaks during viewing and around the conditions that help viewers recover. We use the facilitators to derive practical directions for improving accessibility for people with aphasia, with relevance to communities with complex communication needs, which is explored in later chapters.

4.4.1 Accessibility Barriers for Aphasia

Language understanding shaped the viewing experience. Audiovisual media relied on spoken content to convey information and meaning, which meant that people living with aphasia faced prohibitive challenges with some material. Viewers expended additional cognitive effort to parse speech and to coordinate multiple stimuli so that what was shown and said could be integrated into a coherent account. Difficulties with dialogue and narration interacted with other aspects of viewing, notably with keeping track of the narrative or the storyline. If a viewer missed what a character said, they could lose a vital piece of information and then struggle to recover comprehension. A single lapse could compound later difficulties, since effort spent catching up reduced capacity for ongoing processing. Participants frequently described losing place during fast segments and needing time to rebuild context from whatever cues were available. These interdependencies were consistent with the themes in the taxonomy in Figure 4.6 and they framed language comprehension as a gatekeeper for subsequent understanding.

Barriers were contingent on how content changed over time, with high pace generally reducing comprehension across the measured aspects. The four themes were driven in part by pace and short processing windows. Fast dialogue, rapid narrative shifts such as quick montages or high-intensity action, and limited time to read on-screen text increased the chance of cumulative difficulty. Recovery was harder when the viewer did not control the rate of interaction. This contrasted with prior usability work on social networking sites with people with aphasia, where the pace was user-controlled by design [253]. Focus group participants repeatedly reported better understanding when information was presented in short segments with pauses between units, as in the two documentary examples in **VC1** and **VC5**. Short segments helped orient attention and gave space to consolidate meaning before the next unit arrived. When pacing was not supportive, viewers appropriated pause and rewind to slow delivery and to create space for processing [173, 265]. Such effort demanded active management and was often frustrating, which reduced enjoyment and added cognitive load [52, 327]. Platform-level controls that expose simple, reversible pacing options would reduce the effort cost of coping strategies and make timing adjustments less disruptive to shared viewing.

Barriers in multimodal content interacted – problems in one modality affected others. Pace and timing influenced tracking speakers and understanding speech, while poor speech clarity or distracting audio such as loud background music led to missed visual or narrative cues. Viewers then failed to register humour or overlooked key on-screen text even when subtitles were present. These reports linked continuous listening demands with the need to scan visual elements in parallel, which multiplied processing effort. Some barriers described by respondents were consistent with communication difficulties in everyday contexts, such as noisy public settings [211] and group activities [128, 129]. Supplementary information helped when the primary stream was difficult. Contextual audio or visual cues created redundancy and supported comprehension even if parts of the message were missed, such as a laugh track after a joke. Similar supports appear in other aphasia contexts, such as text-to-speech for reading [154] and orientation aids for smartphone navigation [108]. Recommendations for preparing materials for people with aphasia reflect this need for multiple communication channels [120]. Presenting information in more than one modality increased the chance that at least one pathway remained accessible despite variation in symptoms and severity. Redundancy across these communication channels worked best when cues were paced to coincide with moments of potential information loss. For example, a brief visual signpost aligned with a change of speaker or an audio chime accompanying a dense on-screen element. Designers can combine short segments, simple language, and well-timed cues so that comprehension is supported without reducing content breadth.

These observations pointed to practical implications. People with aphasia benefited when the structure of the content moderated moment-to-moment demands on attention. Segmentation, clear cueing, and manageable pacing reduced the likelihood of cascading effects that began with a single missed element and propagated into broader loss of understanding. Where timing could not be changed in production, lightweight pacing controls at playback allowed viewers to create their own processing windows, though the effort required to operate these controls was itself a burden in the absence of supportive defaults. The cross-modal nature of the barriers suggested that interventions limited to a single modality were likely to offer partial relief. Combining speech clarity improvements with visual signposting and flexible on-screen text timing would offer a more reliable path for maintaining orientation during complex sequences. Taken together, these points suggest that accessible viewing depended less on the presence of any one feature and more on how timing, structure, and cueing were integrated into the overall experience.

4.4.2 Accessibility Facilitators for Aphasia

Facilitators included material intrinsic to the audiovisual media, such as clear and simple language in speech and pacing that left time to process. Facilitators also included support outside the media itself, including non-technology support from a friend or family member who could explain or recap during viewing. The taxonomy in Figure 4.6 grouped these elements and indicated how they improved understanding, including controls over pace and outside assistance. In practice, intrinsic facilitators and social support often worked together, for example clear narration paired with a co-viewer who paused and summarised.

Many facilitators were not always available or required active viewer effort. Using pause and rewind could help but made viewing less passive and added interaction demands [105]. This was pronounced when dialogue was challenging, with viewers sometimes replaying the same words multiple times in a pattern similar to re-reading text to improve comprehension [135]. Presenting information in different forms helped to alleviate these difficulties. Speech, text, and images provided alternative pathways, while clarity improvements such as slowing speech and using plain language reduced processing cost. Structural changes such as repetition and segmentation of ideas further supported understanding. Similarly, short segments, visible speaker cues, and brief recaps reduced the need for repeated manual control.

When barriers in dialogue or narration were present, other viewing aspects were affected, which is common for communities with complex communication needs [276, 277]. Existing technical interventions tended to target a single aspect, such as subtitles presented audio as text [105, 192], or audio description conveyed visual information through spoken words [178, 312]. These interventions were often treated as universal design exemplars [59, 215], yet they introduced challenges for some disabled viewers [45] and for the general population [158]. We have already seen this in detail in Chapter 2 and Chapter 3, seeing that prior work focused on a narrow range of interventions and a limited set of communities. Interventions tailored to barriers faced by people with aphasia were not widely explored, and widely used interventions were not always suitable for aphasia or other complex communication needs.

Aphasia and related communication impairments varied widely, so approaches needed to be variable rather than a 'one size fits all' model. Viewers benefited when they could combine facilitators, such as clearer speech, segmented structure, and simple pacing controls, instead of relying on a single feature. Designing for variation meant exposing lightweight adjustments that could be layered, allowing viewers to tune clarity, structure, and timing without heavy interaction overhead.

4.4.3 Practical Takeaways

Future research that aims to improve the viewing experience for people living with aphasia will need approaches that can adapt to varied access barriers, irrespective of how those barriers present. This need was reflected in how focus group participants adapted their viewing, combining multiple coping methods, interacting directly with the media, and drawing on outside assistance such as family members, often in combination. One direction is greater personalisation that lets viewers shape the content to their own access needs [132, 308, 315]. Prior work reported that people with complex communication needs, including people with autism spectrum disorder, expressed a desire for personalisation and customisation to bridge accessibility barriers [236]. Audio personalisation offers a concrete example. Ward and Shirley

[316], Ward et al. [317] showed that hard-of-hearing viewers could adjust the levels of audio tracks according to narrative importance, which helped isolate dialogue or salient effects and improved viewing. Similar techniques could address barriers that people with aphasia reported with background noise by tailoring audio playback to individual preferences. Personalisation should expose simple controls that are easy to reverse so that adjustments do not introduce additional cognitive effort during viewing.

These ideas can be supported by Object-Based Media (OBM), which represents audiovisual programmes as collections of assets and metadata, rendering a version at runtime that fits a given viewer and context – see Section 2.2 for more detail. While OBM has largely implemented existing interventions such as subtitles and audio description (e.g. [105, 130, 192]), the same principles could support new interventions tuned to aphasia. Examples include offering a choice of newsreader to improve clarity, slowing or staggering delivery to provide processing time, and exposing fine-grained timing controls that reduce cognitive load. Where production time permits, creators can attach semantic metadata that marks speaker turns, key plot beats, and dense text segments so that client players can surface supportive options without manual scrubbing.

Foundation models offer another route. Large language models (LLMs) can operate across modalities such as speech, text, images, and video. Potential applications include automated summaries at scene or episode level to support complex narratives and question answering about the content during playback. These functions echo prior practice in which SLTs provided summaries to support narrative comprehension for people with aphasia [51]. Any deployment should keep outputs concise and transparent, with controls that let viewers choose when and how summaries appear.

Development of such personalisable interventions should involve multiple stakeholders, as end users bring knowledge of everyday viewing and communication strategies. Content creators bring production constraints and editorial aims that will shape what is feasible [21, 181]. Interventions that affect content must avoid amplifying social biases. Participants' understanding of voices was often affected by accent. In our focus group sessions, 'Received Pronunciation' was perceived positively due to familiarity and due to cultural bias in the UK, where such accents are often ranked highly for prestige and 'pleasantness' [267]. If future systems allow voice selection, for instance choosing a different newsreader, designers must consider how defaults and options reflect and reproduce social norms, with the choices made being framed around clarity, rather than prestige to avoid encoding status judgments.

Given the variability of aphasia, participatory design can help create interventions that address real needs [266]. The range of barriers reported by our small sample suggested that different people would require different combinations of support. Design should also consider how technologies will actually be used. People living with disabilities vary in technology literacy [230] and physical abilities [89]. Many people with aphasia may struggle with specific inputs such as touchscreen keyboards because of stroke-related motor impairments including hemiplegia or paralysis [237]. Appropriate techniques are therefore needed in both design and evaluation. For aphasia, tangible design languages can make participation more accessible, and collaboration with SLTs can ensure that materials and procedures remain usable [320]. Interfaces that minimise text entry, provide large touch targets, and support voice or switch access will broaden applicability without requiring separate versions for each profile.

4.4.4 Limitations

Running an online survey with people living with aphasia presents specific challenges and limitations. Recruiting suitable participants is difficult as those active in online communities are typically more comfortable with reading, writing, and using technology, which favours people with milder aphasia. This means voices from people with more severe aphasia were likely under-represented. The overall sample size for the survey included 41 respondents and the focus groups included 10 participants. People with limited familiarity with social platforms were unlikely to be reached, and self-selection further narrows the sample toward people interested in media use or research participation. Taken together, these factors, alongside the modest sample sizes, constrain statistical power and breadth. Caregivers who assist may have influenced responses from those with more severe aphasia. However, this was not common in our respondents.

Designing an accessible online survey for aphasia requires careful choices about question type, presentation, and duration. We addressed this by using short thematic blocks, images, and skip logic to reduce cognitive load, yet completing the survey still demanded sustained attention. Readability may vary across devices, screen sizes, and display settings, which could affect legibility. Responses were self-reported rather than clinically assessed, and we did not classify participants by aphasia subtype, which limits subgroup comparisons. We did not offer a telephone or in-person completion route, so the online mode may have further filtered the sample. Open text fields can be demanding even with dictation, so richer qualitative accounts may be missing, which we addressed through the focus group sessions. Fatigue remains a concern in any language task delivered online, as some respondents may stop rather than pause, which can bias completion.

For the focus group sessions, the video viewing and critiquing activity allowed participants to reflect on concrete examples across formats. The clips were short, so participants did not need to sustain attention for extended periods. We therefore could not examine the impact of fatigue during long viewing, as well as limiting analysis of narrative continuity across longer narrative arcs. Playback controls during the activity were managed by the researchers, so we did not evaluate personal strategies such as spontaneous pausing or rewinding – we explore this in our future work (see Chapter 5, Chapter 6, and Chapter 7). Although reflective workshops captured some information about fatigue, longer video activities would provide a fuller account, with longer sessions which include planned pacing changes would allow measurement of recovery strategies and attention drift. Moreover, group dynamics and SLT support may have shaped turn-taking, which could have influenced what barriers surface. We sought variety in aphasic difficulties within the workshop sample, yet the overall sample was 'WEIRD' (Western, Educated, Industrialised, Rich, and Democratic) [172] compared with the world population. The focus group recruitment happened through a single UK charity, which may reflect local support practices. These constraints shape how the findings should be interpreted.

4.5 Conclusion

This chapter examined how people with aphasia experience digital audiovisual media, combining an online survey with $N = 41$ respondents and two focus groups with $N = 10$ participants to quantify effects and surface specific barriers and facilitators. The aim was to determine whether aphasia affects viewing and to identify where media presentation most often disrupts understanding. The results show consistent deterioration across viewing activities and provide a grounded account of how and why comprehension fails during viewing.

Quantitatively, respondents reported a clear drop in overall viewing experience with aphasia, alongside declines in understanding dialogue, following narrative, reading on-screen text, and tracking speakers. The largest decrease concerned on-screen text, reflecting time pressure and divided attention while scenes continue to unfold. Correlation analysis showed that language understanding relates strongly to overall viewing quality, whereas reading and writing scores are weaker predictors. These patterns place language processing, rather than sensory access alone, at the centre of everyday viewing for this group.

The qualitative findings explain how these declines arise, with understanding speech being the most frequent source of difficulty due to fast delivery, unclear articulation, strong accents, and competing background noise reducing the ability to keep pace with what was said. Scenes with several speakers intensified the problem because participants had little time to register who was talking and to consolidate meaning before the next turn. When clarity and pacing aligned, as in narration with simple phrasing and audible pauses, participants reported lower effort and fewer losses of place. Load increased further when information arrived in long, uninterrupted stretches or when visual action was rapid. Participants described losing focus in sequences with few natural breaks. Once a lapse occurred, recovery was costly because new material continued to arrive, producing a cumulative effect.

Short, clearly bounded units of information reduced this pressure, indicating that opportunities to pause mentally between ideas support comprehension even before any explicit control is applied.

On-screen text introduced distinct barriers, with many respondents having insufficient time to read, and some avoiding the on-screen text to preserve attention for speech and image. Others reported that dense blocks or multiple simultaneous visual elements raised effort even when the words themselves were familiar. Where reading helped, presentation choices mattered: segmentation, simple language, and stable placement allowed participants to confirm what they had heard without exhausting attention.

Narrative tracking weakened when speech was missed or when scenes shifted quickly without cues, with participants describing failing to notice jokes or losing track of actions in sequences with many characters. Signals helped to restore footing, such as a brief audio cue for a comic beat or a visible pause between speakers. Repetition also helped when new material arrived too quickly to integrate with what had just been shown or said.

Together, these themes describe the challenges of comprehension when viewing. Language processing, information pace, and the distribution of attention across modalities interact to determine whether viewers can follow with confidence. A single 'miss' can trigger a cascade of failure, in which later moments demand more effort while yielding less understanding. The survey and focus-group accounts converge on the idea that access is less about whether elements are present, and more about whether there is sufficient time to integrate them.

The chapter also shows that common access features do not always meet the needs of people with aphasia: subtitles can be too fast to read in context and may draw attention away from the picture; sound mixes that prioritise atmosphere over intelligibility obscure key narrative elements. By contrast, narration with simple phrasing and measured pacing reduces difficulty. The taxonomy developed here groups these issues by affected aspect and cause, identifying where comprehension most often fails in routine viewing.

Overall, the findings argue for treating access as the management of effort during viewing rather than the presence of standard features. This chapter therefore establishes the basis for subsequent work on supporting viewing at points where comprehension breaks. The next chapter develops candidate approaches grounded in the barriers and facilitators identified here, drawing on participants' reports of where time pressure and information density most often exceed what is manageable. The focus remains on understanding during viewing, and on how small changes to delivery can reduce effort at demanding moments.

Chapter 5

Envisioning

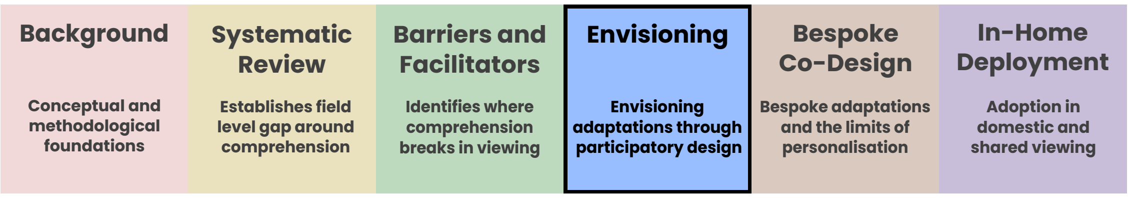

Figure 5.1: Map of thesis - envisioning

5.1 Introduction

Building on the work outlined in Chapter 4, in this chapter we set out to envision specific accessibility interventions for digital audiovisual media with people with aphasia. We ran two workshops following a diamond design process, using a cultural probe [99] in the form of postcards to support reflection between sessions. This work sits within our wider RtD methodology, using successive prototypes to test how proposed changes support comprehension during viewing in concrete examples. Through these workshops, we developed multiple mid-fidelity prototypes that represented the different intervention ideas, which were used in a critiquing activity.

This chapter aims to answer two key design questions. First, what form should accessibility interventions take. Here, we are looking at the ways in which the audiovisual media content can be manipulated to make it more accessible to people with aphasia, and builds on the facilitators identified in the previous chapter (see Figure 4.6). Second, what are the likely impacts of these interventions of the viewing experience. Here, we are also trying to understand how these interventions would be used in social viewing contexts.

5.2 Methodology

5.2.1 General Methodology

We structured the research as a sequence of activities that moved from idea generation to consolidation. This sequence comprised two co-design envisioning workshops and a postcard-based cultural probe completed at home. People with aphasia were involved throughout as end-users and as experts on their own accessibility requirements [21]. The flow in Figure 5.2 shows how these activities built on earlier analysis of barriers to audiovisual access. We adopted divergent and convergent thinking to support creative ideation [168] and to shape proposals into actionable intervention concepts [65]. Workshop 1 (WS1) used divergent thinking to elicit a broad set of intervention ideas, including unfeasible “magic” options [65, 74]. Discussion focused on changes to specific elements of audiovisual content that could remove barriers. A postcard probe then extended reflection beyond the workshop, with participants documenting everyday viewing situations and the interventions they wanted access to [99]. Workshop 2 (WS2) applied convergent thinking. Participants selected postcards they judged most relevant and discussed a narrower set of important barriers and more realistic interventions [65]. Ideas were refined towards concepts suitable for representation through mid-fidelity prototypes for later critique. We linked the WS2 discussion to the barrier and facilitator taxonomy from the previous chapter to maintain continuity of concepts across the thesis, see Chapter 4.

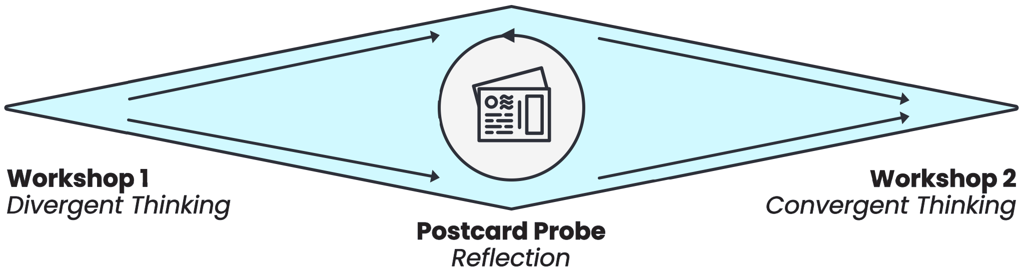

Figure 5.2: Representation of the two workshops and the postcard probe kit.

5.2.2 Participants

Recruitment was carried out through Dyscover, a charity in South-East England that supports people with aphasia through regular activities and groups. We had worked with this cohort in a prior study, and those individuals agreed to continue. Prospective participants received a plain-language information sheet before enrolment. They were invited to ask questions and to discuss participation with family or friends. Informed consent was obtained before the sessions began, with time set aside to confirm understanding and to address any concerns.

The workshops were held at Dyscover, four weeks apart, in line with recommendations to work in familiar settings [181]. Scheduling matched the time of existing weekly support sessions to minimise additional travel and effort. Each workshop lasted two and a half hours, including a planned break to manage fatigue. Activities took place in a room separate from other group work. Participants sat around a large table facing a projector screen, which supported shared viewing and discussion, see Figure 5.3. Tangible communication aids were available throughout.

Six people with aphasia took part, see Table 5.1 for their details. Four researchers attended both workshops. One researcher was a licensed SLT with experience supporting communication with people with aphasia and provided assistance across all activities. WS1 also included a Dyscover staff member as an additional assistant. Participant ages ranged from 56 to 71, with a mean of 60.8 and a standard deviation of 4.9. Time living with aphasia ranged from 4.5 to 16 years, with a mean of 8.7 and a standard deviation of 3.8. All participants had been fluent in English before stroke. None used augmentative and alternative communication during the sessions beyond the tangible aids we supplied, such as pen and paper. Each participant received a 40 GBP Amazon voucher to compensate for time and expertise across the two workshops.

Figure 5.3: Photo from the two workshops. Participants were seated around a table in front of a projector. The top right photo shows a participant holding a tangible prop created for the session; a wooden tablet representing the recap intervention on a 'second screen'. The bottom right photo shows a researcher reading a postcard.

5.2.3 Workshop 1 - Divergent Thinking

The first workshop had four parts. First, we did a briefing on the project, followed by a video-based critique and brainstorming activity, an ideation and critique session using a generative AI tool, and instructions for the postcard probe kits. The briefing, delivered with slides, set out the motivation for the research, the longer plan for the project, the categories of technologies we intended to explore, and the stages of future work. Presenting this overview at the start framed the subsequent activities and encouraged broad thinking about accessibility interventions rather than immediate solutions to single barriers.

The second part revisited barriers reported in Chapter 4 with the same participant group, including "*lack of pauses*", "*loud background noise*", and "*multiple*

Name	Gender	Age	Years w/ Aphasia
P1	Female	61	4.5
P2	Male	58	6
P3	Male	56	9
P4	Male	58	16
P5	Female	61	6.5
P6	Male	71	10

Table 5.1: List of participants in the workshops, along with demographic data. Researchers will be denoted as R1-4 in quotes.

people talking". Each barrier was paired with a short video clip prompt [225] and with quotes drawn from the previous sessions to anchor the discussion in familiar experiences. Clips were projected in front of the group and covered a range of broadcast formats (e.g. films, news broadcasts, and documentaries), with varied levels of audiovisual complexity [201]. After each clip, participants discussed what they would change to address the barrier. The discussion focused on specific elements such as dialogue pace, turn-taking, sound mix, and the availability of visual cues. To help with creative ideation, we introduced a cardboard magic button" prop, see Figure 5.4. Participants were invited to describe what should happen after pressing the button, which encouraged imaginative suggestions without requiring immediate technical justification. Tangible props can support envisioning with people with aphasia by providing a physical anchor during discussions [320]. We then had a 30-minute break to manage fatigue.

The third part asked participants to critique accessibility interventions generated by a large language model chatbot, '*ChatGPT-4o*' [1], and to review accompanying text-to-image visualisations. Using generative tools provided many ideas quickly from consistent prompts and reduced presenter bias in the examples shown. We requested intervention concepts that address barriers faced by people with aphasia when accessing audiovisual media. From the generated images we prepared simple representative prototypes, which allowed participants to react to concrete artefacts

[1] chatgpt.com

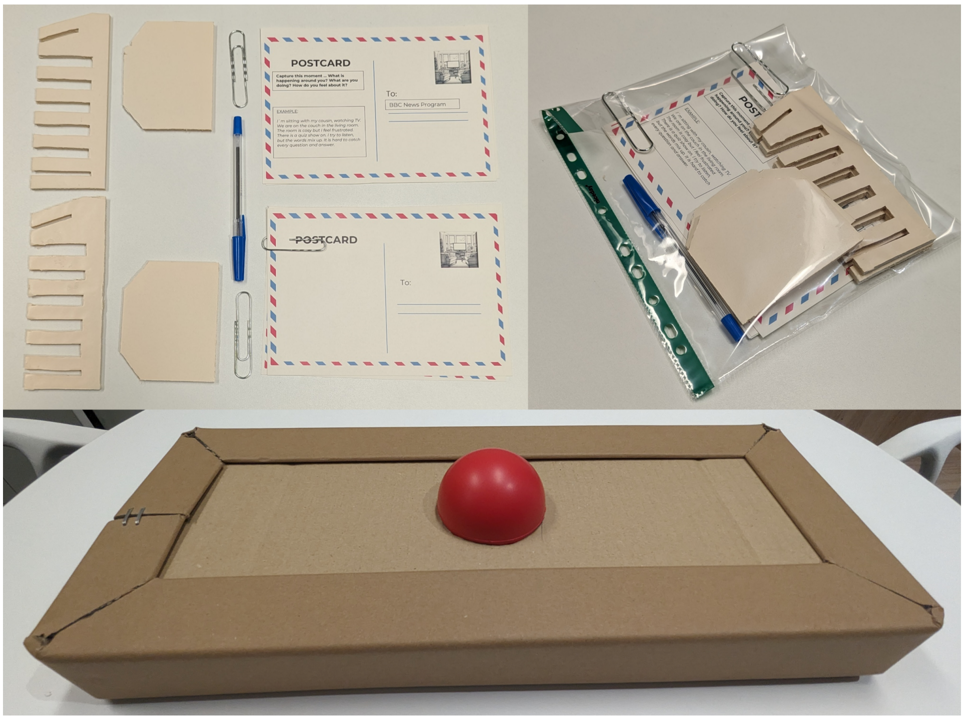

Figure 5.4: Photos of physical props we brought to WS1. The top two images represent the postcard probe kits, including the postcards themselves and their stand. The bottom image is the “magic button” used to prompt participants.

rather than abstract descriptions. Locating the idea source in an artefact shifted responsibility for the initial proposals away from the co-design team and supported open critique in a manner similar to persona-based evaluation [200]. To avoid priming the earlier brainstorming, the AI activity followed the divergent discussion. We introduced the tools with a short demonstration using a participant’s prompt about Liverpool Football Club to illustrate how ideas can be generated on demand. We then presented each AI-derived concept in turn. For each one we showed a brief text description of the intervention, the companion image, and the representative prototype. Participants commented on likely usefulness for their own viewing, raised concerns about elements that did not work for them, and suggested modifications that would better fit their needs.

The workshop closed with instructions for the postcard probe kits and distribution of materials for home use, see Figure 5.4. Participants were asked to record barriers they encountered in everyday viewing and to state what they wanted to happen differently to make viewing easier. The prompt invited examples such as fast dialogue, dense on-screen text, or distracting background noise, and asked for preferred changes such as slower delivery, added pauses, clearer speaker cues, or alternative audio mixes. Each postcard was addressed to a chosen entity to indicate where participants expected change to occur. Entities included producers, broadcasters, streaming platforms, and tool providers, reflecting whether the suggested intervention required adjustments during production or could be applied over existing content (e.g. subtitles). Each participant received one completed example and five blank postcards. They could seek help from family or friends to write or type their cards. The completed postcards were collected at the next workshop and used to guide convergent discussion of barriers and candidate interventions.

5.2.4 Postcard Probe Kits

The postcard probe kits were designed to capture barriers encountered during everyday audiovisual viewing and to document preferred ways of addressing them. Postcards invited participants to describe specific situations, propose an intervention, and indicate how that intervention should interact with the media. The kits also gathered details about viewing beyond the workshop setting, including device choice, household routines, and the presence or absence of others.

Participants placed the postcards on a stand near their main viewing area at home, such as in front of the television, so they were within easy reach when difficulties arose. This setup encouraged immediate note-taking at the moment a barrier was experienced, with cards available throughout the study period for repeated use.

Social viewing was a recurrent focus. The cards supported reflection on interactions with family or friends when barriers occurred, and on how viewing patterns differed between independent and shared sessions. Each postcard asked participants to address it to an entity responsible for change (e.g. producer, broadcaster, platform), making clear whether a proposed intervention required adjustments at production or could be applied over existing content. Keeping the kits available over time gave space to consider which issues mattered most, articulate desired changes, and return to the cards as new examples emerged. This enabled participants to curate their experiences by selecting instances they felt strongest about.

The format supported personal accounts. Participants could write short descriptions, add sketches, or combine both, which reduced reliance on extended prose and invited playful engagement. Overall, we received 15 completed postcards. Main topics concerned the social aspect of viewing (N=9), the pace of content (N=5), and a wish for understanding to return to pre-aphasia levels (N=4). Six postcards were completed with help from friends or family, such as assistance with writing or typing.

5.2.5 Workshop 2 - Convergent Thinking

The second workshop comprised two main activities. First, a structured presentation and discussion of the postcard probe kits. Second, a critique of mid-fidelity accessibility prototypes. For the postcard activity, each participant selected one or two cards they considered most relevant and presented them to the group. Where reading was difficult, a researcher read the card aloud so that the content could be discussed collectively. Several cards had been completed with help from family or friends, including typed versions or printed extras that were easier to handle and read – see Figure 5.5. After each reading, we asked targeted questions about the barrier itself, its impact on the viewing experience, and the social context in which it occurred. We also asked how participants would prefer to address the issue in practice, including whether the solution should change the content at production or operate on top of existing material. Once all cards had been discussed, we collected the full set for subsequent analysis and took a 30 minute break.

The second activity introduced a suite of mid-fidelity prototypes for critique – see Table 5.2 for summaries. These prototypes were derived from WS1 divergent thinking and mapped to barriers that participants had described. A recurring concern was pace in action, narrative, and dialogue. Fast delivery can trigger cascading difficulty, with attempts to keep up increase cognitive load and raise the chance of missing additional information. Missed content then requires effortful recovery, which further taxes attention. To explore pace-focused ideas, we prepared three interventions that explicitly adjust delivery. These were a playback speed control, automated pauses at speaker turns, and a step control for dialogue progression, see

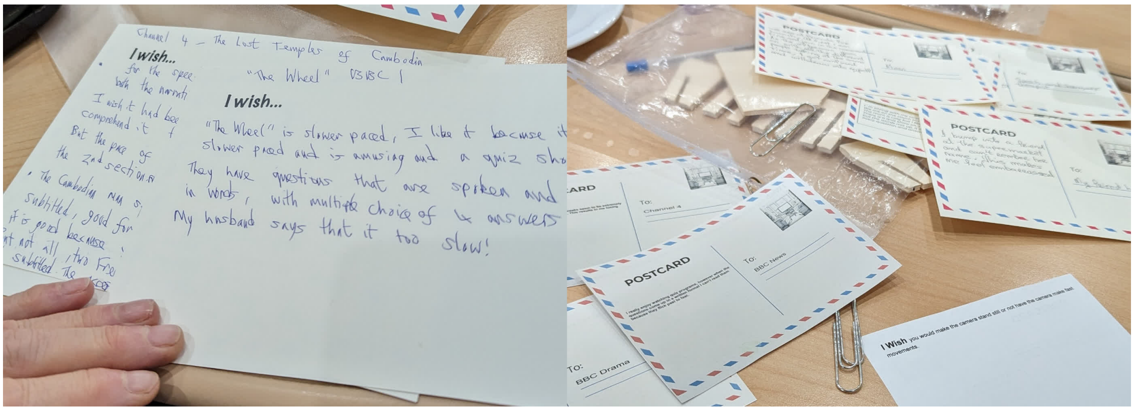

Figure 5.5: Examples of postcards created by the participants. One side of the postcards allowed participants to describe the barriers they experienced while expressing what they wished for to facilitate their viewing on the other side.

Figure 5.6 for examples of the control layout. Other prototypes addressed speech-in-noise, speaker identification, and language complexity by allowing background level adjustment, speaker highlighting, and alternative voice-over settings.

Each intervention was implemented as a page in a multi-page web application built with Next.js [2]. Pages shared a consistent layout to reduce interaction overhead, with each page having a central video player with control elements below. Some systems relied on manually generated metadata, such as timestamps for speaker changes, pauses, or salient plot beats. Others were presented as edited videos that behaved as working artefacts in the session, for example the speaker highlighting prototype. We projected the application for the group so that demonstrations and discussion could proceed in a shared view, and we introduced each intervention with a brief description of what it changed and how participants could operate it.

Feedback was organised around three prompts. First, would the prototype facilitate viewing in practice, and in which situations. Second, would participants use it in social contexts, for example when watching with family, and would its operation disrupt shared viewing. Third, how would they want to control the intervention. We probed device choice, level of control, and preferred effort. Participants considered a spectrum that ranged from continuous, moment-to-moment interaction to a set-and-watch configuration. We also asked whether triggers should be manual or automated

[2] https://nextjs.org/

Name	Description of the intervention
Slow Down	Video clip from the film "The Social Network" in which two characters speak in a busy bar. The intervention allows the viewer to control the playback rate of the video through a slider control.
Step Control	Video clip from the film "The Social Network" in which two characters speak in a busy bar. Intervention allows for speaker-to-speaker dialogue step control, in which the viewer controls the pace of the dialogue in a video by pressing the "Next" button to continue to the next piece of dialogue. The viewer can also press the "RePlay" button to re-watch the previous piece of dialogue.
Automated Pauses	Video clip from the film "The Social Network" in which two characters speak in a busy bar. The intervention automatically inserts pauses into the dialogue after each speaker, with the duration of the pause controlled by the viewer.
Background Noise	Video clip of a BBC News broadcast in which the journalist speaks over busy road noise. The intervention allows the viewer to control the volume level of the speaker and the background noise.
Highlight	Scene from the film "The Social Network" in which two characters speak in a busy bar. The intervention highlights the on-screen character that is currently speaking.
Simplified	Video clip from the film "Richard III" in which the character gives a monologue in 'Shakespearean' English. The intervention allows the viewer to toggle between the original dialogue and a simplified version read out by text-to-speech.
Accent	Video clip from the TV series "Limmy's Show" in which two characters with strong Scottish accents speak to each other. The intervention allows the viewer to change the voice-over of a video from the strong original accent to Received Pronunciation read out by text-to-speech.
Recap	Scene from the TV series "Doctor Who" showing the end of a scene. The intervention gives the viewer the ability to receive a scene recap when they pause the video.

Table 5.2: List of all WS2 mid-fidelity prototypes. The interventions consisted of hard-coded interactive elements or edited videos. These were presented through a video player with interactive components allowing control over video elements or toggling on the intervention.

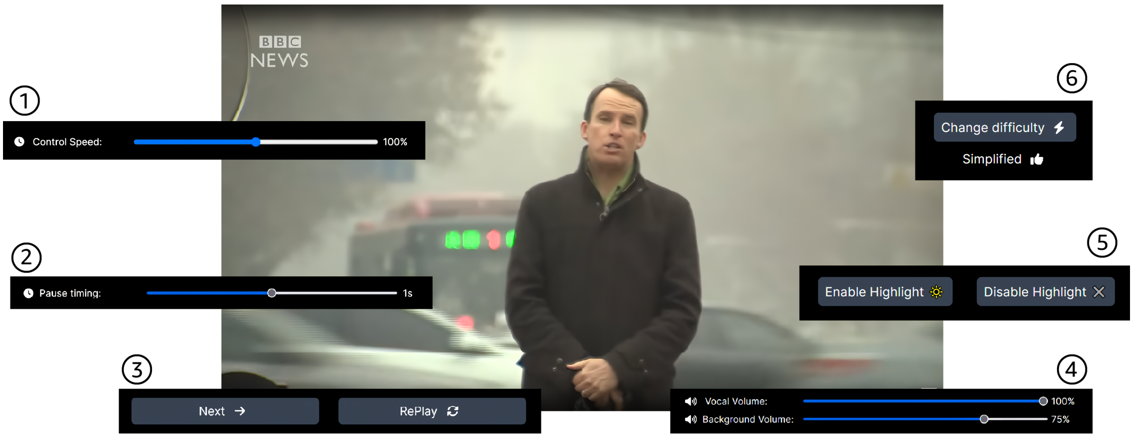

Figure 5.6: Examples of the mid-fidelity accessibility intervention prototypes created by the researchers. These interventions allow the participants to interact with elements of the video, such as: (1) the video playback speed, (2) the duration of automated pauses in dialogue, (3) dialogue step controls, (4) speaker and background volume levels, (5) enabling speaker highlighting, and (6) changing voice-over difficulty. The video still comes from the BBC News YouTube channel – link to video.

and what level of transparency was required when content elements were altered during playback. These prompts ensured that comments covered feasibility, social fit, and control strategy, linking WS2 discussion back to the barriers and preferences surfaced by the postcards and WS1 brainstorming.

5.2.6 Data Analysis

We video and audio recorded both workshops. Participants chose how their image could appear in outputs. Options were not shown, shown with faces blurred, or fully visible. Two researchers transcribed the recordings in NVivo 14. Transcripts captured verbal and non-verbal communication because many people with aphasia rely on alternatives to speech; for example, **P3** had limited verbal abilities and used non-verbal cues and tangible communication aids.

The first author conducted an inductive thematic analysis of the transcripts. We used reflexive thematic analysis [37], coding what participants said, noting any extra meaning they suggested, and refining codes step by step. Coding focused on accessibility barriers, preferred ways to facilitate viewing, and aspects of the viewing experience including social contexts. We did not compute inter-rater reliability, consistent with reflexive TA where researcher subjectivity is treated as a resource rather than a threat to validity [38].

Postcards were then transcribed and analysed using the same approach, aligning codes with those used for the workshop material to support comparison. We used methodological triangulation across the transcripts and postcards to check consistency of barriers and facilitators and to refine theme boundaries. Theme review and naming followed multiple passes through the dataset, checking candidate themes against coded extracts and against the full set of materials. Exemplar quotes and brief descriptions of non-verbal communication were retained to evidence each theme. Following the initial analysis of transcripts and postcards, themes and sub-themes were reviewed with the other authors to check coverage and refine definitions. We also wrote short positionality statements reflecting the analysts' backgrounds and prior work with aphasia, as recommended for reflexive TA reporting – see Section 1.4.2.

5.3 Results

We present results from thematic analysis of the workshop transcripts and the postcard probe, combining reflections from activities and discussions with accounts from everyday viewing. Across these materials we identified three themes: *bespoke adaptations* in interventions, the *social fabric of audiovisual media*, and *translating audiovisual content for accessibility*.

5.3.1 Bespoke Adaptations

Participants asked for personalisable interventions that give them direct control over content, so they can adjust or remove elements that act as access barriers during viewing.

Navigating Audiovisual Pace Variability

Pacing across speech, narrative progression, and visual action was a central focus of intervention ideas. Participants wanted ways to moderate timing when sequences felt too fast to process or when dialogue moved on before they could consolidate meaning. In WS1, this led to concrete proposals for control at the point of delivery, with **P1** suggesting to "*put pauses between each uh sentence or pair of sentences*", and later adding that she would "*love to be able to control the pauses*". These requests positioned pause as a practical tool for creating brief processing windows rather than lengthy stops. In WS2, the appeal of pausing was clear, with several participants already using available controls whenever possible: "***P5:** You have a remote that can do this, so I would [...] pause, let me get my head around this, okay, carry on*". Pausing and rewinding were described as effective recovery strategies that let viewers rebuild context and, where needed, obtain quick clarification from a co-viewer before continuing: "***P5:** The pause button, I said oh quickly explain for me, and the she will... he will explain it and then is... we carry on*". These practices emphasised short, targeted interruptions that restore orientation without derailing the session.

Prototypes that addressed pace in WS2 – see 'Step Control', 'Automated Pauses' and 'Slow Down' in Table 5.2 – were well received. Even the simplest option, slowing the playback rate, drew positive reactions: "***P3:** [thumbs up, smirk on his face] **P5:** I mean it's much better than it was*". Participants noted that pacing tools are most helpful when they fit natural viewing rhythms. Some preferred to watch longer continuous segments before intervening, rather than introducing frequent small stops: "***P2:** Longer sentences um longer sentences um pause,* 20 *[seconds], pause*".

Participants also cautioned that automatic or frequent pauses can disrupt the flow of conversation, especially in fast exchanges where utterances build on one another: "*Because they're talking so fast, it... she says something, and he says something, you can't actually split it, because what she says, or- and she- he feeds on to what he says and it- it flows like that*". This tension between needed processing time and maintaining dialogue continuity shaped preferences for control style and frequency. Several participants favoured manual control over pauses or a simple global slow-down, rather than automation that might pause at moments they did not perceive as difficult. Overall, participants sought timing controls that created brief processing space, preserved conversational momentum, and matched their preferred cadence for interaction.

Reducing Audiovisual Complexity

Participants also considered interventions that alter or remove elements of the audiovisual stream that act as access barriers, prioritising direct interaction with the content for greater personalisation. A straightforward case came from WS1, where **P6** explained difficulty with long end-cards in documentaries, stating that he "*can't track the words properly*" and would prefer if the information were shown as "*bullet points I think... I am better at reading bullet points*". In the same workshop, while watching a clip of a journalist presenting a news story, **P2** focused on visual clutter and background motion, noting that the "*picture [of the journalist] is bad for me, because [points at journalist] is there, but [imitating the cars behind journalist] background [waves hand] [...] delete them all, the background*". Reducing competing visuals made it easier to attend to speech.

Clearer visual cues also supported speaker tracking. With the speaker highlighting intervention enabled, participants reported less confusion about who was talking: "***P6**: Whereas before, I thought the two people talking at the back of the room... confusing the... um protagonists*". **P5** reinforced this point, adding "*you know exactly who is speaking, you concentrate on him, you block out some of the noise* ". We further examined focus on the speaker with a prototype that exposed background audio controls. Lowering ambient sound made the primary voice easier to follow: "***P1**: It's just such a simple thing to do and it just makes so much difference, it's uh when you um decrease the um slowly it will uh I could... the benefit of the uh doing that*". Across these examples, participants tended to prefer simplifying existing elements (e.g. bullet-point summaries, reduced background motion or audio) rather than adding new overlays or dense text.

Localisation Through Temporal Adaptation

The choice of which intervention to use, and when, was difficult. Participants wanted increased content localisation – adapting existing content to meet their own needs – with the adaptation varying by context. Scene recaps from WS2 were seen as useful when confusion arose: "***P2**: Complicated **P5**: Yeah, I agree, it depends **P6**: Only if it's complicated, yeah*". **P5** suggested combining mid-fidelity interventions with rewinding, such as highlighting the speaker or changing the voice-over, so she could first watch the version intended by the artist and then a more accessible pass for herself: "*To play two minutes of this [video], and then maybe go back again, and play it again, so you get the interaction between [the characters], but then you can start again and just recap*". Deciding when to enable other interventions was harder, especially those that alter the content, like replacing a strong accent with an alternative voice-over: "***P6**: Hard to locate when or if you might turn it on because you don't want to watch the uh rhythm of the play too much*". We also probed views on automated enabling based on the content and momentary complexity. Participants were cautious about automation, expressing concern that automatic changes could spoil the experience or act at the wrong time: "***P6**: How can we be certain that it always works for us*". Overall, preferences favoured lightweight manual control or simple global adjustments, with automation acceptable only if it was transparent, reversible, and easy to override.

Frictionless Audiovisual Viewing

The interventions envisioned by participants involved a high level of personalisation that required interaction to adjust them to individual needs, which raised questions about the experience of viewing while using such tools. Such interventions can introduce friction by subverting expectations and by changing aspects of the content during playback. While discussing the 'automated pauses' intervention, **P1** noted that unexpected stops could feel jarring even if helpful: "***P1**: I uh I think on balance it is useful but um I uh had a uh the moment you showed it to us uh I thought um the stops um were um unnatural*". This issue is pronounced for 'time-consuming' interventions that take time away from the content, such as pausing dialogue or providing a synopsis of events.

One suggested refinement was to give viewers manual control or to increase the interval between automatic activations, with this duration being customisable. Participants also warned that time-consuming interventions can interrupt concentration and cause loss of place: "***R2****: If you have long pauses between turns, you might even forget what was the happening* ***P5****: Yes, but that's what... you have a remote that can do this*". These points indicate that automatic enabling may hinder the viewing experience. Intervention use would also depend on context, including content type, interest, and whether others are present, with **P5** in WS2 describing a willingness to continue without support despite barriers: "***P5****: So I would- this [content] I would let go* ***R2****: So you also making some sort of compromise* ***P5****: Yes, yes*".

5.3.2 Social Fabric of Audiovisual Media

Audiovisual viewing often occurs with others. In our discussions, preferences varied between participants and neurotypical co-viewers such as family and friends. We describe two patterns. One is brief, seamless support from close family that maintains the shared experience. The other is situations where usual routines and preferences take precedence over accessibility.

Seamless Integration of Social Support

Our findings indicate that audiovisual viewing is highly interdependent. While close family members offer support, it often involves minor adjustments to our participants' needs. This support most commonly takes the form of explaining missed information: ***P5****: Both of us have a remote, and I would stop it and say to him, please explain what- what they said, and he would explain, and then we would carry on again*". Support can also reinforce each other's understanding and improve the shared experience: ***P6****: And you can operate from one another* ***P5****: Yeah, and he looks, and he says, you don't know what's happening, and I say yes*". Such mutual support can make some interventions less useful in social contexts, for example synopsis or recap features becoming redundant: ***R1**: So, for example, when you're watching with your husband, and there is a scene that is complicated, instead of zoning out and picking up your iPad and looking at pictures, you could look at [the recap intervention] and... **P5**: Yes, but normally he would read it out to me **R1**: So, he would do this [points at the wooden tablet] for you? **P5**: Yes, but if it was on the screen, he would read it for me*". Other forms of support include accessing language elements that would otherwise be inaccessible, notably reading on-screen text. This is not always possible, such as when there is a large amount of text and insufficient time to read it: ***P6**: When I asked [my wife] to read them out loud, her reading slowed down, as is everyone's case, and she was unable to read fast enough because you read more with your mind than you read aloud*".

Challenging Act of Balancing Neurotypical Tendencies

The interdependent nature of viewing involved continuous negotiation about what to watch and how to watch. Preferences differed within households, captured in **P1**'s postcard about pace in quiz shows: *My husband says [the quiz show] is too slow, but I like it because it's slow*". Negotiation also covered whether to intervene during playback or pause to seek clarification, as **P1** reflected: ***R2**: Do you sometimes ask your husband like, hey can you explain this? **P1**: I didn't in that case, but I uh I do uh normally*". Time of day influenced engagement and stamina for social viewing, with **P5** noting: "*In the morning, I am very alert, by the time my husband comes back from-from the office and he wants to tell me about his day, and we have dinner, and then we want to watch a movie, I am a bit slow by then*". This negotiation often prioritised neurotypical viewing patterns, which could exclude the person with aphasia if barriers arose. **P5** described disengaging when shared viewing became challenging: ***P5**: Yeah, if [husband] and I are watching, then I do something else, but if it's me on my own I change the channel*". Disengagement sometimes meant leaving the shared activity entirely: ***P5**: Okay, then I will pick up my iPad [...] and [husband] finishes the- the movie*". Reliance on others for understanding could pass on gaps, as **P6** put it: "*you miss what they miss*". Social dynamics could add barriers, such as interruptions that broke concentration: "***P6**: Um yeah uh because uh interruptions at the key moments, [my wife] has a penchant... for talking through big classic lines, you know?*". These moments affected continuity even when support was well intentioned. Participants weighed whether to use accessibility interventions during shared viewing. **P5** indicated she would avoid intervention use to keep the session smooth with her husband, but would use it when watching alone: ***P5**: Yes, but with [husband], I would let it play [...] but in my- on my own, yes I would use the [intervention]*". Maintaining togetherness often implied accommodating the preferences of neurotypical co-viewers, including acceptance of adjustments like reducing background noise: ***R1**: How do you think they would feel if you were [...] removing the background noise **P2**: Fine, fine. **P5**: No, [husband] would be fine*". Balancing access needs with household preferences led to frustration at times and to reflections on lost capabilities. Postcards captured these feelings, including **P5**'s wishes that *all the numbers made sense to me*" and that she *could remember things like [she] used to*".

5.3.3 Translating Audiovisual Content for Accessibility

Participants discussed interventions that personalise content in two ways: by altering the content itself to meet the viewer's needs, or by providing support that runs in parallel to the content. Both approaches change how the media is consumed and interact with the intended vision and message.

Making Accessible Audiovisual Content Culturally and Linguistically Appropriate

Examples of content-altering interventions included changing camera work and voice characteristics to improve intelligibility. In a postcard, **P5** asked for a steadier camera, writing "*I wish you could make the camera stand still*" rather than "*going around and around*" and "*making fast movements*", because she was "*concentrated on the- the- the camera going around... I can't... I can't... I can't listen to them speak*". This request framed motion as an access barrier that could be mitigated by alternative shots or reduced movement, so speech and action remained easier to follow.

Altering the voice was a second route to accessibility. Participants discussed replacing strong regional accents with alternatives that they found easier to understand, with **P6** preferring "*BBC English, because it's easier to understand*". We explored this idea in WS2 with a prototype that changed a strong Scottish accent to a voice closer to received pronunciation (RP). Reactions were mixed. **P1** said "*it makes it easier to understand*", while **P6** cautioned that such changes could "*lose part of the film's character*". These divergent views emphasised that clarity gains had to be balanced against changes to the perceived identity of the material.

Decisions about whether, and how much, to alter content depended on context. Participants weighed content type, their interest in the material, and whether they were viewing with others. Cultural factors also shaped preferences, with David Attenborough cited repeatedly as a presenter whose accent was, in **P6**'s words, *easier to understand*". By contrast, some 'Northern' UK accents or foreign accents were described as harder to parse, and their voices could be replaced to improve access. The same prototype that replaced the Scottish accent of comedian Limmy highlighted this variation. **P1**, unfamiliar with the show, reacted positively and stated that "*it makes it easier to understand*", whereas **P6** could not imagine Limmy not being Scottish and viewed the change as "*destroying*" the creators' intention. When asked what producers might think, **P6** added they "*would think you are destroying it*".

These discussions sit alongside everyday viewing practices that already diverge from how content was originally intended to be experienced. Participants described constructing workable experiences to meet their own needs and preferences: using second screens while watching [252], repeatedly rewinding and re-watching scenes, asking others for support, or engaging in parallel activities. Such practices alter timing, attention, and interpretation, even without changing the underlying media assets. Content-altering interventions therefore extended familiar strategies, bringing adjustments into the playback itself rather than leaving them solely to viewer effort and social support. Many viewers with aphasia benefit from adapted camera work, clearer speech, or accent substitution, yet some regard these changes as compromising artistic vision or character. Participants navigated this tension pragmatically. They supported options that were easy to toggle, reversible, and scoped to specific moments, so that clarity could be improved when needed without a permanent shift in style. They also preferred interventions that simplified existing elements over those that added dense overlays. Across these accounts, content alteration was acceptable when it preserved narrative coherence, remained sensitive to creator intent, and respected the viewer's control over when and how changes occurred.

Semiotic Audiovisual Adaption

The balance between access and preserving the original meaning shaped how participants talked about personalised adjustments. They drew a clear line between support that leaves the media intact and changes that rewrite parts of it. Non-altering support focused on timely summaries that help regain orientation without touching the soundtrack or visuals. Participants expressed interest in a recap that appears on pause and can be listened to rather than read: ***P6**: I would like the facility whereby it's read to me*". They also stressed selective use. A recap should be available when needed rather than persistent: ***P5**: Maybe one scene you want it and another scene you don't want it, so you move on*". These comments frame recaps as on-demand scaffolding that maintains the existing cut, mix, and style while offering a brief route back to the thread of the story.

Views shifted when interventions proposed to alter the content to improve access. Participants were open to targeted modifications in some contexts but wary of broad or permanent changes. Objections were strongest where they judged the work to be artistically significant. When shown a mid-fidelity prototype that simplified Shakespearean language, the group pushed back: "***P3**: No... [thumbs down, makes a 'bad' sound effect] No [shakes his head] **P2**: No, no... [it loses the] ambience*". The same people suggested that similar transformations would be acceptable for didactic or ephemeral material such as online courses, documentaries, or live news, where clarity may matter more than stylistic fidelity. This content dependence placed editorial intent and genre at the centre of decisions about where adaptation fits.

Priority between access and authorial intention was contested but not uniform. Asked directly which should come first, **P1** prioritised being able to watch in practice: ***P1**: Yes but otherwise people wouldn't watch the film and uh... if it helps people to um watch the film*". Participants also noted that alternative versions can widen accommodation across disabilities, as **P2** put it: ***P2**: No because um speech, language, others... other ones*". These remarks indicate that some viewers would accept alterations when the trade-off yields usable access, and that benefit may extend beyond aphasia.

Mechanism and control were central to acceptability. Several participants supported automated adaptation in principle, including AI-driven adjustments, but only with clear signalling that a change has occurred and with safeguards for personal data. Hidden modification and opaque data use raised concerns: ***P6**: I would um feel uncertain of that fact, you know, who is that going back to*". Others preferred to keep adaptation entirely under their control, rejecting automation even when it might be accurate on average. **P5** stated a firm preference for manual operation: ***P5**: I could not take from that thing [automated content pacing] **R3**: So, you would turn it off? **P5**: Yes, all of it*".

Domesticating Audiovisual Content

Participants discussed altering content in terms of translating the original into an aphasia-friendly version that retains intent while removing barriers. They wanted translation to "*stipulate aphasia friendly*", as **P6** put it, and to be designed into production rather than added post-hoc. **P5** argued that such work should start "*right at the beginning when they shoot it*", so that alternative materials exist alongside the main cut rather than being retrofitted under playback constraints.

Voice alternatives were a frequent example. For accents that some viewers found hard to parse or for dense language, participants proposed recording more than one reading at production so that an intelligibility-focused track could be selected when needed. **P1** suggested that "*you could have two versions... and uh it's all about diversity*", framing multiple takes as parallel products rather than as an afterthought.

This approach fits content domestication, described by Chaume [55] as influencing creation from an early stage so the programme can be assembled to meet specific needs. Participants applied the same reasoning to camera work. Rapid motion and complex blocking were described as exclusionary when attention had to be spent tracking movement rather than meaning. In postcards and discussion, **P5** called for steadier footage at production to support comprehension: *The camera must stand still, it doesn't matter what way we... the front, the back, the everything, but it just stands still*".

Translation raised limits. Some changes that increased access could also reduce enjoyment if they altered the feel of performance or timing. Even minor overlays were cited as affecting delivery, with **P6** remarking that subtitles can reduce conversational punch: "*But, you lose something by listening to subtitles because it the punchiness of the uh conversation, you know? Loses it's punch*". Participants therefore framed production-time alternatives as preferable to heavy on-screen additions during playback, since alternatives can preserve rhythm and tone while improving intelligibility. Participants also favoured options that were reversible at any moment, with clear indication of which version was active.

Agency in creation and selection was important. Participants wanted people with aphasia to be involved in shaping alternative assets and in deciding how those assets should be surfaced to viewers. **P5** offered to participate directly: "*And if you want to go with the... to BBC, we will all go with you*". In practical terms this meant providing accessible scripts to actors for second takes, capturing steadier shot sequences in fast-paced scenes, and attaching metadata that marks where an alternative would help. It also meant building players that allow a viewer to choose the original or the translated version for a scene, a character, or a segment without losing continuity with co-viewers.

5.4 Discussion

5.4.1 The Promise of Personalisation

Personalisation in digital environments, including internet-distributed television, enables changes in consumption that can produce accessible viewing for people with aphasia. It refers to the extent to which content reflects a viewer's distinctiveness across interests, history, and relationships [220]. In accessibility contexts it must also reflect lived experience with disability, as many access needs arise from everyday viewing situations rather than from static profiles.

Through personalisation, viewers select options that meet individual requirements. This selection can yield new content [289] and can position the viewer as a *source* in the interaction [145]. Workshop discussions treated the viewer's selection and timing of adjustments as part of the experience rather than as a separate configuration task. Interfaces should minimise linguistic demand so that people with aphasia can operate personalisation without extended text entry.

Personalisation differs from user-initiated customisation. In customisation the viewer adapts existing elements without changing the underlying content or its message. Many personalisation systems place control with external agents that use personal data to adapt content to data-derived requirements, while customisation describes services controlled by the user that rely on adapting the content's data [18, 88]. Systems that use personal data should provide opt-in consent, clear provenance, storage details, and simple override to meet privacy expectations.

Adapting content to a viewer's needs can produce accessible versions of the content [132, 308, 315]. Changes can address pace, intelligibility, and visual clarity so that processing effort remains manageable during continuous playback. Personalisation can operate at different levels of granularity, from global adjustments that affect an entire programme to momentary changes that apply to a scene or a speaker. It can leave the original assets intact and surface alternatives as selectable tracks or metadata-driven variants that fit viewing rhythms. Where content alteration affects perceived character or intent, reversible alternatives should be available at all times.

These distinctions set out the terrain for accessible personalisation of audiovisual media. Personalisation introduces agency through selection, timing, and scope of change, while customisation remains limited to adapting existing elements. For people with aphasia, the practical benefit lies in targeted adjustments that fit shared routines and do not require continuous verbal input. Personalisation that respects privacy and preserves optionality can produce usable access without imposing a single mode of viewing.

System Controlled Adaptive Personalisation

System-controlled adaptive personalisation is becoming more feasible with advances in AI and other machine learning approaches. Tools that generate text and images from prompts now extend to text-to-video systems [176]. These capabilities can support runtime adjustment of audiovisual elements. However, audiovisual viewing is complex. Our participants reported varied practices that depend on setting, co-viewers, preferences, and distractions. Localisation affects what counts as television and how it is experienced [298]. This variability matters when automating adjustments for people with aphasia.

Black-box interventions could adapt content to access needs. A system might remove distracting audio in dialogue-heavy scenes or slow delivery when processing time is limited. Such automation raises questions about transparency, control, and failure modes. Changes to the soundtrack or edit must be visible to the viewer. Systems should show clear on-screen indicators when a modification is active and provide a one-click revert to the original.

Encountering a barrier can lead to disengagement. If an automated intervention behaves unexpectedly, the viewer needs to regain control and adjust settings. Interaction patterns that resemble continuous prompting of generative tools are demanding. They require linguistic input, sustained attention, and rapid decision making. These demands run against the preference for seamless, low-friction support expressed by our participants. Interfaces should minimise text entry, rely on large, simple controls, and offer presets that match typical viewing rhythms.

Automated personalisation must fit household practices, with viewers needing reliable behaviour during shared sessions and consistent operation across devices. Indeed, default behaviour should fail safe to the unmodified track when confidence is low. Scene specific granularity is preferable to global switches so that changes apply only where they are needed. Additionally, personal data used for adaptation should be opt-in, with visible provenance and simple override. When automation cannot meet these expectations, manual control that is lightweight and reversible remains necessary.

Flexible Media

Flexible media offers an alternative to automated, system-controlled adaptation by treating programmes as assemblies of assets whose selection is deferred until playback, with metadata guiding how elements are combined to produce a personalisable experience [17]. As outlined in Section 2.2, object-based media provides the technical substrate for this approach; here we focus on how runtime assembly supports accessibility without restating production details. By manipulating constituent audiovisual elements such as parallel audio and video tracks, the player can assemble versions that meet a viewer's needs at runtime, yielding accessible variants without re-authoring the source [130]. Established pipelines already deliver interventions such as subtitles [105] and audio description [192], and support personally adjustable audio that lets listeners rebalance dialogue against background elements [316, 317]. In practice, this requires metadata that marks speaker-turn boundaries, dialogue salience, competing sound objects, and shot stability, so that a player can expose intelligibility-preserving mixes and steadier visual selections.

Extending this approach enables interventions that reshape presentation rather than only layering parallel tracks, which includes interactive narrative configurations [302] and controls over pacing and delivery applied at scene level [63]. For people with complex communication needs such as aphasia, these non-linear, perceptive controls change how information is delivered so barriers can be addressed within the flow of viewing. Participants in our workshops described abandoning content when camera motion or dense sound made comprehension difficult, whereas exposing steadier alternative shots or substituting a clearer voice for a given character allows viewers to pre-select supportive options or toggle them at the point difficulty arises. Crucially, the original assets remain available, so personalisation can be scoped to a speaker, a sequence, or a setting while preserving the unmodified version for co-viewers. Interfaces that present these choices should avoid heavy text entry, offering presets and reversible toggles aligned to typical viewing rhythms, and platform profiles should apply consistent preferences across devices while retaining a single-control return to the original mix.

Bespoke Co-Design

Co-designing highly personalised media interventions in group settings proved limited in our workshops because participants brought variable demands that shifted with content, device, and social context. Bespoke co-design addresses this by applying personalisation within the design process itself, working one-to-one with participants to examine their viewing routines, identify the moments where barriers occur, and prioritise changes that make the largest practical difference. This approach is suited to people with aphasia, whose profiles vary in language abilities and may include motor or cognitive sequelae after stroke. It also extends beyond the individual to the wider viewing context, since significant others influence what is watched, when interventions are operated, and whether adaptations remain acceptable in shared sessions. Indeed, sessions should include co-viewers where possible and take place in familiar environments such as the home so that preferences and constraints reflect everyday practice.

Bespoke co-design raises feasibility questions, notably how to generalise beyond single cases without flattening individual requirements. One route is to iterate interventions with multiple participants who present related barriers, comparing where settings converge and where they diverge, and documenting the situational factors that drive those differences. Rather than fixating on a single optimal configuration, the design can expose parameter ranges and presets that map to recurring situations such as dialogue-heavy scenes, dense on-screen text, or busy sound mixes. Reports should specify the trigger conditions, control granularity, and effort required to operate each intervention, and share artefacts that include timings, speaker turns, and mix priorities so others can replicate and refine them. Where automation is proposed, record how confidence thresholds and override behaviour affect trust and continued use.

New audiovisual forms (e.g. live streaming and short-form social media) have accelerated hyper-personalisation and the rise of personal media economies [111], enabling creators and platforms to tailor experiences to nuanced preferences and interactive behaviours. These shifts can support accessible consumption when controls are scoped to scenes, voices, or mixes and remain reversible, yet they can also exclude viewers if defaults privilege speed, novelty, or dense interaction over comprehension. Platform audits should examine recommendation and player defaults for their accessibility impact, and co-produce guidance with viewers, including those who have aphasia, so that personalisation strategies remain usable in shared contexts.

5.4.2 Redefining Audiovisual Access

Maintaining Social Viewing Experiences

Audiovisual media are changing as devices become more interconnected and applications multiply, introducing shifting methods of interfacing with content [148]. These shifts unsettle familiar heuristics of use and require viewers to adjust repeatedly, which adds interaction overhead and increases cognitive demand. For people living with aphasia, continual adaptation to new controls, layouts, and timing patterns compounds existing language-related difficulties, as the effort spent navigating interfaces competes with the effort needed to follow speech, action, and on-screen text.

Parallel viewing offers a strategy that absorbs this variability by giving viewers temporal agency [54]. Rather than treating playback as a single, fixed timeline, parallel viewing recognises that viewers may pause, rewind, segment, or consult companion materials on a second device, stretching the time available to process dialogue and narrative while maintaining participation in the main session. Chambers [54] argues that such tangential browsing does not fragment social viewing, and instead it enables co-present engagement in a shared space, with each participant managing attention according to need. In our workshops, participants described using short pauses, brief rewinds, and quick checks on a second screen to regain orientation without abandoning the session, which preserved togetherness even when barriers arose.

When content presents challenges for a viewer with aphasia, second screens can provide concise prompts, names, or short summaries that restore context without interrupting the shared flow. This arrangement keeps the primary display clear while supplying just-in-time support on a companion device, reducing the need for extended verbal explanation and limiting disruption to co-viewers. Adapting the shared space to fit these practices matters for accessibility and inclusion. Households can agree simple rules for when pausing is acceptable, assign light-touch controls to whichever person finds them easier to operate, and keep supportive materials close enough to consult quickly. Researchers exploring the domestic viewing environment can draw on 'relaxed' shared media experiences designed for autistic audiences [14] to inform configuration of pace, sensory load, and interaction, ensuring that adjustments remain practical in everyday settings and that viewers with aphasia are not excluded from shared sessions.

Enhancing Participatory Content Creation

Ensuring participatory practices in audiovisual media production helps address the idiosyncratic needs of people living with aphasia, reflected in participants' calls for domesticated content that can be selected when needed rather than imposed at playback. Using a second screen and engaging in parallel activities can draw attention away from the experience originally intended by creators, and personalisation during viewing can alter elements that shape meaning and style. To preserve creative choices while supporting access, production can prepare alternatives in advance, with decisions about shots, voices, and timing made alongside the principal cut – this aligns with flexible media because artists can supply parallel assets, such as steadier camera takes or alternative voice-overs, that remain selectable without rewriting the work. This translation may not have formal or dynamic equivalence to the original, yet it produces a companion rendition that augments the source and can improve comprehension [55]. For instance, participants reported difficulty parsing certain accents and losing narrative threads; offering a clearer voice option changes the presentation but can restore intelligibility while retaining the original track as a selectable choice for co-viewers. Producers can also attach metadata for speaker turns, salient plot points, and mix priorities so players expose alternatives at appropriate moments.

Accessible alternatives that preserve artistic or semantic equivalence can engage people with aphasia in media they would otherwise avoid, with creators participating in preparing versions that fit different viewing contexts. These practices broaden ways of consuming and interacting with audiovisual products and connect with the "audience's turn" described by Chaume [57], which frames audiences and creators as joint actors in assembling the experience. Participants' discussions of aphasia-friendly content show a preference for reversible options and maintained continuity, indicating interest in active involvement as co-creators rather than passive recipients. By taking part in creation, individuals with aphasia can shape accessible assets, specify where alternatives are appropriate, and determine how they should surface during playback, expanding their role in defining inclusive media. Co-creation activities can include trial recordings of alternative readings, review of camera movement in dialogue scenes, and agreement on labelling conventions so that viewers can identify versions quickly during shared sessions.

Cross-Disciplinary Engagement

Research on supporting audiovisual media viewing with people living with aphasia should move beyond purely technical solutions and involve stakeholders across clinical, design, and production contexts. Speech and language therapy offers established practices for enabling everyday access [182, 209] that can inform interventions targeted at barriers encountered during viewing, including how information is structured, paced, and presented. When developing support that operates in parallel to the content, such as second-screen applications with varying levels of detail [201], these aids can be aligned with SLT approaches to presenting relevant information concisely and at appropriate moments, reducing language demand while maintaining orientation. Communication scholarship is also pertinent for translation and high-level personalisation, as it addresses how meaning can be preserved when delivery changes and how alternative renditions can be prepared to fit different contexts.

Supported viewing has social implications that extend beyond a single user. Interventions that offer concise prompts or summaries can reduce reliance on verbal explanation in shared sessions and can benefit other communities with complex communication needs, for example viewers who wish to lower distracting sensory stimuli. Future work should therefore engage a wider range of stakeholders, including SLTs, people with aphasia and co-viewers, accessibility and communication researchers, platform and player developers, and production partners for whom commercial viability and generalisability are central, while examining cultural and political responses that shape adoption [104]. Studies should report how stakeholder input changes requirements, specify workflow and metadata needed for production-time alternatives, and evaluate social acceptability and operational effort in domestic settings.

5.4.3 Limitations

Envisioning abstract interventions with people who have complex communication needs can be demanding because participation in such activities requires sustained cognitive effort [320]. To support engagement, we presented concrete examples through tangible props and mid-fidelity technical prototypes, which grounded discussion and allowed participants to explore intervention ideas in a more immediate form. The choice of video clips may have influenced opinions of specific interventions. For instance, the highlighting prototype was demonstrated in a dark scene, which could favour visibility and may not reflect performance in brightly lit material. Similarly, background-noise controls were shown with a news clip featuring prominent traffic sound, which may elicit different reactions than quieter genres.

All demonstrations used short video clips, which enabled coverage of varied content but limited observation of how interventions behave over longer sessions, including fatigue, attention shifts, and cumulative effects on comprehension. The postcard probe was intended to extend reflection into everyday viewing, yet probe data captures high-level accounts and relies on post hoc recall, which constrains detail about timing and moment-to-moment interaction.

Our participant sample was small and drew from similar backgrounds, approximating a WEIRD cohort [172] in a global context, similar to the in-person study in Chapter 4 (see Section 4.4.4) which was conducted at the same charity. The group did, however, vary in language abilities and communication strategies, which provided a range of perspectives within aphasia.

5.5 Conclusion

In this chapter we worked with people with aphasia in envisioning ways to make digital audiovisual media more accessible. Two co-design workshops and a postcard probe were used to surface barriers, facilitators, and think of candidate directions for interventions. The work centred participants' accounts of where comprehension breaks and how small, targeted adjustments might keep material within reach without turning viewing into continuous effort. The activities produced themes that link practical breakdowns in comprehension with preferences for how support should appear in the moment, while also documenting tensions between comprehension-first interventions and routine viewing.

Participants consistently asked for highly personalised adjustments that respond to the specific moment rather than a single, fixed solution. Pace was a recurring barrier, and people described relief when they could create brief processing windows by pausing, re-watching short spans, or easing the overall rate, yet they also noted that frequent or poorly placed interruptions can fracture continuity. Pacing support was therefore valued when it preserved the line of thought, with control used sparingly at points where comprehension breaks rather than becoming an extra task layered onto viewing.

Alongside pace, participants focused on reducing audiovisual complexity in concrete, situated ways. They wanted help to hold onto the current speaker and to hear what mattered when background elements competed with dialogue. They also asked for text that conveys essential information without requiring sustained reading while images change, preferring segmented and stable presentation over dense blocks. Across these examples, the benefit lay in lowering momentary load so that attention could follow events, not in adding more elements to the screen. At the same time, participants recognised that changes to mix or layout affect tone, so relief from effort must be balanced against what gives a programme its character.

Support was also framed as situational, such as through short cues that recap what has just happened, but rarely desired as a constant layer – the same intervention that restores orientation in one scene can distract in another. When asked about automatic enabling based on content characteristics, several participants were cautious about systems that act on their behalf without clear signalling or easy override. Usefulness depended on what was being watched, when difficulty appeared, and how much disruption the viewer was willing to tolerate at that point in the programme. Since participants' viewing often occurs in shared contexts, the sessions showed how support is negotiated with others. Family members often supplied quick explanations that allowed viewing to continue, making some envisioned features redundant in shared contexts. In other cases, household routines led participants to let challenging material run or to disengage rather than ask for changes, especially when stopping would impose on others. Accessibility is therefore embedded in the social fabric of viewing, where decisions about when to intervene are shaped by habits, roles, and tacit norms as much as by individual need.

Discussion of content alteration exposed further tensions, with some adaptations were seen as legitimate translations that preserve meaning under constraint, such as steadier camera work in demanding sequences or a more intelligible reading of a line. Others, including replacing a distinctive accent or simplifying stylised language, raised concerns about losing authorship, ambience, or a sense of place. Judgements varied by genre and attachment to the material: adjustments acceptable for live news or instructional content were questioned for comedy or Shakespeare. Participants also noted that everyday practices already reshape timing and emphasis through pausing, re-watching, and second-screen use, making the boundary between following and transforming it porous in real-life viewing.

These perspectives suggest a role for alternative versions prepared alongside the main cut, allowing viewers to select material that is easier to process without discarding identity. Ideas included parallel readings for difficult voices and calmer shots for fast-moving sequences, with control left to the viewer over when to draw on them. While some participants were open to automation, they wanted transparency about what was changing and why, and expressed unease about hidden modification or opaque use of personal data. Others preferred to keep decisions in their own hands. Across positions, the shared requirement was that any translation remain recognisable as the same work, not a different one. The combination of idea generation, reflection in place, and convergent critique created conditions in which preferences, trade-offs, and limits could be expressed plainly.

Taken together, the findings treat access as the management of effort during viewing rather than the presence of specific accessibility features. Participants asked for ways to ease spikes in demand, reduce competition between audiovisual elements, and regain footing when comprehension slips, without turning programmes into a series of stops and starts. They also situated change within social routines and within a view of authorship that accepts alternative versions in some settings and not in others. The themes of bespoke adjustment, social negotiation, and careful translation provide a working organising framework of where interventions are likely to matter and what forms of change viewers judge acceptable in practice. This organising framework anchors subsequent work that narrows scope onto concrete expressions of these ideas. The emphasis remains on understanding during viewing and on adjustments that can be applied when pressure is highest, with attention to how preferences vary by person.

Chapter 6

Bespoke Co-Design

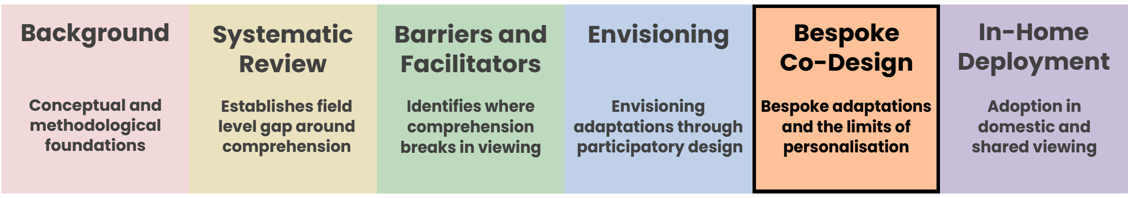

Figure 6.1: Map of thesis - bespoke co-design

6.1 Introduction

Building on the intervention concepts and mid-fidelity prototypes developed in Chapter 5, this chapter examines bespoke accessibility interventions for people with aphasia through targeted co-design. We focus on personalising delivery during playback, extending prior work on manipulating audiovisual elements. This framing focuses on comprehension during viewing, with intervention value assessed through whether viewers with aphasia can follow what unfolds with manageable effort. Earlier engagements indicated that barriers, preferred adjustments and acceptable trade-offs vary between people with aphasia, so we used one-to-one co-design to capture each participant's priorities. We therefore worked individually with four participants to envision their ideal audiovisual media accessibility interventions. We developed mid-fidelity prototypes of these interventions, and iterated them between envisioning workshops. We then evaluated the interventions practical use with eleven participants at a different aphasia charity, exploring the interventions use in shared social experiences, their effect on creative intent, their interaction complexity, and their feasibility for content producers.

The chapter addresses two design questions. First, how should personalisation operate for an individual's access needs during continuous playback. Here we examine controls over pace, the mix between speech and background sound, navigation through complex dialogue, plus subtitle alternatives, linked to previously reported barriers. Second, we assess implications for domestic and social viewing, creator intent, as well as production and runtime assembly. The remainder of the chapter sets out the co-design method, the interventions, the iteration, and the evaluation, and reports findings on usefulness, effort, and acceptability across different contexts.

6.2 Co-Designing Bespoke Accessibility Interventions

In this section, we discuss the approach we took in envisioning these bespoke interventions and the steps taken to evaluate them. We first discuss the overall methodology, introduce our four co-designers and the four interventions they envisioned. To ground the bespoke nature of these interventions in the co-designers lived experiences, we present representative vignettes. Following this, we describe the development process, as well as the evaluation and data analysis.

6.2.1 Methodology

To examine how audiovisual personalisation could facilitate viewing for people with aphasia, we advanced personalisation to a bespoke level, developing individual interventions through participatory co-design with each person. In the workshops, participants articulated preferred ways to make viewing easier and co-designed simple technical artefacts embodying those ideas. We used research through design (RtD) to investigate futures of personalisation, focusing on what such experiences could and should be [330, 332]. Through the design, creation, and annotation of these interventions, we examined how personalised delivery changes the viewing paradigm for people with aphasia [100]. As described in Chapter 1 and Chapter 2, we defined personalisation here as adapting the presentation and control of existing audiovisual elements to individual needs during continuous playback. We report a study that applies bespoke audiovisual personalisation with people with aphasia, using mid-fidelity prototypes to ground discussion and refine concepts.

6.2.2 Participants

We recruited four participants through Dyscover, a charity that provides group activities and support for people with aphasia. We had collaborated with this cohort previously at the same organisation, so the participants were familiar with the research team and routines, which reduced introductory friction. Earlier conversations with the group covered barriers encountered when viewing audiovisual media and typical viewing contexts, and these inputs shaped the workshop plan.

Participants received a plain language information sheet and had opportunities to ask questions. Informed consent was obtained in advance of the sessions. The study received approval from the King's College London Ethics Committee. Sessions were scheduled to coincide with the participants' weekly support meetings at the Dyscover centre to minimise travel effort and keep the environment familiar. Activities were run in a room adjacent to the main group area to limit distractions and allow focused discussion.

We conducted two workshops spaced six weeks apart. The interval provided time to translate individual ideas into mid-fidelity prototypes and to prepare materials for iteration between sessions. Two researchers who had worked with the group before facilitated both workshops and maintained continuity in communication practices.

Participant characteristics are reported in Table 6.1. All participants were fluent in English before their stroke, their mean age was 64 (SD=4.9), and they had been living with aphasia an average of 10 years (SD=4.4). All four had right-side hemiplegia with limited use of the right arm, which informed practical arrangements during activities, with seating and device placement being arranged to favour left-hand use.

None of the participants used augmentative and alternative communication during the workshops. Each participant received a 20 GBP Amazon voucher per workshop as compensation for time and expertise. This arrangement was communicated in advance and applied consistently across sessions. The combination of familiar setting, established rapport, and a defined schedule supported engagement with bespoke co-design and the development of the prototype interventions.

Name	Gender	Age	Years w/ Aphasia
Jane	Female	62	5.5
Simon	Male	59	17
Sally	Female	62	7.5
Bart	Male	72	11

Table 6.1: List of participants in the workshops, along with demographic data. Participant's names have been changed.

6.2.3 Individual Intervention Ideation

The first workshop (WS1) focused on co-designing individual interventions grounded in the accessibility barriers to audiovisual media reported earlier in Chapter 4. Sessions were run one-to-one (see Figure 6.2) to centre the individual, giving each participant sustained time to articulate preferences without interruption, and to keep communication simple and free of distractions. This arrangement was particularly helpful for participants with more severe expressive aphasia, where expressing complex ideas can require extended time and gradual clarification.

Discussion opened with the participant reviewing the barriers they encounter during everyday viewing and identifying those that most affect their experience. Once a barrier was selected, the participant was invited to propose any changes they would want to see, using pen and paper to sketch or list ideas, without constraining thinking to immediate feasibility or available technology. Researchers prompted for concrete examples (e.g. moments of fast dialogue, dense on-screen text, or heavy background sound) so that proposed adjustments remained tied to practical situations.

Figure 6.2: Images from the envisioning workshops, including an example of the storyboard, a participant envisioning, and a demonstration of the mid-fidelity prototype.

Following this ideation, the set of suggestions was synthesised into a single technical intervention concept that captured the participant's priorities. We then asked the participant to storyboard how the intervention should operate during playback, using a prepared template that stepped through setup, activation, and typical use. Where drawing was difficult or undesired, the participant directed a researcher to complete the storyboard while verifying each element. The one-to-one activity lasted about half an hour per participant, yielding a clear record of targeted barriers, proposed adjustments, and preferred modes of control that informed prototype development for the subsequent workshop.

6.2.4 Understanding Bespoke Interventions Through Vignettes

We use four vignettes to situate how participants with aphasia co-designed bespoke interventions, describing everyday viewing, concrete barriers, and the contexts in which adjustments are desired. Each vignette draws on multiple sources of information, following guidelines and prior work that employs them [31, 67, 83, 127]. These include the storyboarding activity conducted in the workshops, earlier engagements with the same participants across research activities that began in June 2023, and materials produced during those prior sessions.

These sources provide detail about factors that facilitate or impede viewing in practice, participants' views on the role of original creative intent in shaping acceptable adaptations, and how social viewing with family or friends affects whether and when an intervention should be used. The qualitative corpus comprises 15 hours of video-recorded discussion, storyboard artefacts, and cultural probes in the form of postcards that document everyday barriers and preferred changes, including who or what entity a participant expects to act on those changes. We analysed the material inductively. Codes were generated from observed examples of difficulty during viewing, short accounts of relevant life events, and direct quotations that express preferences and reasoning. The aim was to trace barriers to specific audiovisual elements or interaction moments and to capture how participants framed workable adjustments without imposing technical assumptions.

Vignettes offer a practical way to organise large qualitative datasets while preserving the distinctiveness of individual contributors [31]. They present cohesive narratives that remain grounded in empirical materials and retain the participant's voice. The viewing experiences of people with aphasia are under-represented in accessibility research, particularly when investigating intervention design. The vignettes therefore provide concrete, situated accounts around which interventions can be specified, prototyped, and critiqued, supporting design decisions that reflect lived experience rather than abstract requirements. Participant identifiers used in the vignettes refer to the workshop cohort and do not disclose personal information.

Vignette 1 – Jane

It is late in the afternoon when Jane comes back home from the shop, where she bumped into a friend whose name she had forgotten. As she starts putting away the groceries, her husband arrives home after work, and the two recount their day. They both sit down in front of the TV to watch the BBC News broadcast and unwind. As the newsreader rattles off the day's main events, Jane has to focus on each word and quickly finds the exercise tiring. The news broadcast then opens with the first main news story, a new government proposal. Jane listens attentively but finds it difficult to follow the news story, getting distracted by the news ticker describing a different

Jane

Simon

Sally

Bart

Figure 6.3: Illustrations of the four bespoke co-designers.

event and the sound of cars behind the on-field reporter. After a short while, Jane picks up her remote, pauses the broadcast and turns to her husband. *"I can't watch it because I'm looking at the text... can you explain to me what they are saying now?"*. He briefly explains how the government wants to put in place some sort of restrictions on cars, but does not add more detail and uses his remote to continue watching. Jane leans back on the couch and tries to continue watching, but she has lost interest and is feeling tired. She picks up her iPad and checks her Facebook instead. She wishes that she could remove the distractions, both in the video and audio, so that she could focus on what's important.

Vignette 2 – Simon

Simon puts on a comedy skit show on Netflix as he assembles a piece of furniture, glimpsing at the TV between each repetitive task. Usually, the skits involve two comedians, but occasionally other characters are also involved. The episode, however, sees a guest comedy troupe in each skit, with many of the jokes overlapping with others. Simon finds this difficult to follow, so he puts down the screwdriver and uses the remote to go back in time. The simple 10-second rewind, usually sufficient, does not help much, and he has to continuously press the button. Due to his hemiplegia, Simon cannot use the remote control while holding the screwdriver, so progress on the furniture assembly is slow. He turns off the TV and concentrates on the assembly in silence. Simon wishes he could have simply gone back to the moment before he got confused so that he could press the button once and keep going – even better if he did not have to press a button.

Vignette 3 – Sally

Sally talks to a friend who recommends a new quiz show on TV. A fan of quiz shows and the challenges they bring, she decides to tune in later in the evening. Unlike most shows, this one is slower-paced, something that she appreciates as it gives her an opportunity to answer the questions. The host is taking his time to read each question slowly and with great clarity, a little too slowly for Sally's husband. One of the participants, however, speaks quickly and with poor enunciation, making it difficult to comprehend what she is saying. After answering a question on Angkor Wat, a Buddhist temple in Cambodia, this participant starts talking about its history, a topic of great interest for Sally as she plans to visit later that year. Sally concentrates on each word but struggles to understand what she is saying. She wishes that she could slow down the speaker and hear each word spoken one by one, with a clear voice, like the presenter.

Vignette 4 – Bart

Bart and his wife sit down to watch an episode of a new TV series. The episode follows various characters preparing a secretive scheme, with many scenes involving the characters speaking in hushed voices, which Bart finds difficult to make out, so he turns on the subtitles. This does not help much, since the language is getting quite complicated and he cannot keep up before the next set of subtitles appears on the screen. At this point, he pauses the episode and asks his wife if she can read the captions whenever the dialogue is indiscernible, which she starts doing. They continue watching, but his wife can only read out loud so fast, and she quickly finds herself lagging behind and missing the end of the dialogue. They finish the episode, at which point he turns to his wife and asks her to explain what had just happened, which she happily does. Bart wishes he could have changed the subtitles to his liking – he doesn't need them to convey the plotter's dialogue verbatim and would prefer them to support his understanding by conveying the gist in a simple manner. He also wishes that he had more time to read the subtitle, without having to constantly stop and start the playback, something that annoys his wife.

6.2.5 Developing the Interventions

All bespoke interventions were implemented by the first author as a browser-based application that combined a video player with a separate controller. The controller could be operated from a second device such as a tablet or phone (see Figure 6.5). The application was built with Next.js [1]. Each intervention page paired a small library of clips with controls that enabled the feature and adjusted its parameters, keeping operation consistent across prototypes.

The clip set comprised four short items: a *BBC News* segment (02m15s), a round of the quiz show *The Chase* (02m52s), a dialogue scene from the TV series *Industry* (02m11s), and a scene from the film *Devil Wears Prada* (01m50s). These clips were chosen to span broadcast formats and genres and to cover different levels of

[1] https://nextjs.org/

audiovisual complexity [201]. Short durations made it possible to present a complete narrative unit within a single demonstration, which supported comprehension during evaluation and aided envisioning activities [171].

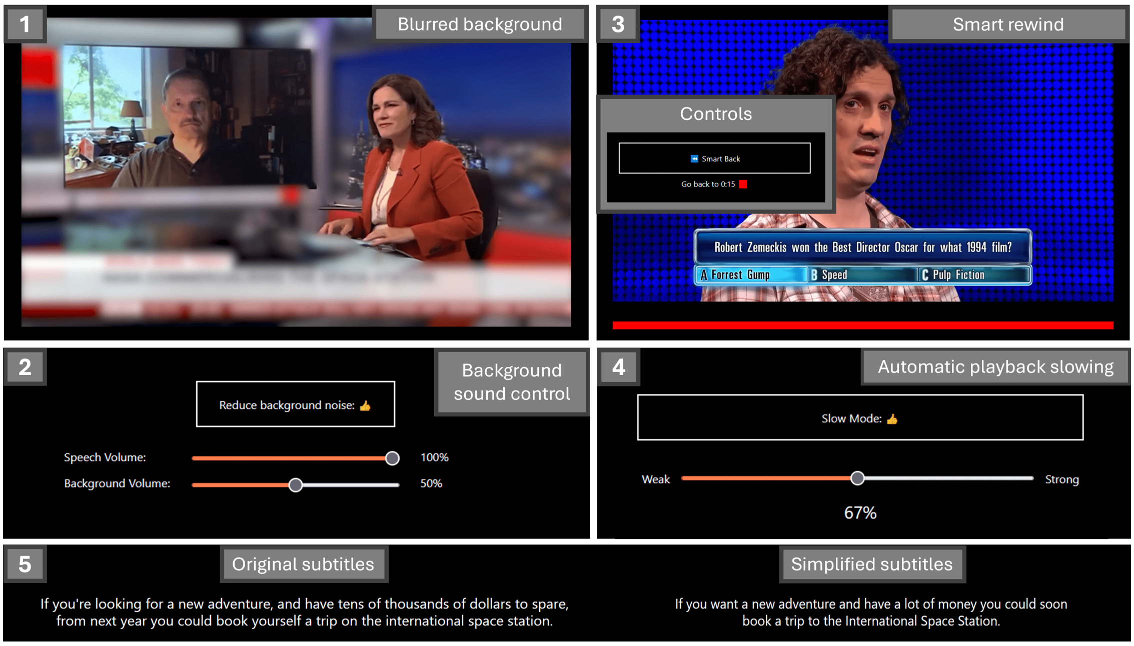

Figure 6.4: The bespoke interventions. Numbers 1 and 2 are Jane's interventions. Number 3 is Simon's intervention. Numbers 4 and 5 are Bart's interventions. Sally's intervention could not be represented visually. See this paper's video captions for a full demonstration of all these interventions.

Jane's Intervention

Jane reported difficulty maintaining attention when distractors were present in both sound and image. In discussion and in her storyboard, she proposed two features to moderate these effects. The first was control over the mix, with separate adjustment of salient speech and ambient elements such as music or environmental sound. The second was a way to concentrate visual attention on key on-screen material. Jane considered two variants for this: applying a background blur that leaves principal subjects in focus, or zooming the frame to isolate the primary action. After comparing the two, she preferred the blur, judging it less abrupt and more compatible with retaining contextual cues while softening visual clutter (see Image 1 in Figure 6.4). For operation, she asked for simple manual controls so that she could enable support when needed and keep it off when watching socially with her husband. The implemented panel therefore provided an on-off toggle, along with sliders to set relative levels for speech and background tracks (see Image 2 in Figure 6.4).

The blur was realised using a Wizard-of-Oz approach. Each clip was edited in DaVinci Resolve [2] to track principal characters and objects, applying blur to the remainder of the frame. During demonstrations, the player switched between the original and the edited version to represent the feature in use. Background audio control relied on source separation. We processed each clip through the lalal.ai platform [3] to derive two stems, one dominated by speech and one by residual ambience. The controller then exposed two independent sliders that set the playback gain for each stem, allowing the viewer to privilege dialogue or lower competing sound without altering the underlying content assets.

Simon's Intervention

Simon proposed a rewind mechanism that would return to a meaningful point in the programme rather than a fixed interval, guided either by narrative structure or by measured audiovisual complexity. He is a regular user of standard rewind controls on streaming platforms and wanted a more targeted tool that reduces repeated tapping and guessing. We implemented a smart rewind feature driven by dialogue complexity and, after early trials, incorporated the pace of delivery so that rapid speech increased the likelihood of a jump.

Subtitles for each clip formed the basis of a metadata file that stored the text, start and end times, and a derived complexity score. The initial score used the Flesch-Kincaid readability metric [152], which is commonly applied when preparing materials for people with aphasia [120]. Adjacent subtitle units were concatenated to meet the 100 word threshold required for reliable computation, and a rolling score was maintained as the viewer progressed through the clip. To account for delivery speed, we added a multiplier based on words per minute (WPM), using the formula $1 + \log_{10}(\max(\mathrm{WPM} - 150, 1))$. The multiplier grows quickly just above 150

[2] DaVinci Resolve https://www.blackmagicdesign.com/products/davinciresolve

[3] https://www.lalal.ai/

WPM and then tapers, which increases the combined score for passages that are fast without overwhelming longer segments.

In operation, *Smart Rewind* jumps to the start of the most recent segment whose combined score meets or exceeds 8, consistent with accessibility guidance for aphasia-friendly materials [120]. This behaviour privileges moments that are both linguistically demanding and delivered at a speed that reduces processing time [134]. Thresholds and window sizes were set conservatively for the workshop prototypes and can be tuned to individual preference or content type. The result is a rewind that targets difficult passages rather than uniform time slices, supporting re-listening with fewer interactions.

Sally's Intervention

During the design conversation, Sally described several situations she finds difficult, including sequences with fast dialogue and rapid visual changes, but chose to focus on dialogue pace. She explored approaches for reducing delivery speed without fragmenting meaning and proposed breaking speech into smaller units so that words and short phrases arrive with more separation. Translating this idea into a working feature, she preferred a text-to-speech reader to perform the lines, since natural speech tends to bind words together and inserting artificial pauses between them can sound odd, reduce enjoyment, and bring limited gains in intelligibility. She did not want fine controls for rate or segmentation, as these invite constant tweaking and draw attention away from the programme. She also asked to disable any alterations during shared viewing with her husband. The resulting intervention therefore supplies a single switch that replaces the original voice with text-to-speech and adjusts playback so that audio and picture proceed at a slower, more manageable pace.

Implementation used the browser's built-in text-to-speech to speak the subtitle text. Subtitle timings were edited to align with the reader's cadence, and sentence boundaries were kept intact to maintain coherent prosody. Slower delivery requires synchrony between the spoken line and the image, so the player reduces video speed to match the duration of the text-to-speech for each caption block. The text-to-speech engine did not provide reliable duration estimates in advance, so durations were measured manually during preparation and stored in a metadata file alongside the subtitle text, start and end timestamps, and the original unmodified duration of the same block. This metadata allowed the player to cue the video accurately and to avoid mid-line drift or truncated phrases when the reader was slower than the original. The intervention remained simple at the point of use, with a single toggle governing activation and no additional parameters exposed. In a production workflow the same timings could be derived automatically from scripts and audio stems, with predicted read times removing manual steps while preserving alignment.

Bart's Intervention

Bart centred his intervention on subtitles. In earlier conversations he described persistent difficulties with on-screen text, including typography that was hard to parse and captions that vanished before he could finish reading. He therefore proposed three features. First, a way to simplify the subtitle language so that captions present a concise summary rather than a verbatim transcript, allowing him to use the text as supportive guidance during demanding passages. Second, an option to have the current subtitle read aloud with text-to-speech when he watched alone or when reading felt tiring. Third, a control that slows playback while captions are active to provide extra time for processing the text.

For operation he asked for simple, reliable controls. Toggles would switch between original and simplified captions and enable slowing during subtitles. This arrangement would let him operate the system independently in solo viewing, yet keep interventions unobtrusive when watching with his wife or other family members. He expressed interest in some adjustment of speed when the slowing feature was active and suggested a single slider to set the maximum reduction. He did not want granular control over the degree of language simplification, preferring an on/off choice that avoided constant tuning. For text-to-speech, he wanted a button that pauses the programme and reads the current line, then returns to normal playback once the reading is complete.

To generate simplified captions, the original subtitle file was processed with ChatGPT [4], using a prompt aligned with aphasia-friendly writing guidance [120]. The output was converted into a metadata file containing, for each caption block, the simplified text, the start and end timestamps, and a complexity score used elsewhere in the system (see Figure 6.2.5). This metadata structure kept original and simplified versions aligned in time so that switching did not affect synchrony with the picture.

The slowing feature used the complexity score associated with each caption block to adjust playback rate. Higher scores triggered a stronger slowing effect, combined with the value chosen on the user's slider, and capped at half speed to avoid excessive elongation of scenes. This approach gave Bart additional time to read when the language was dense or technical without imposing a global slow mode. The text-to-speech function relied on the browser's built-in reader and default voice. Pressing the read button paused motion, played the spoken version of the on-screen caption, and then resumed the programme, keeping interaction predictable and limiting disruption to the viewing flow.

6.2.6 Iteration and Critiquing

We held a second workshop (WS2) to present the implemented interventions and obtain detailed feedback on their operation. The session took place at the same venue as WS1 and included all original participants except Sally, who was absent. Interventions were projected at the front of the room, and each participant operated their own prototype using a laptop. We had planned to use a tablet for control, since a touchscreen would be easier to manipulate than a mouse, but connection issues with the web application led us to switch to the laptop. Each prototype was introduced briefly before participants explored it with four prepared video clips, and we prompted for comments at regular intervals on usefulness, effort, clarity of controls, and any confusion. Participants interacted only with the intervention designed for them rather than viewing others. They paused playback to explain reactions and used rewind to demonstrate specific moments. Engagement varied by feature, with Simon repeatedly invoking *Smart Rewind* to revisit segments, while Bart

[4] ChatGPT: https://openai.com/index/chatgpt/

preferred to toggle his intervention and then watch with minimal further adjustment. Each one-to-one activity lasted about half an hour.

Direct use revealed several refinements, as summarised in Table 6.2. The player initially exposed separate play and pause buttons, which participants found awkward since switching state required precise pointer movement, so we merged them into a single play/pause toggle. Participants also reported uncertainty about whether an intervention was active or what it was doing at a given moment. Simon, for instance, could not tell where *Smart Rewind* would jump. We added visible status and feedback to reduce this ambiguity, including a colour-coded cue on Simon's player that marked higher complexity regions (see number 3 in Figure 6.4), while the controller displayed the destination of the next rewind before activation. Similar indicators were added for other features to clarify on/off state and magnitude without introducing heavy overlays.

Intervention	**Description**	**Changes made**
Jane - Audio Control	Control over the speaker and background audio levels	None
Jane - Visual Blur	Blurs the video background	None
Simon - Smart Rewind	Rewinding to the start of a moment of high complexity	Added visual cues to the video player and controller and incorporated speech rate to the complexity score
Sally - Slowing Speech	Replaces the speaker's voice with text-to-speech	None
Bart - Simplified Subtitles	Simplifies the subtitle text	None
Bart - Slower Subtitles	Slows down the media based on subtitle complexity	Incorporated rate of speech to the complexity score

Table 6.2: List of the four bespoke interventions and a description of their features. Also included are the changes made to the interventions following WS2.

Participants asked that interventions blend with the content rather than dominate the experience. They preferred subtle cues with consistent control placement, while animations were restrained rather than large or intrusive. The text-to-speech features raised concerns about the "robotic" voice, which several participants found distracting. Suggestions included replacing it with a more natural reading voice or, ideally, preserving the original speaker's voice while slowing and clarifying delivery.

While such behaviour could be achieved with AI voice synthesis, participants raised ethical issues around impersonation and consent and requested transparency and user control if synthetic voices were introduced. Overall, WS2 provided concrete direction on control simplification, feedback design, and the balance between visibility and subtlety needed for domestic use.

6.3 Evaluation of the Bespoke Interventions

To assess the utility of the bespoke accessibility interventions for a broader group of people with aphasia, we ran three evaluation sessions with a different cohort at Aphasia Re-Connect. We selected a new charity to obtain an un-primed perspective, with participants having no prior exposure to the earlier workshops or to the individuals who co-designed the interventions. In total we recruited 11 participants (3F, 8M), with a mean age of 65.4 (range 45 – 85) and an average of 11.0 years living with aphasia (range 6 – 20) – see Table 6.3 for details. Groups comprised either 3 or 4 attendees per session, and each person received a 20 GBP Amazon voucher.

Name	**Gender**	**Age**	**Years w/ Aphasia**	**Aphasia evaluation**
P1	Male	58	8.5	ASR: 3 - Receptive and expressive aphasia
P2	Male	54	10	ASR: 2 - Receptive and expressive aphasia
P3	Female	84	20	ASR: 2 - Expressive and receptive aphasia
P4	Female	62	13	ASR: 2 - Receptive and expressive aphasia
P5	Male	73	6	ASR: 3 - Receptive and expressive aphasia
P6	Male	52	12	ASR: 2 - Moderate expressive and receptive aphasia
P7	Male	65	14	ASR: 4 - High level, mild aphasia
P8	Female	85	6	ASR: 4 - High level, mild aphasia
P9	Male	67	16	ASR: 2 - Expressive and receptive aphasia
P10	Male	74	6	ASR: 2 - Moderate receptive and expressive aphasia
P11	Male	45	10	ASR: 2 - Receptive and expressive aphasia

Table 6.3: List of participants in the evaluation sessions. The aphasia evaluation includes the participants' Aphasia Severity Rating (ASR) [144] and a textual description. This evaluation was done by a speech and language therapist who works at the charity.

Most participants reported engaging in social viewing (N = 7), typically with family members. Sessions coincided with weekly group meetings at the charity to keep the setting familiar and reduce additional travel. Two researchers facilitated sessions, introduced the study, gathered demographic information, then demonstrated the four interventions. Participants were invited to operate the controls via a tablet to reduce interaction effort compared with a mouse (see Figure 6.5). Participants commented on ease of use, perceived benefit, and any points of confusion while viewing a set of short clips that spanned multiple genres.

Figure 6.5: Photo from the first evaluation session, showing the setup used to demonstrate the interventions.

Data collection combined structured prompts for discussion with a brief survey. The survey used 5-point Likert scales to capture general judgments of usefulness and clarity and included open text boxes for additional remarks, examples from everyday viewing, and suggestions for control refinements. This mixed approach provided both numerical summaries and context-rich accounts of how the interventions might fit everyday practice.

6.3.1 Data Analysis

With participants' consent, all sessions were video recorded, yielding 4 hours and 15 minutes of material from a 360° camera that captured the whole room during the evaluations. We also recorded the iPad screen and the main display showing the video content. The first author reviewed the recordings in full, then conducted an inductive, reflexive analysis, coding salient points raised by participants. Codes addressed general viewing experiences, accessibility barriers encountered, reactions to the bespoke interventions, and further remarks about audiovisual personalisation. After this initial coding pass, broader themes were constructed from related codes. To support rigour, three authors met to review the coding, discuss discrepancies, and refine theme definitions and boundaries.

6.3.2 Results

We now present the results of the evaluation session, looking at each intervention separately. The overall Likert results representing the participants perceived usefulness of each intervention feature can be seen in Figure 6.6.

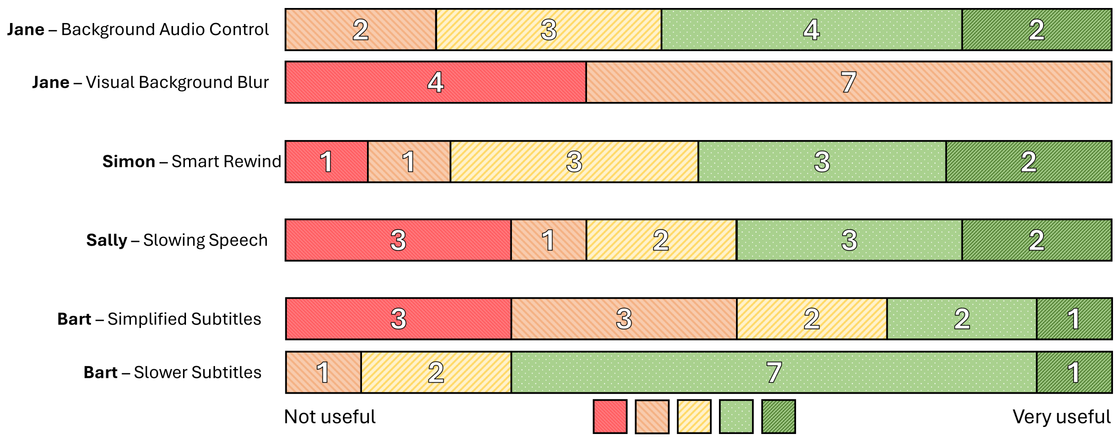

Figure 6.6: Likert data on the perceived usefulness of each intervention feature. The values represent the answer to the question "Did you find the intervention useful to help understand?" A score of 1 representing 'Strongly disagree' reflected in red and 5 representing 'Strongly agree' reflected in green.

Jane's Intervention

Jane's intervention comprised two features, control over background audio levels and a visual background blur. The ability to adjust ambient sound relative to dialogue was judged useful for supporting viewing (see Figure 6.6). Participants described how emphasising speech can make it easier to stay with the thread of a scene when attention is strained, noting specific situations where this matters: "***P1**: Yeah, good... brilliant [...] on a soap... soaps... lot of noise going on there*". Utility was linked to sequences with overlapping speakers, music, or environmental sound where competing audio increases confusion and recovery effort. At the same time, reservations were expressed about lowering ambience because it can remove elements that contribute to tone and meaning: "***P5**: Background noise builds up the tension*". Several framed background sound as a deliberate creative choice placed by those making the programme, not extraneous material that can be discarded without consequence: "***P11**: It is all part of the same... same... I mean, otherwise, they wouldn't have it there in the beginning*". Views therefore balanced access against aesthetics and varied with content type, interest, and momentary difficulty. Participants described a situational trade-off in which clarity gains may be worth minor losses in atmosphere, while recognising that preference shifts with programme and mood: "***P8**: You lose something, but you gain something else*". When considering shared viewing, most said they would rarely operate mix controls with others present in order to avoid introducing friction. Active manipulation was also associated with taking a quasi-directorial role that some did not want during leisure: "***P5**: That is a bit like me being my own director, and that takes a bit away from it*". In practice, participants preferred light-touch operation, clear indication of state, and settings that do not demand continuous adjustment once a workable balance has been found for a given scene.

On the other hand, the blur feature drew little support. Most participants reported that blurring itself attracts attention and introduces effort rather than reducing load: "***P6**: It... it... it is distracting*". When blur was active, several described focusing on the altered regions instead of the intended subject, which undermines the goal of simplifying the visual field: "***P8**: I would be constantly concentrating on the blurring*". Others said that de-emphasising the background can remove cues needed to follow interactions across the wider scene and weaken a sense of place, making comprehension harder: "***P5**: I would find it hard to understand... if I didn't understand the interaction with the other people*". Questions also arose about control logic and agency, including how the system determines which elements are blurred and whether viewers could choose them. While response was generally negative, there was acknowledgement that a similar tool could be useful for people with more severe aphasia, particularly soon after stroke when processing capacity is reduced and fewer simultaneous cues are tolerable: "***P8**: For me, it is not the thing, but... for others, I think [nodding]*". Some welcomed having blur available as an option for specific co-viewers even if they would not use it themselves, emphasising reversibility and unobtrusive design: "***P7**: If you can do it... then why not do it? You don't have to use it... but... if it's... if it's no problem to do it, why not do it?*". Despite openness to optional provision, the prevailing view was that blur would be used rarely, if at all, during shared sessions, and only with clear, immediate control and minimal visual disruption so that attention remains on the primary content.

Simon's Intervention

Participants regarded an advanced rewind that targets difficult segments as useful for supporting viewing. Many already use platform controls such as 10-second jumps, so the concept of returning to a specific point for re-listening was familiar and easy to adopt. Considering complexity when selecting the jump point was seen as a practical way to make re-watching meaningful rather than mechanical: "***P1**: It is just slowing it down to an accessible pace*". Several valued the option to move back to a well-defined moment that mattered for understanding, rather than stepping back by fixed time. One participant asked for direct access to segments that contain data-heavy material, such as numeric graphics, so revisiting concentrates effort where comprehension is most strained: "***P8**: I'd like to see the graphs [...] because my problem is numbers*".

Participants asked for clear feedback about where the control would land to avoid losing orientation. Without visible indication of the destination, the feature risked adding uncertainty at the point it is meant to reduce it: "***P6**: I couldn't understand where the... the smart back was... where?*". Several said they would use smart rewind alongside other adjustments, returning to a challenging passage and then enabling an intervention that reduces competing sound or slows delivery during the replay: "***P5**: You can go back and actually play it without [background noise]*". Unlike interventions that alter voice or image, participants did not view targeted rewinding as compromising the original meaning or tone, since it operates on navigation rather than content.

In social contexts, the feature was judged acceptable in moderation. Participants expected to use one or two jumps to restore context without disrupting the shared flow, while avoiding repeated activation that could frustrate co-viewers: "***P9**: One or two, yes, but three or four or five... [chuckles]*". Preferences favoured simple, low-effort operation with a visible marker for the pending jump and consistent behaviour across clips. Overall, the mechanism was understood and welcomed as a way to regain footing when pace or density outstrips processing time, provided destination cues are explicit and repeated use remains limited during shared viewing.

Sally's Intervention

Participants described speech pace as a central barrier and responded positively to an option that slows delivery to a more accessible rate, reporting clearer understanding when the feature was active: "***P4**: Yeah, yeah... I think it's clearer, yeah*". Appreciation tended to be strongest in sequences with dense dialogue or rapid exchanges where processing time is limited. Several participants said that slower speech made it easier to follow the thread of a scene without resorting to repeated pausing or rewinding.

Reactions to text-to-speech were mixed. While some found the reading helpful, others judged the synthetic voice to be distracting, feeling that it detracted from the experience: "***P1**: I just don't... the computer voice... no, can't... I don't like it at all*". Suggestions for improvement focused on voice quality, including the ability to select a clearer narrator similar to well-known presenters (e.g. Stephen Fry or David Attenborough), whose delivery are widely perceived as easy to understand, in part due to their Received Pronunciation. Participants also recognised trade-offs between intelligibility and performance, noting that synthesised speech can sacrifice expressive detail even when it supports comprehension: “***P8**: With the robot voice, we could lose the sort of the humorous intonation [...] the nuance of speech, but it would definitely help a lot of people by slowing it down*”.

The perceived usefulness was judged as being context-dependent. Participants said they would enable the feature for complex or informational material, while avoiding it when the performance is central, preferring unmodified delivery unless comprehension breaks down: “***P7**: If I slowed it down and I could understand, it’s a no-brainer*”. Several linked the feature to learning contexts, describing potential value as a comprehension aid in speech and language therapy where controlled pacing can reduce stress and support practice: “***P8**: Comprehension when you first get aphasia is very stressful*”. Participants also addressed the interaction, with a simple toggle, clear indication of the active state, and reliable synchronisation between audio and video being favoured over fine-grained controls that demand continuous adjustment. Participants agreed they would rarely use text-to-speech during shared viewing, citing the risk of disrupting collective engagement and drawing attention away from the programme. In social contexts, they preferred unobtrusive support, reserving slowed delivery and synthetic reading for moments when clarity is necessary, as well as ensuring that co-viewers are comfortable with temporary alteration.

Bart’s Intervention

Bart’s intervention comprised two features – subtitle simplification and playback slowing triggered by subtitle complexity. Reactions to simplification were mixed. This was especially true for participants who do not use subtitles at all because of aphasia or visual impairments, for whom changes to wording offered little practical benefit: “***P10**: I watch to one side [...] I used to watch [foreign films] with subtitles, but now I can’t do it*”. Where subtitles remained usable, several participants valued simplified wording as supportive information rather than a verbatim transcript, using shorter phrasing to preserve orientation during dense dialogue or technical explanation: “***P7**: I like the idea of simplifying [the subtitles] because you don’t always need every word to get the gist of it*”. One participant suggested making captions shorter still, presenting keywords on screen to anchor attention without overloading reading. Concerns centred on maintaining the intended message; simplified text can remove nuance and tone. Participants asked for an explicit visual cue whenever the simplified track was active so that viewers could distinguish it from the original and judge acceptability for the current scene.

The option to have captions read aloud with text-to-speech raised similar issues to Sally's intervention, notably the perceived artificiality of the voice: "***P7**: But they sort of tend to pronounce words in a weird way*". Preferences favoured a natural-sounding narrator, ideally matched to the speaker and varied across characters to preserve identity within a scene: "***P9**: First the girl [reported] is talking, then the guys [text-to-speech] is speaking, it's [waves hand around]... that sounds... a disaster*". Participants discussed social viewing and differed on whether they would enable simplification or spoken captions with others present. Some judged the change minor enough to be acceptable, while others preferred to avoid alterations that could draw attention or disrupt shared rhythm, especially where style and performance were central to enjoyment.

The slowing feature was, by contrast, judged the most useful single element in Likert responses. Participants reported that reduced pace made following dialogue easier, particularly when speech was fast, complicated, or laden with specialist terms: "***P8**: I take more time*". One participant also suggested using the feature to support language practice soon after stroke, appropriating the intervention for learning goals that emphasise controlled pacing and repeated exposure to words: "***P8**: We all want to extend our speech to use more colourful and interesting words, and I think television and radio helps us to do this*". A visible indicator of when slowing was active was appreciated, since automatic reduction can alter pronunciation and timing. Participants asked that cues be clear yet unobtrusive so that the state of the feature is evident without masking content.

Effort costs associated with extended running time were raised across multiple sessions. While shorter sequences were acceptable, longer programmes risked feeling protracted, particularly in film-length content: "***P9**: But um long um 30 minutes, then 40 minutes... yeah... 3 hours long, oh!*". These considerations were pronounced during shared viewing, where some were reluctant to make co-viewers spend longer on a programme. Preferences favoured reversible, scene-scoped slowing with straightforward controls, consistent behaviour across clips, and an instant return to normal speed when comprehension was restored. Overall, participants positioned subtitle simplification as situational support with clear labelling and text-to-speech as contingent on voice quality, while they treated pace reduction as broadly useful provided operation remained simple, feedback was visible, and total viewing time did not increase beyond what the household accepted.

6.4 Discussion

We now discuss what bespoke interventions reveal about making audiovisual media accessible for people with aphasia. Our findings indicate where personalisation produces practical gains, where it introduces friction, what this implies for design, domestic use, and production workflows. We relate these points to earlier chapters by treating comprehension as the design target, while locating change at the media object rather than only at the interface. Where relevant, we reference concepts introduced in Chapter 2 to orient the reader between design ideals and implementation constraints.

6.4.1 Personalisation and Viewer Agency

Personalisation shifts the viewer's role from configuring peripheral settings to selecting when and how core elements of delivery change. Across the bespoke interventions, participants preferred controls that are lightweight, reversible, and scoped to scenes or speakers, with clear feedback with minimal continuous adjustment. Automation was acceptable when transparent, otherwise manual operation was favoured. These preferences align with an agency-centred view of personalisation, and expand on the distinction between adaptable controls and adaptive behaviour introduced in Section 2.2.

Empowering the Viewer

Offering technical interventions that assemble accessible versions of audiovisual media through personalisation positions the viewer as an active participant during viewing while inviting artists to prepare alternative renditions that remain artistically or semantically equivalent. This arrangement supports a flexible mode of consumption that changes how viewers engage with programmes, matching what Chaume [57] describe as the "audience's turn". Interventions that let viewers tailor content to their needs can shape actions and perceptions by enabling purposeful operation even when interaction is light [288]. Our participants' responses to Simon's Smart Rewind illustrate this shift, using targeted navigation to restore orientation and maintain enjoyment without altering the source assets.

The medium influences understanding when it becomes personalisable at the level of delivery rather than surface presentation. Controls that translate the media for comprehension increase a sense of agency and allow viewers to manage effort during continuous playback [145]. This accommodates diversity without assuming homogeneity across communities that include older adults [242] and people living with CCNs. Reactions to the bespoke concepts varied. Some judged features potentially useful for other people with aphasia, for earlier stages of recovery, or for wider audiences such as non-native speakers. This echoes how established interventions spread beyond their original scope and become part of everyday practice.

Reducing distracting background information was widely understood as beneficial in principle, yet preferred methods differed. Feedback on Jane's Visual Background Blur expressed disagreement about visual suppression as a route to focus, with some participants preferring audio-level control over ambience rather than manipulating the image. Participants recognised variation in each other's aphasia profiles, and described adapting their own viewing to accommodate co-viewers, treating assistance as part of shared practice. Where household preferences dominate, simple presets that fit typical routines can reduce negotiation effort.

Utility extended beyond immediate aims. Like subtitles, several features can support communities outside the initial target when paced delivery or simplified text aids comprehension. Participants discussed features as resources to be appropriated for personal goals, such as language practice or recovery of confidence after periods of reduced viewing. Interventions provided value through adaptability across contexts, either by repurposing single features for specific barriers or by combining multiple adjustments to create a tailored configuration. Indeed, as discussed in Chapter 4, many of the accessibility barriers experienced occurred in conjunction with one another, often exacerbating the overall accessibility of the media. Intervention combinations were favoured when a single change did not suffice, for example pairing smart navigation with reduced background audio in dialogue-heavy scenes.

Such flexibility keeps technologies workable as needs change over time. Access requirements vary within and across diverse communities and call for different configurations to maintain understanding during everyday viewing. Lightweight operation supports continued use because it minimises friction. Reversibility preserves trust because a viewer can return to the original quickly. Clear feedback makes state visible without drawing attention away from the programme. Taken together, these preferences point to personalisation that privileges comprehension over continuous micromanagement and that can scale across household arrangements, genres, and devices without diluting creator intent or social fit.

Appropriating Technology to One's Needs

Designing technologies so that people can appropriate them to their own needs is essential when working with diverse communities. In our study, participants discussed the co-designed interventions not only as viewing support but also as tools that could be used in speech and language therapy. Increased control over speech pace through features such as Simon's Smart Rewind and Bart's Slower Speech was viewed as helpful for re-learning language understanding, giving space to process words and phrases at a manageable rate. Interactive tools can strengthen motivation to engage with technology by increasing perceived competence and autonomy [287], and participants framed these controls as practical means to manage effort during demanding sequences.

Technology appropriation commonly extends beyond the original use case, particularly among marginalised communities for whom tools are rarely designed. People adapt available systems to fit their circumstances, whether by reconfiguring interaction sequences, repurposing interfaces, or changing basic input arrangements such as the layout of a computer keyboard [203]. Personalised interventions that expose simple, reversible controls enable this kind of adaptation without heavy configuration. They let viewers select when support appears, decide how long it remains active, and return to the original delivery instantly when the aid is no longer needed.

Providing these affordances aligns with calls to design for appropriation rather than prescribing a single correct way to use a system. For instance, Spiel et al. [280] argue that technologies should invite situated re-use and reconfiguration so that people can shape experiences to fit their abilities, preferences, and contexts. The interventions developed here follow that direction. They operate at the level of delivery, offer clear feedback, and avoid continuous micromanagement, which makes it feasible to appropriate them for aims beyond entertainment, including practice, confidence building, or collaborative support with significant others. In everyday use, such flexibility allows individuals to assemble workable viewing arrangements that reflect changing needs over time.

6.4.2 Inherent Tensions in Personalisation

Personalisation improved comprehension and control for many participants, yet it introduced trade-offs that shaped willingness to use the features. These tensions clustered around social fit in shared viewing, perceived shifts in creative intent when delivery was altered, interaction effort when controls required constant adjustment, and practical limits in production and tooling. Participants asked for transparency about what had changed, reversible operation, and cues that were visible without dominating the screen. They also weighed clarity against duration, since slowing delivery increased running time. This section examines these tensions and identifies design choices that reduce friction, including scene-scoped controls, restrained signalling, and conservative automation that is easy to override.

Social Tensions

Viewing is a social activity that involves co-presence, shared attention, and conversation [96, 204, 324], yet much HCI work has treated accessibility as an individual configuration task rather than something negotiated in households and social groups. Personalisation features operate within relationships, routines, and norms, consistent with Shinohara and Wobbrock [272] who state that 'technology use does not happen in a social vacuum'. Across our participants, engagement patterns differed, with some expressing a preference for greater independence during co-viewing, while others described interdependent support as the most workable arrangement for maintaining comprehension without fragmenting the joint experience. Several participants noted that household norms about pace, pausing, and interruptions shape whether interventions feel acceptable during shared sessions.

A recurring account involved choosing not to modify playback when watching with others even when this decision reduced comprehension or enjoyment, a choice often framed as avoiding deviation from what was perceived as the default experience of neurotypical viewers, which accords with reports of stigma around assistive technology use [71, 226] and appears in our vignettes in Section 6.2.4. The decision not to intervene can produce a dependency arrangement in which a partner or family member provides on-the-fly explanation while the viewer remains present, yet partly excluded from the content, since narrative details are filtered through another person and opportunities to re-listen or slow delivery are forgone to preserve the shared flow. Additionally, participants linked repeated reliance on such explanation to fatigue and disengagement over longer programmes.

Participants described strategies that limited friction in social contexts by keeping changes small in scope and short in duration, such as localised rewinds to recover a missed line or subtle rebalancing of audio so that speech remains clear without flattening ambience, and many preferred to reserve more assertive alterations for specific moments of difficulty rather than applying them throughout a programme. Second-screen control was seen as a practical way to operate features privately, reducing visible signals on the main display, which aligns with accounts of temporal agency in companion media that allow the flexible occupation of shared time [54]. Similarly, several asked that any on-screen indicator of state be restrained so that attention remains on the programme rather than the intervention, in line with their demand for privacy.

Other participants treated social viewing as intrinsically interdependent and framed requests for brief explanation, a short replay, or slower pacing as legitimate co-viewing moves rather than evidence of deficit. This is consistent with work that positions interdependence as central to assistive technology use [28]. They described the conversational interaction around a difficult passage as part of the programme experience rather than a departure from it, emphasising that joint decisions about when to adjust delivery can preserve enjoyment for everyone while sustaining access for the person who needs support, with some households reporting routines that alternate control or agree thresholds for pausing so that decisions feel fair across viewers.

These social tensions have direct implications for how personalisation should operate during shared viewing, because acceptability depends on involving all stakeholders. Based on our results, reversible, low-effort controls that can be applied discreetly and withdrawn instantly could help reduce these social tensions. Indeed, clear behaviour that remain predictable across viewing patterns is crucial. Features that alter performance or timing need to have visible labelling so that co-viewers can decide together whether the change suits the current scene. Companion controllers can carry private status cues and controls, so that the main screen remains free of distracting overlays. These controls could include a ‘social mode’ which would limit intrusive operations, cap cumulative slow-downs, and prefer brief, local actions to preserve rhythm. Presets or defaults that can be invoked silently on a companion device, together with a single ‘return to original’ control that is always available, can help keep negotiation light, while retaining agency. The inclusion of consent prompts for interventions that substantially change the media could support agreement among co-viewers before activation, which reduces surprise and maintains trust in households that value continuity of performance.

Creative Tensions

Personalising audiovisual media raises tensions about shifts in creative vision, intended viewing experience, and potential changes in meaning, with participants expressing concern when adaptations departed from the original message. They described a boundary between accessible delivery and authored performance, where personalisation that modifies voice, timing, or composition can feel like a move away from the work that artists intended the audience to interact with. These concerns were more pronounced when changes were persistent rather than applied briefly to overcome a specific difficulty, with several participants suggesting that personalisation be framed as an alternative rendition, rather than an invisible edit, so that viewers understand how the experience differs.

Examples from discussion made these tensions concrete. A busy background soundscape can raise effort for dialogue comprehension, yet ambient sound is often placed to convey tone or situational detail, so aggressive rebalancing risks removing cues that make the scene legible. Participants worried that viewers might watch nominally the same programme yet come away with experiences that are hard to compare, which complicates post-viewing discussion. Concerns intensified for informational content (e.g. news or documentaries) where modification could misrepresent [85]. Altered track could therefore carry provenance metadata accompanied by a visible indicator so that audiences can judge the suitability of the accessible version for their purpose.

Creators may resist personalisation they do not author because it reduces control over performance, rhythm, or audiovisual composition, particularly where character identity or sonic texture carry narrative weight. A practical route is to involve artists in producing explicitly accessible versions through flexible media techniques, ensuring alternatives remain aligned with the original intent even when they are not formally equivalent. This approach shifts change from after-the-fact manipulation to planned assembly, enabling scene-scoped alternatives for pace, mix, or wording that respect narrative boundaries while maintaining coherence. Where voices are replaced or synthesised, consent as well as attribution should be embedded in the artefact, with clear presentation to viewers before activation.

Future work should engage audiovisual artists alongside production teams to characterise what constitutes acceptable transformations at the level of scene, sequence, or programme, while developing workflows that carry timing, structure, or semantic metadata required for reliable assembly. Co-design in adjacent domains demonstrates value when reimagining accessibility with communities who have complex communication needs [33]. Similar collaboration can ground acceptable ranges of personalisation, keeping authors involved in specifying alternatives that preserve meaning. Evaluations with households that include both disabled as well as non-disabled viewers can test whether labelled alternatives support conversation or joint decisions about use, informing guidance on equivalence classes that remain robust for everyday viewing.

Interaction Tensions

Personalising audiovisual media creates an interaction problem because each alteration exposes multiple parameters that many viewers wish to tune, which expands the space of possible operations during playback. During demonstrations, participants quickly moved from trying features to proposing ways to reshape them for their own needs, which mirrors findings with viewers who have ADHD where audiences request adjustable sound channels plus bespoke playback controls [139]. A single programme already contains numerous audiovisual elements, so author-provided alternatives create further branching. Each new control multiplies the paths a viewer can take, turning occasional choices into a mesh of decisions that cascade. This raises a basic question: which changes produce access for a given person, and which technical capabilities are necessary to deliver those changes without overwhelming the session.

Participants often cited controlled pace as a practical route to improved comprehension, yet there are many ways to achieve slower delivery. For instance, playback rate can drop globally, pauses can punctuate difficult passages, a clearer voice-over can replace rapid or mumbled delivery, and navigation can return to the point before confusion. Each method depends on thresholds, durations, or segmentation rules, so parameter load rises quickly. The ideal of fine-grained personalisation must coexist with a desire to sit back. Many viewers prefer to settle into a programme rather than manage continuous adjustments, consistent with the account of select-and-settle leisure [103]. Hyper-personal tailoring invites substantial effort from the viewer, which undermines expectations of seamless accessibility during domestic viewing, with prolonged tweaking being seen as work rather than relaxation.

Once interaction reaches high dimensionality, a trade-off emerges between access and usability. A narrow control surface can lift comprehension while staying workable, while a broad control surface risks friction because choices interact in non-obvious ways. As features accumulate, complexity does not grow linearly as interactions intersect and different effects mask one another. Viewers then chase a moving target, tuning one parameter after another in search of stability. To contain this growth, lessons from interactive media research are useful. Work on player agency examines how systems present meaningful actions at the right granularity so that choices remain intelligible without exhaustive menus [291, 294]. Applying that lens to accessible playback suggests privileging actions that map to perceptible units of the programme such as scenes, speakers, or narrative beats. Controls that operate at those units reduce the need for continuous adjustment, since each activation carries a clear scope, with our participants preferring inputs producing a complete, bounded effect.

These observations place a constraint on how far personalisation can extend before it impairs the experience it seeks to improve. Systems that invite reflection about when to intervene, that temper parameter growth through coherent groupings, and that keep operations intelligible at the point of use are more likely to support sustained viewing. Personalisation remains valuable, yet its interaction must respect leisure rhythms so that comprehension gains do not come at the cost of excessive control work during ordinary sessions.

Production Tensions

Producing personalised audiovisual media requires a marked shift in how content is produced and delivered, since existing pipelines optimise for linear media rather than versions reconfigured at playback. Armstrong et al. [17] describe the effort involved in a single piece of flexible media, a radio documentary built with object-based audio that supports variable length, which demanded bespoke processes across recording, editing, and rendering. Artists and producers must anticipate how discrete elements will interact once manipulated during viewing, so that meaning persists across alternative rendition.

The move toward an adaptable media collides with institutional realities. Many broadcasters operate workflows shaped by decades of linear practice, a layer of technical debt that resists significant changes. As Cieciura et al. [58] report, projects that target highly personal experiences often begin by creating tools, workflows, and metadata models from scratch, which fragments practice across teams, leaving artefacts that do not interoperate. Automation can streamline portions of the process, yet building or maintaining these systems is costly in time and budget, as well as requiring specialist expertise across different production roles.

Recent technological developments illustrate the scope of change required, such as DanmuA11y [328], which shows that converting visual-centric designs into accessible formats depends on context and user experience choices that must be captured within the production process itself, rather than bolted on after export. Legacy organisations face additional pressure. Public broadcasters, as described by McChesney [186], carry national service mandates and legacy infrastructure, while on-demand platforms (e.g. Netflix, Amazon Prime Video) introduce distinct expectations that shift audience norms [297]. The result is a production landscape where experimentation occurs, yet large-scale adoption stalls unless infrastructures evolve alongside creative practice.

Further work should focus on foundations that make personalisation reliable in everyday use, with producers engaged early to surface constraints. People with disabilities should also be involved at the outset to specify changes that deliver improved access [214]. A bespoke co-design approach can supply grounded cases that expose where, for example, the alternative tracks, timings, or mixes are necessary, yielding concrete requirements for production teams. Early engagements can identify elements suitable for manipulation, which helps production teams plan assets that assemble predictably at runtime. They can also reduce incompatibility between projects, improving the odds that personalised versions remain portable across distribution channels.

6.4.3 Personalisation and Bespoke Interventions

The bespoke co-design showed that needs and preferences vary widely, with some features proving transferable, while others remaining highly individual. Personalisation was most workable when controls were lightweight, presets matched typical situations, and changes were confined to moments of difficulty rather than applied globally. This section synthesises principles from the interventions and evaluation, outlines where individual settings generalise to families of cases, and describes requirements for consistent behaviour across viewing contexts without burdening the viewer with continuous configuration.

Ideals of Personalisation

Personalisation of digital audiovisual media permits changes to the media object as well as to the viewing experience, which can address diverse preferences across audiovisual and informational dimensions. The intent behind such personalisation often diverges, with platforms, including broadcasters, seeking novel products to match market dynamics [303], whereas people living with disabilities call for greater access suited to varied needs. Our findings point to demand for tighter control over pace, key audiovisual elements, and the complexity of presented information, which aligns with the barriers reported by people with aphasia during viewing seen in Chapter 4. Participants in the evaluation described deeper engagement when they could influence delivery through adaptive slowing or targeted navigation. For instance, Simon's *Smart Rewind*, once its complexity metric reflected access needs more closely, helped viewers regain footing by returning to moments that mattered for understanding, which increased temporal agency during ordinary sessions.

The route by which personalisation reaches the viewer determines what changes are feasible and how they feel in practice. Personalisation applied implicitly to already produced material introduces compromises because finished assets are static. Adjusting playback to slow delivery can distort speech timing or prosody, as Bart's experience showed when pace reduction affected the natural flow of dialogue. Sally's text-to-speech mitigated timing issues, yet it departed from expectations about voice and performance, which broke immersion for several viewers. An explicit approach prepared during production can avoid such trade-offs. Flexible media techniques with richer metadata make it possible to assemble accessible versions that have been planned from the outset rather than patched after completion. Where dialogue is dense or fast, an alternative recording with clearer wording and moderated pace can be authored to serve comprehension while the original remains available. Such alternatives do not need to claim formal equivalence, instead they can augment the work in ways that preserve intent and keep creators involved in specifying acceptable variation [55]. This shifts effort upstream, but it yields delivery that behaves coherently at playback, since adjustments operate on assets designed to be interchangeable rather than on fixed mixes.

Personalisation, in this framing, becomes a property of the content rather than a post-hoc fix. Viewers gain options that map to their access requirements, creators retain oversight of how those options relate to meaning, while platforms obtain artefacts that support consistent assembly across devices. The result is a viewing experience in which change serves comprehension without collapsing tone or narrative structure, provided preparation occurs during production rather than only in the player.

Scalability of Personalisation

Developing bespoke accessibility interventions that meet an individual viewer's needs raises questions about transfer to others and about scale in industrial settings. Large platforms operate within entrenched structures shaped by legacy practices, which makes integration of novel approaches slow. The push to deliver a perfect experience for every person sits uneasily with the logic of scalable production described by Tsing [296], where output grows without reworking underlying methods, while pursuit of a single common denominator can flatten difference. As Pfotenhauer et al. [232] note, such framings tend to sideline alternative problem definitions that support minority groups. Oliver [217] offers transport access as a case in which calls for scale were used to justify designs that many could not use. The relative absence of interventions for people with complex communication needs on mainstream services suggests a similar pattern within HCI for audiovisual media, where translation between modalities dominates while delivery-level adjustments remain rare.

Addressing this imbalance requires evidence that links lived barriers to workable alterations rather than abstract principles. Involving people with aphasia through bespoke co-design provides such material because it exposes concrete breakdowns during viewing and ties them to specific manipulations that restore comprehension. Individual features may appear narrow in scope, as with Jane's background blur, yet the underlying problems recur across the population. Treating each bespoke outcome as an instance of a wider class allows researchers to map families of barriers and the transformations that respond to them. This aligns with Harper [118] on design-for-one, where uniqueness is the starting point for systematic adaptation rather than an exception to be averaged away. It also matches arguments by Hilgartner [123] that innovation often begins at vanguard sites where participants challenge defaults shaped by business logic. Constraining inquiry to a single intervention type per person further reduces the search space, which supports depth over breadth, thus yielding sharper requirements for content handling across viewing contexts. Consistency at playback remains crucial so that controls behave predictably across material – manipulation must preserve a clear heuristic of use [174]. Shared exemplars that tie scenes to concrete adjustments could help production teams plan assets that assemble reliably at runtime, reducing guesswork during implementation. Additionally, publishing small, well-specified cases with accompanying metadata would enable comparison across projects and gradually build a library of patterns that speaks to both producers and disabled communities.

This perspective reframes scalability. Rather than promising a universal solution that fits everyone, platforms can accumulate supported cases grounded in real viewing, each with defined limits of operation and known effects on meaning. Over time, coverage increases without erasing difference, creators retain involvement in how alternatives relate to intent, and viewers gain options that reflect the variety within disability, rather than a single imagined user.

6.4.4 Limitations

This study reports bespoke co-design of highly personal accessibility interventions with people living with aphasia, and contains several limitations. The cohort was relatively homogeneous, as participants were drawn from WEIRD countries [172], lived in urban locations, and had familiarity with technology. The sample does not capture wider diversity within disability communities, particularly those with complex communication needs [205, 230].

Use of the co-designed bespoke accessibility interventions in everyday settings remains uncertain. Sessions relied on short video clips to elicit reactions while keeping tasks manageable. This design constrains insight into longer viewing, cumulative fatigue, and persistence of effort over full programmes. Tasks were paced with breaks to support participation, which further limits observation of sustained use during continuous playback.

Broader social viewing was not examined. The interventions were evaluated outside domestic co-viewing, so effects on household routines, negotiation over control, or tolerance for alteration remain unclear. Field studies in real contexts would address these gaps by observing activation patterns, frequency of use, and longer-term uptake across typical schedules.

These limitations place this work as an initial account of bespoke personalisation with this group of people with aphasia. Subsequent research should recruit broader demographics, include viewers with limited technology access, and observe viewing across complete programmes. Longer deployments would permit assessment of fatigue, effort costs, and stability of preferences once novelty fades, while studies run with co-viewers would document how personalisation fits social practice.

6.5 Conclusion

This chapter examined bespoke audiovisual interventions co-designed with four people living with aphasia and evaluated their transfer, following a 'design-for-one' approach. Across design and evaluation, interventions were thought of in real-life viewing contexts, addressing the needs participants had in their day-to-day viewing. Usefulness varied by person, content, and situation, arguing for families of adjustments rather than a single solution.

The four interventions offer different ways of addressing a shared problem space. Adjusting dialogue relative to background elements addressed situations where ambience competed with comprehension. Smart rewind targeted recovery by returning to the start of a difficult segment rather than by fixed steps. Slowed delivery using synthetic speech treated fast or indistinct voices as a timing constraint that could be eased. Subtitle simplification and complexity-tied slowing reframed text as support for following speech and plot rather than as verbatim transcription. Together these ideas position access around moment-to-moment manageability, not overarching solutions to barriers.

Evaluation with eleven people with aphasia produced differentiated responses. Background audio control was often seen as useful because it made the current speaker easier to hear in busy mixes, whereas blurring the visual background drew attention to the manipulation itself and was generally disliked. Smart rewind was valued when the interface showed where the jump would land; without feedback it created uncertainty. Slowing delivery improved clarity for some, but distracted others because of the synthetic voice. Simplified subtitles helped when treated as an additional cue, though concerns were raised about altered meaning. Slowing tied to subtitle complexity was frequently helpful, yet viewers worried about the extra time it could add to long programmes or when watching with others. Overall, interventions worked when they lowered effort at the right moment while being perceptible.

Control strategy and interaction cost strongly shaped acceptance, as participants preferred toggles and presets to continuous fine-tuning, wanting clear local indicators when an adjustment was active, and asked for a single way to revert to the unmodified presentation. Automation was acceptable only with transparency and easy override. Even features that ran 'in the background' needed minimal cues so expectations aligned with what was seen and heard.

Social viewing further conditioned use, as our participants would avoid visible or time-expanding changes when watching with others. Short explanations from co-viewers were common and sometimes preferred to using an on-screen aid, keeping the viewing experience social. Features that altered voices or dialogue pace were seen as least acceptable in shared sessions. Accessibility is therefore embedded in household routines, where willingness to intervene depends on how changes affect companions as well as the viewer who needs them.

Concerns about tone, character, and authorship were explicitly mentioned by participants as concerns. Lowering background elements risked removing atmosphere used to signal setting or tension. Replacing a distinctive voice with a generic synthetic reading was rejected in some genres but tolerated for informational material. Where wording was simplified, participants wanted a visible state so they always knew which version was active. These views support the value of alternative versions prepared alongside the main cut, with selection left to the viewer, so intelligibility improves without discarding what gives the work its identity.

The evaluation showed broad interest in navigation that returns to meaningful boundaries, in slowed delivery when language is dense, and in simple controls for dialogue prominence. It also revealed polarisation, especially around more intrusive adaptations like the background blur and synthetic voices split opinion. This variation reflects the heterogeneity of aphasia and argues for a wide range of adaptation options rather than fixed settings, so viewers can shape a version that fits their needs.

Several design directions follow. First, timing and navigation aids scoped to scenes or speaker turns are promising, provided interfaces make their operation clear and reversible. Slowing tied to locally estimated complexity with unobtrusive indicators fits this pattern. Second, balance controls that favour dialogue over background have practical value when offered as quick presets rather than continuous mixing tasks. Third, affording additional cues that support viewing without adding needless distraction, such as focusing gaze or giving simplified subtitles.

Finally, the chapter records tensions these comprehension-first accessible personalisation implementations must accommodate: visibility in shared viewing, risks of altering meaning or feel, the effort of control, and production realities when preparing alternative versions. Addressing these tensions requires designs that change little by default, make adjustments explicit, and let viewers decide when a change is worth the trade-off. The next chapter builds on these findings in domestic settings, studying a focused set of interventions during everyday viewing.

Chapter 7

In-Home Deployment

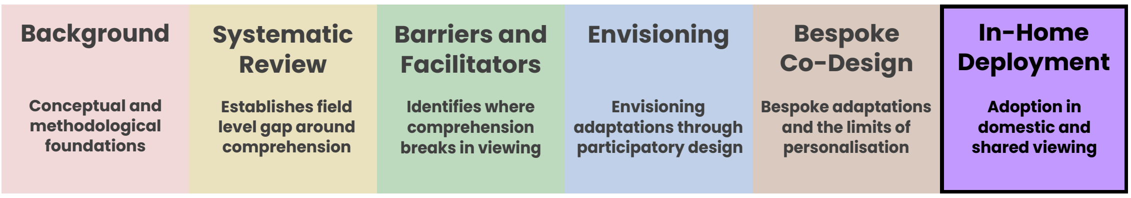

Figure 7.1: Map of thesis - deployment

7.1 Introduction

Building on the bespoke interventions and mid-fidelity prototypes reported in Section 5.2.5 and Section 6.2.5, along with their evaluation in Section 6.3, this chapter examines how accessibility personalisation operates in everyday viewing through an in-home deployment with nine people living with aphasia. This chapter directly addresses limitations raised in Chapter 5 and Chapter 6 by moving from short clip demonstrations to whole-programme viewing in domestic settings where comprehension breakdowns occur, observing co-viewing in practice, and examining effort and persistence over a multi-week period. This extends our RtD methodology by placing intervention artefacts into routine use and observing how participants adopt, adapt, or abandon them.

We deployed a simple content delivery platform in participants' homes for two weeks with nine participants ($N = 9$), providing a small library of broadcast material and a custom player with four accessibility features aligned with prior co-design. This custom player included four main interventions: simplified captions, speaker spotlight, playback speed controls, and multi-track volume controls. Participants used a tablet and, where preferred, cast to their TV with a Chromecast. Usage was logged and event-triggered ecological momentary assessments captured experience during viewing.

This chapter asks how personalisation can reduce interaction latency during continuous playback in domestic contexts and how adaptations are negotiated when watching with others. The remainder of the chapter sets out the system deployment, reports quantitative and qualitative accounts of the system's use, and examines implications for designing accessible personalisation that fits everyday practice.

7.2 Methodology

This section outlines the methodology for the in-home deployment that examined personalised audiovisual accessibility features for people with aphasia. We first give a system overview covering the four accessibility features in the player, then sets out the study design, recruitment, demographics, and ethical approvals. We then describe the in-home setup, equipment issued to participants, and additional support materials. Following this, we outline the data collection and analysis methods used, including usage logs, event-triggered ecological momentary assessment (EMAs), and pre- and post-deployment interviews.

7.2.1 System Overview

The deployment used a browser-based application with a compact video library and a custom player that exposed four accessibility controls. Figure 7.2 shows the control layout of the application and Table 7.1 lists the features and their descriptions. The interface followed feedback from sessions described in Chapter 5 and Chapter 6. The feature controls were placed on a page separate from the video to suit tablet and casting use, keeping the TV image uncluttered. The video library contained five items supplied by the BBC: a nature documentary, a travel cooking episode, a talk show episode, a comedy episode, and a quiz show episode. These were chosen to represent different genres and production styles using real broadcast content, and the average duration was 39 minutes (range 22 to 58 minutes). All participant interactions within the application were logged to a database. The system also included event-triggered EMAs [269]. The EMAs appeared when the viewer exited full screen, asking for a 5-point Likert rating of the general viewing experience. This design avoided text-heavy, specific questions that could be inaccessible for people with aphasia on a small screen.

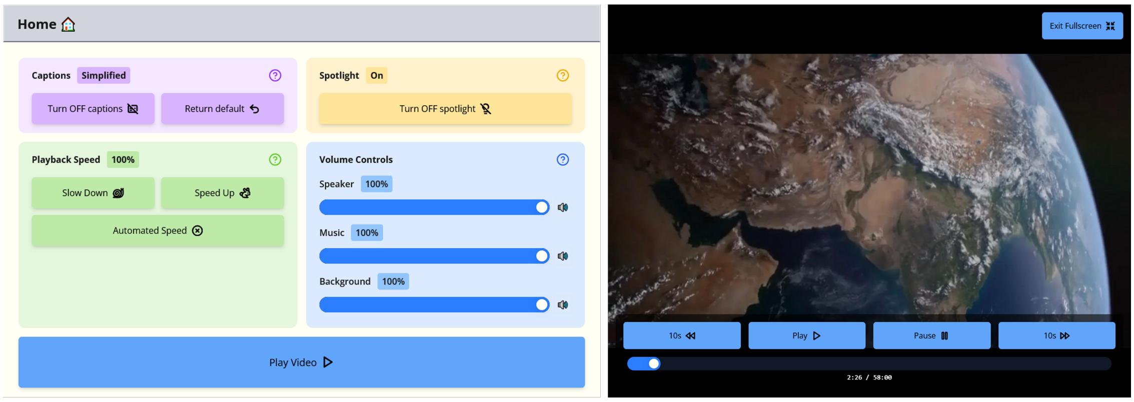

Figure 7.2: Screenshots of the system. On the left is the feature controls page and on the right is the video player.

The features had no default settings, with all controls being set to neutral values, so viewers adjusted them from the start for any effect. The subtitle controls switched between verbatim and simplified text. The simplified subtitles were created programmatically with a large language model with explicit instructions to align to aphasia-friendly guidance. The simplified subtitles were manually checked, along with programmatic checks to ensure appropriate reading level. Original and simplified subtitles shared the same timings. The spotlight added a visual emphasis on the current speaker. The spotlight overlays were prepared manually in advance by editing each video, with the activation of the feature switching to the edited stream. Playback speed was available as manual 5 percent steps or automation. The automated slowdowns used the same pre-processing as Figure 6.2.5, combining subtitle-derived complexity with estimated words per minute to trigger local changes.

Features	Description
Simplified Captions	The user can turn captions on or off. When the captions are on, the user can choose to see a version of the captions that replaces verbatim speech with shortened and simplified text.
Spotlight	The user can toggle on spotlighting, which highlights the current speaker on the screen.
Speed Controls	The user can manually adjust the playback rate of the video by 5% increments, or allow it to automatically slow down the content based on pace and complexity of speech.
Multi-Track Volume Controls	The user can control the volume of the speaker, music, and background audio tracks.

Table 7.1: List and description of the four accessibility features available to the participants.

Automated pace was off by default. The multi-track volume controls included a volume slider for speech, music, and background sound, with the different audio tracks being separated programmatically using a custom Python script that used the Audio Separator package [1].

These interventions were selected based on our prior work in the co-design and evaluations reported in Chapter 5 and Chapter 6, selecting the preferred interventions from families of interventions. Multi-track volume controls addressed frequent reports of speech masking by music or ambience. Pace controls responded to difficulties with rapid delivery and dense dialogue, with manual adjustment for increased control and automation for less active viewing patterns. Subtitle simplification provided concise support where reading speed or lexical complexity limited comprehension. Spotlight supported speaker identification in scenes with multiple talkers, facilitating comprehension.

Participants were given a Samsung Galaxy Tab A9+ and a Google Chromecast. We offered to connect the Chromecast to their television so the tablet screen could be mirrored to the TV. This arrangement standardised interaction with the system while keeping viewing in a familiar setting. We provided an aphasia-friendly user guide

[1] beveradb/audio-separator

with step-by-step instructions for using the tablet, casting to the TV, and operating the four features, as well as including contact details for support. See Figure 7.3 for the materials given to participants.

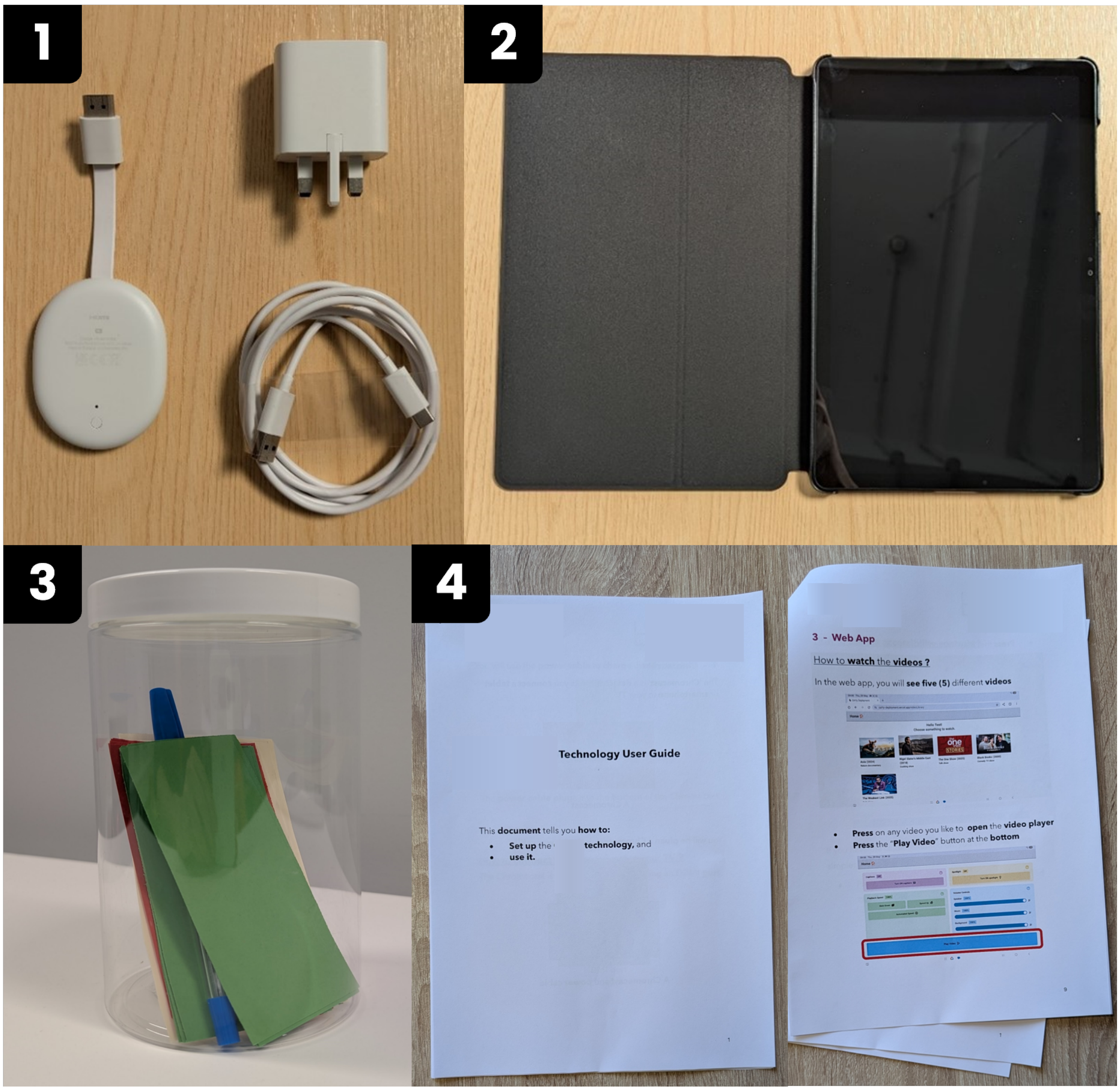

Figure 7.3: Equipment given to participants for the deployment: (1) the Chromecast, (2) the tablet, (3) the notes jar, and (4) the user guide.

7.2.2 Study Design

In our prior work in Chapter 5 and Chapter 6, alongside co-authored work [33], we co-designed personalisation features for people with aphasia. These studies, however, did not capture long-term or everyday viewing in the home since activities occurred in neutral third placed, such as our participants charity drop-in sessions.

Deployments in real contexts can deepen engagement and reflection because participants interact with systems within their routines [151, 169]. This matters for people with aphasia who often depend on contextual cues and support from familiar places and people to aid comprehension [143]. Lab evaluations usually lack these affordances which reduces ecological validity, as we note in Section 5.4.3 and Section 6.4.4. Additionally, research has shown that use contexts shape how people operate and make sense of interactive systems [27, 179]. Embedding technology in everyday environments can also support ownership, reflection, and advocacy among participants which is valuable when working with marginalised groups [165]. Collecting data through context-aware and event-triggered mechanisms in the home can surface situated experiences that are rarely available in lab studies [124].

In this deployment we treated the system as a knowledge probe – an artefact placed in everyday life that invites interpretation and appropriation, generating insights through its use rather than through controlled tasks [100]. Our player acted as such a probe by offering adaptable features during ordinary viewing, inviting reflection. Participants could adjust, ignore, or combine the different features, while the system recorded these choices and their timing. Participants annotated the system following Gaver's account of annotation as knowledge production [100]. Annotation occurred through event-triggered EMA ratings when exiting full screen, through in-app logs that captured feature activations, setting changes, and session boundaries, and through aphasia-friendly interviews that elicited concrete episodes of difficulty and repair. Several participants kept brief notes during the deployment by using the notes jar, as well as their own note taking. These traces and accounts served as annotations attached to the artefact, linking behaviour and interpretation to specific moments of use. By running the deployment in participants' homes, we enabled them to familiarise themselves with the technology, explore features at their own pace, and engage with content in a setting aligned with their natural viewing practices. This preserved ecological validity and revealed how participants adapted and responded to the technology across individual and shared viewing.

7.2.3 Participant Recruitment and Demographics

We recruited 9 participants (2F, 7M) for the in-home deployment through Aphasia Re-Connect, which provides group support for people with aphasia, and through the City St George's, University of London mailing list. Ethical approval was granted by the King's College London Ethics Board. Recruitment operated under strict inclusion criteria. We sought people with aphasia who self-reported difficulty watching television at home, and who were willing and able to host the system for a multi-week deployment, including the two home visits. Individuals who did not report viewing difficulties were excluded during screening.

At the weekly drop-in sessions of Aphasia Re-Connect, members of the research team introduced the study to attendees, answered questions, and provided an aphasia-accessible participant information sheet [254]. Screening occurred at the same sessions. We asked brief, concrete questions about everyday viewing, such as whether participants struggled with fast dialogue, dense sound, or on-screen text, and whether they would be comfortable running a system at home for several weeks. Those who expressed interest and met the criteria received the information sheet to review with their support network in their own time. Research materials were prepared with input from a speech and language therapist with 18 years of experience working with people with aphasia and followed appropriate guidance [120]. Communication with prospective participants used supported communication techniques [143], including structured prompts, visual aids, and time for clarification.

At least one week after first contact, those who remained interested met a researcher, optionally with a supporter of their choice. The researcher reviewed the information sheet using supported communication, answered questions, provided clarification, and checked understanding. Only then did we ask whether they wished to take part. Those who agreed completed an aphasia-accessible consent form in writing. After consent, we collected demographic data and practical details needed for the home deployment. These included the participant's home address, contact information, next of kin contact, the presence of animals in the home, and any pre-existing medical conditions the team should be aware of. The same process applied to mailing list recruits, except meetings occurred online or by phone, with documents sent in advance and collected at the first home visit. Screening for the mailing list pathway mirrored the criteria used at Aphasia Re-Connect, excluding respondents who did not report barriers when watching television or who could not host the system.

Participants had a mean age of 60.6 years with a range from 39 to 77, and had lived with aphasia for an average of 12.4 years with a range from 1.5 to 50+ years, as can be seen in Table 5.1. Experience with technology varied, as well as the patterns of video use, from occasional film watching to spending much of the day with television and online video. All participants were fluent in English before their aphasia. **P3** and **P8** had non-English mother tongues, Luganda and Kirundi. All participants had aphasia following stroke except **P4**, who had an intracranial subdural haematoma. Some participants also presented with co-occurring conditions. **P6** had dysarthria and dyspraxia, **P7** had cognitive communication disorder, and **P8** had apraxia. Five participants, **P3**, **P4**, **P5**, **P7**, and **P8**, had right-sided hemiparesis. None used high-tech augmentative and alternative communication during our interactions. Each participant received a 100 GBP voucher for a shop of their choosing in recognition of the time commitment involved in hosting and using the system across the deployment period.

Name	Gender	Age	Years w/ Aphasia
P1	Male	71	14
P2	Male	53	12.5
P3	Female	58	50+
P4	Female	73	1.5
P5	Male	44	11
P6	Male	77	3
P7	Male	60	5.5
P8	Male	39	2
P9	Male	70	12

Table 7.2: Deployment participant demographics.

7.2.4 Deployment Procedure

Following the participant recruitment, we arranged a convenient home visit to install and demonstrate the system. We began with a pre-deployment semi-structured interview. This was audio recorded only, which limits the non-verbal communication common with many people with aphasia, yet it reduced intrusion in the home. No more than two researchers attended each visit to avoid crowding.

The interview had two parts. First, we assessed self-esteem using the 10-item Visual Analogue Self-Esteem Scale (VASES) [44]. We used VASES at the start of the deployment to establish how participants felt about themselves in relation to everyday activities that include watching television. We did this because we wanted to understand whether interacting with the system and its accessibility features affected how participants felt during viewing, as well as whether repeated use would relate to any change in self-reported confidence about watching. Second, we ran a semi-structured interview supported by short 5-point Likert items. These covered everyday viewing routines, understanding of video content, how participants communicate about what they watch, how they feel while watching, and the technologies they use to access video at home. We asked about barriers, workarounds, and current accessibility feature use.

We then set up the technology and demonstrated basic operation. This included connecting the Chromecast to the television and opening the web application on the tablet. After our walk through, we asked participants to repeat the steps themselves to confirm that they could carry out core actions without assistance. Any issues or questions the participants had were addressed in the moment. The researcher did not leave until the participant could operate the system independently.

Participants received clear instructions for the deployment. They could choose when to watch and how to watch, and they could re-watch segments or whole videos as they preferred. Participants did not have to watch every video if they did not want to. They could watch directly on the tablet or cast to the television using the Chromecast. They could pause or stop at any time, or adjust feature settings in any way that suited them, and they did not need to keep features active if these were not helpful. If they encountered problems, they could contact the team and request a visit or a call.

To support later conversation, we provided each person with a plastic memory jar, coloured cards, and a pen. The memory jar was a simple aid that participants could use to capture brief notes during their everyday viewing. Green cards were available for positive points, red cards for negative points, and white cards for neutral points. Participants were invited to write down short descriptions of specific moments that they found helpful or difficult, including what feature they used, what happened on screen, or how they felt. These notes were intended to help them remember and discuss particular events during the post-deployment interview by providing concrete anchors rather than general impressions. For those who found writing difficult, they could take photographs on their phone to capture a moment they wanted to discuss later, or have a viewing partner help them write down their thoughts. The memory jar notes and photographs were treated as prompts to structure the conversation rather than as formal diaries.

A researcher contacted each participant after one week to check in, address any issues they had, and schedule the post-deployment interview and equipment collection. Where difficulties were reported, these were resolved with the participant before the final interview. The post-deployment interview mirrored the structure of the first. We again administered the 10-item VASES to understand how participants felt at the end of the deployment and to compare with the initial scores [44]. We then conducted a semi-structured interview that revisited the same topics, focusing on viewing while the system was in the home. We asked about the web application, overall experience, and the accessibility features. We discussed which features supported understanding, how participants used the controls, and any moments where the features were avoided. We also asked about shared viewing with family or friends, and what changes participants would want to make to the features or controls to better fit their needs. During the interview, participants used the notes from their memory jar and, where applicable, their own notepads, to bring up specific scenes or situations. These materials helped locate discussion in concrete examples.

7.2.5 Data Analysis

As part of the deployment, we collected quantitative and qualitative data. The quantitative data comprised usage logs, Likert responses, EMA ratings, plus VASES scores recorded before and after the deployment. We summarised feature usage and session metadata to characterise viewing behaviour. VASES scores were compared using a Wilcoxon signed-rank test and a Bonferroni-corrected alpha for multiple comparisons. The small sample did not yield significant differences, so we additionally report descriptive statistics without inferential claims.

For the qualitative data, all interview audio was transcribed using the Nvidia Parakeet automatic speech recognition (ASR) model. People with aphasia often produce speech with filled pauses, hesitations, atypical prosody. Many ASR systems suppress such language tokens, which can remove cues needed to interpret effort, timing, or repair. The selected ASR retained these filler words and pause markers, preserving conversational rhythm. We then reviewed each transcript in NVivo 15 [2], corrected non-recognitions by the ASR model, restored missing or wrong words, and aligned speaker turns.

We conducted reflexive thematic analysis following Braun and Clarke [37]. The manual pass fixing the transcript also served as familiarisation with it, as recommended for such analysis. Coding was inductive at the semantic level. Early themes addressed feature adoption, interaction effort, perceived benefits, drawbacks, effects of shared viewing. Themes were iterated in analytic meetings, with codes, boundaries, and exemplars discussed, revised, or merged to improve coherence and fit with the data. Disagreements were resolved in these discussions.

7.3 Results

This section reports results in two parts. The first presents quantitative measures of engagement from feature usage logs alongside the event-triggered EMA ratings

[2] https://lumivero.com/products/nvivo/

to characterise viewing behaviour over the deployment. The second presents the thematic analysis of the qualitative data, organised as themes that describe adoption of accessibility features, operation of the player, perceived benefits, reported difficulties, interaction effort, and the role of co-viewing in shaping use. Representative quotations from interviews ground each theme in specific moments, and we relate themes to usage traces where relevant.

7.3.1 Quantitative Results

Participants generally reported that the accessibility features supported understanding of video content, as can be seen in Figure 7.4. The *multi-track volume control* was the most valued feature, with eight participants rating it as helpful or very helpful. This evaluation aligns with usage data, where *multi-track volume controls* were active for 92.7% of total watch time. When this feature was in use, viewers most often lowered the background and music tracks, with background reduction present during 90.4% of watch time and music reduction during 88.6% of watch time. See Figure 7.5 for a summary of feature use.

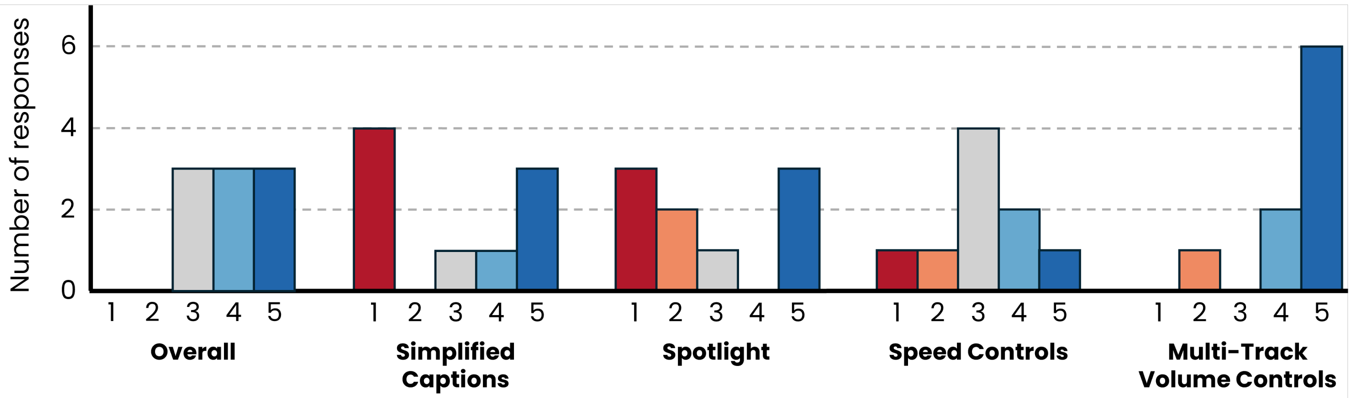

Figure 7.4: Histograms of Likert scores for how helpful the overall system was, as well as for each individual accessibility feature.

Speed controls supported comprehension in moments of rapid delivery rather than acting as an essential feature throughout. They were enabled for 65.9% of total watch time. Manual adjustment accounted for 36.6%, while automated slowing accounted for 29.3%. When manually slowing playback, viewers reduced speed by

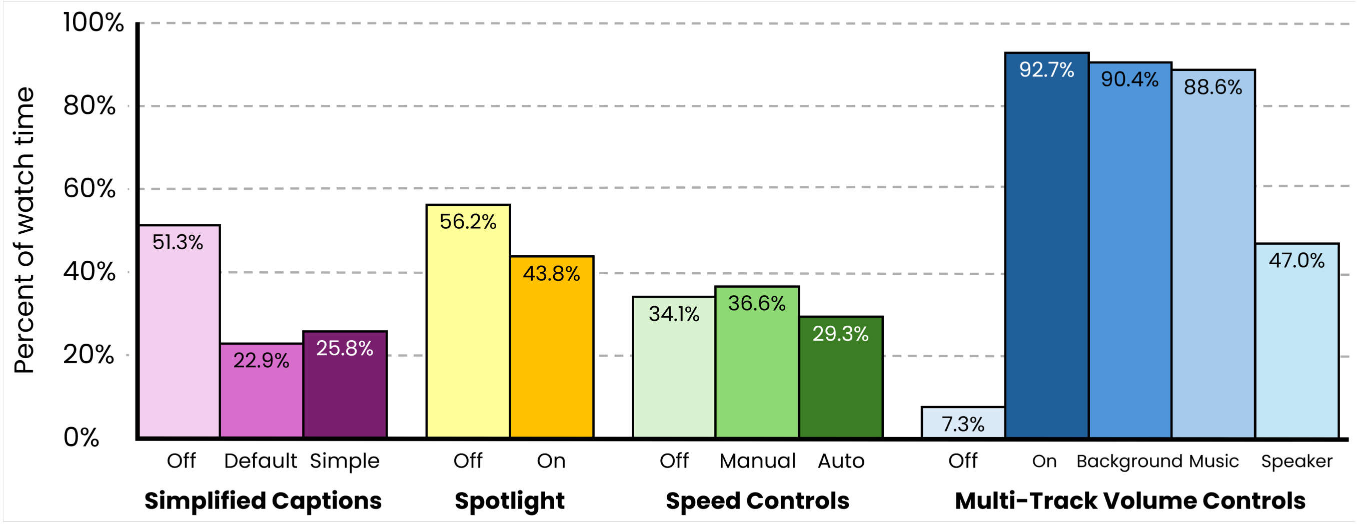

Figure 7.5: The total use of the accessibility features and their settings as a percentage of the total watch time.

an average of 21%. Subtitles showed varied adoption, with *simplified subtitles* being enabled for 25.8% of watch time, standard subtitles for 22.9%, and subtitles were off for 51.3%. *Spotlight* showed mixed uptake, enabled for 43.8% of watch time. These patterns reflect different strategies for managing effort during viewing and show that most adjustments were light touch and situational rather than global.

Usage logs indicate that participants watched an average of 4 hours and 18 minutes across the two week deployment. These logs provide a stable account of how features were incorporated into everyday viewing. The predominance of *multi-track volume controls* points to the central role of audio balance in making dialogue intelligible without sacrificing the overall programme. The use of *speed controls* suggests that pace often required local management. Subtitle behaviour demonstrates heterogeneity in reading comfort and preference among people with aphasia. Examining individual feature use, three viewing patterns emerged – passive engagement with features, sustained active adjustment during viewing, gradual convergence on settings that suited the viewer. **P6** illustrates passive engagement, making few changes after an initial setup – see Figure 7.6 for an exemplar trace of the *multi-track volume control*. **P3** exemplified active adjustment, frequently modifying settings while watching, while **P9** adjusted across sessions to locate values that worked for him.

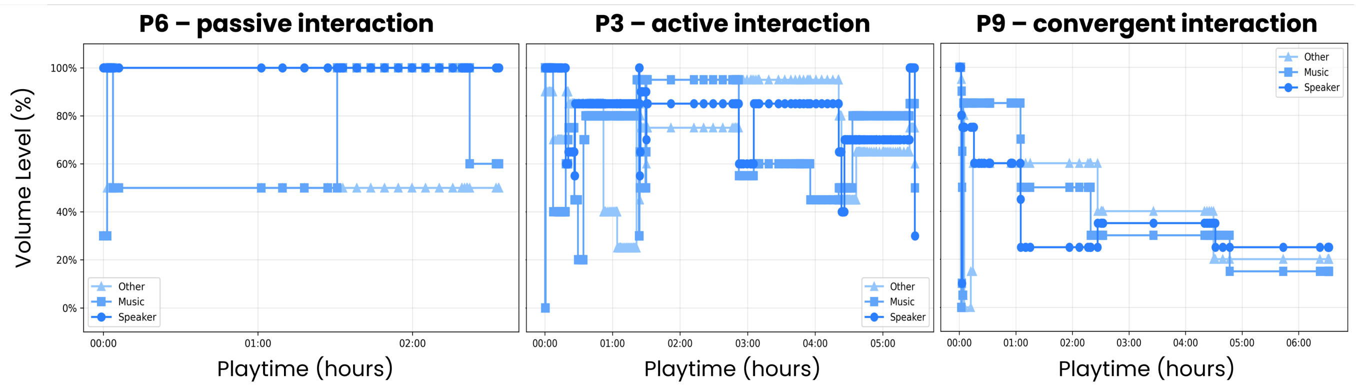

Figure 7.6: Line charts showing the settings of the multi-track volume controls for three representative participants (P6, P3, P9), showing how much these participants interacted with the settings across their watch time.

EMA ratings varied across sessions and showed a small increase over time as participants became more familiar with the system – see Figure 7.7. To make the aggregate EMA plot comparable across viewers who had different numbers of sessions, we normalised each participant's series by session count. Normalisation was computed programmatically using the formula $(sessionNumber - 1)/(maxSessionCount - 1) \times (userSessionCount - 1)$ for all $sessionNumber$ for all users. This converts each session number into a value from 0 to 1, then scales it to each user's total number of sessions. When the normalised position fell between two actual sessions, ratings were linearly interpolated between the adjacent values. The procedure avoided over-representing heavy users and produced trajectories that were legible on a shared axis. The resulting plot shows that small improvements tended to occur as viewers explored the system and settled on workable settings, although individual paths remained varied.

Turning to the VASES, pre-post comparisons were conducted using a Wilcoxon signed-rank test using Bonferroni-corrected α of 0.005 for the ten conditions. None of the comparisons reached significance under the corrected threshold, we therefore report descriptive summaries – see Table 7.3 for all VASES statistics. Several items showed little to no change, such as feeling cheerful, outgoing, mixed up, intelligent, angry, optimistic, and frustrated. Participants reported slight increases in feelings such as being understood (+0.88) and confident (+0.25), while feeling trapped decreased (−0.63). Standard deviations indicate notable variability across participants,

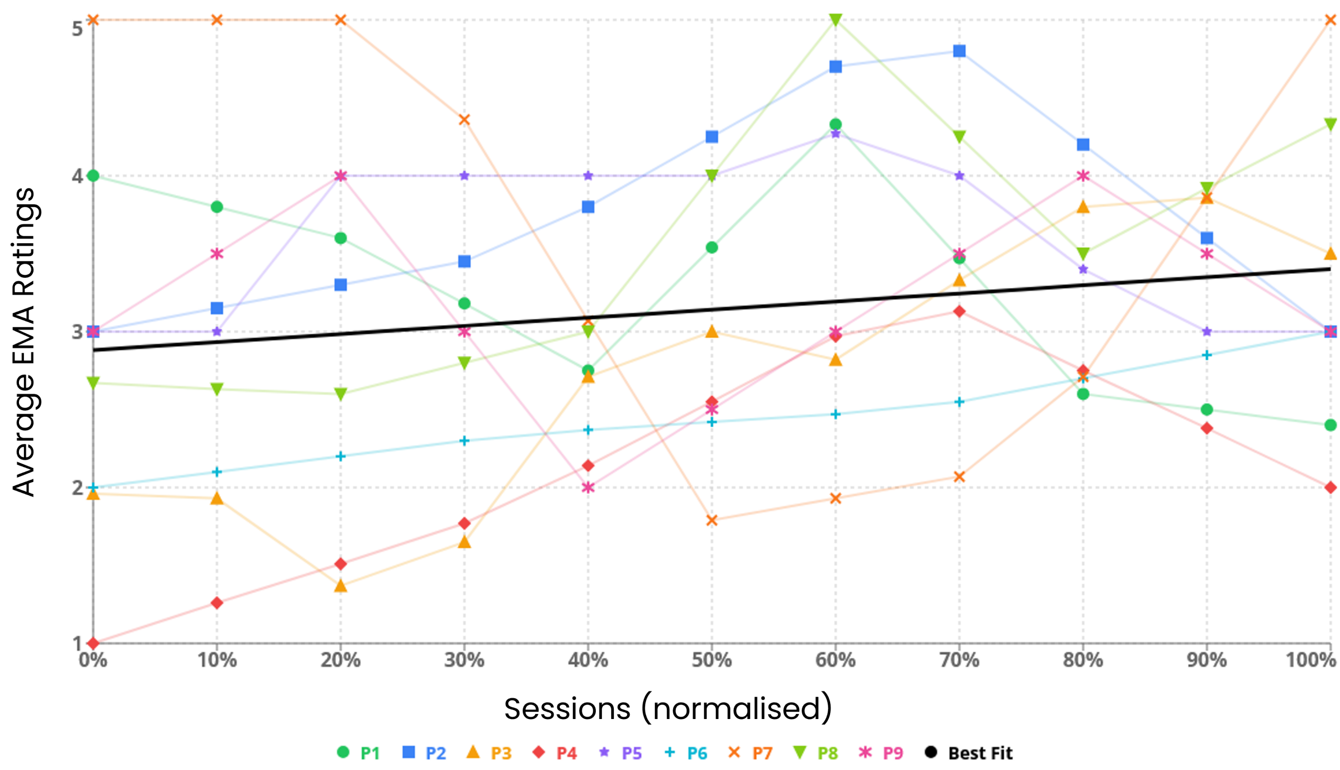

Figure 7.7: Participants' self-reported EMA ratings across viewing sessions. The viewing sessions are normalised to prevent participants with anomalous session counts from skewing results.

which is consistent with a cohort that differs in aphasia profile, viewing habits, and comfort with technology, as well as the small sample size. These descriptive outcomes suggest that extended use did not produce large shifts in self-reported feelings over two weeks, yet small positive changes were present for several items.

7.3.2 Qualitative Results

We now turn to how participants used and made sense of the features during everyday viewing, looking at initial exploration, settling into routines, breakdown and repair, and negotiation during co-viewing. We describe perceived benefits, points of friction, and the ways that context shaped decisions to adjust or ignore the features.

Emotion	Mean difference	Standard deviation	Wilcoxon Signed Ranked Z-value	p-value
Understood	0.88	0.93	2.10	0.036
Confident	0.25	1.30	1.05	0.295
Cheerful	-0.13	0.78	0.00	1.000
Outgoing	0.13	1.54	0.91	0.361
Mixed up	0.13	0.93	0.37	0.715
Intelligent	-0.13	1.17	0.84	0.402
Angry	-0.13	0.33	0.00	1.000
Trapped	-0.63	0.99	1.83	0.068
Optimistic	0.00	0.71	0.67	0.500
Frustrated	-0.13	0.78	1.07	0.285

Table 7.3: Descriptive statistics of the VASES.

Patterns of Feature Adoption

Participants followed different accessibility feature use patterns. They familiarised themselves with the accessibility features in different ways, and settled into viewing habits that met their own preferences and needs.

Exploration Prior to Adoption. Participants began by exploring each of the four features and comparing their effects to baseline viewing. **P1** described this initial phase as "*experiment with various things*", while **P9** tried the *spotlight* feature "*just to compare*", a pattern that aligns with the early spike in interactions visible in Figure 7.6. Many approached this exploration on a per-programme basis, testing features against different genres to judge fit. As **P7** put it: "*I tried every program with it on. Didn't find it useful, but did try it on each program*".

Participants often proceeded through iterative adjustments rather than a single decision. They tuned settings, watched a little, then revised. **P4** captured this stepwise process: "*So, in the end we ended up with, after trial and error, we ended up with* 65% *[speed]. Spotlight on. Uh... volume, speaker, uh... a hundred percent, music...* 50%, *background...* 50%. *And then that's, that's the ones we... we tried*". Through these cycles, participants formed workable understandings of what each control changed and when it helped.

Exploration also clarified mismatches. Reflecting on simplified subtitles, **P7** concluded that the feature did not map to his needs, explaining that he "*tried them and I think, no... that's not my disability. So I, um... switched them off*". In contrast, *multi-track volume controls* encouraged targeted experimentation to separate sources that interfered with dialogue. For **P7**, laughter was a recurrent irritant, leading to tests that isolated its origin. He reported that he "*hated it with the laughter on*" and eventually "*realise that the music and the laughter track are different things*".

Learning Curve to Adapt to Own Needs. While initial exploration revealed which features were useful, a learning curve remained for selecting features and navigating the system as a whole. When first trying to use the controls, **P6** struggled to identify what each feature did, stating "*well, we didn't at all have any success the first time so I phoned up.*". Similar confusion appeared for **P9**, who described early use as "*to use it now... turn off... turn on... sometimes this was a bit confusing*". With repeated use, familiarity grew. Participants learned how each feature adapted content and where settings were located. **P3** captured this progression, stating "*Because when you left... I... I'm... I'm confused. And then again and again and again and then... I got it*". This pattern aligns with gradual improvements seen in EMA ratings over the deployment. Early confusion imposed cognitive load. Operating several personalisation options required memory for control locations, plus an understanding of effects across different types of content. This overhead reduced enjoyment during learning and points to a need to streamline onboarding. **P9** reflected on this barrier: "*I found it hard to activate it at the start [...] because I didn't know.*".

Curated Feature Use. As familiarity grew, participants selected a subset of features that supported their own viewing, setting aside options they judged unhelpful. One participant summarised this choice: "***P1**: Captions, no. It didn't seem to make any difference [...] uh... I couldn't see what it was trying to do*". Selection reflected which elements of the programme each person considered necessary for comprehension. For **P6**, most non-speech audio felt distracting, leading to a narrow configuration:

"*Well, the speaker is always on. Music, I don't really care about the music, and the background is... No, I don't care about that.*". Feature use also depended on whether the perceived benefit outweighed the extra work required to operate adjustments during viewing. **P5** weighed the trade-off explicitly: "*It was quite useful... But... But... Um... I don't know... Um... I could have done it, um... I could have done it... I could have done it... Without speeding up, I mean slowing down...*". In practice, features were kept active when they produced clear gains with minimal effort. Others were left off to avoid adding interaction steps that disrupted the flow of watching. Content shaped preferences as well, where some programmes already matched a comfortable delivery, so additional slowing was unnecessary or counterproductive. As **P1** noted: "*So for the Asia documentary with David Attenborough, his normal pace is [slow]... So slowing it down was a bit too slow.*".

Habitual Viewing Affecting Feature Use. Habitual viewing shaped which features participants integrated into their routines. This was most apparent with captions, which many people with aphasia consider difficult to use because reading can be tiring and slow. Even when captions were available, several participants described pausing or rewinding to try to read them in full: "***P2:** [Captions] on the screen, I didn't use that... because I can't keep up with them. Always known... If it's something that I've got... Um... I would have to go back and... and check... the writing... again*". In contrast, some were comfortable reading subtitles and preferred to keep them enabled at all times: "***P5:** I've always done it... Um... With the subtitles anyway. So, I couldn't ever watch TV without it*". These accounts show how long-standing habits influence whether text support is used during everyday viewing. Existing device behaviour also influenced choices. **P8** reported that he did not need the *multi-track volume controls* because his television already improved speech clarity automatically. He explained: "*I tried [multi-track volume controls]... yes, but, in the end... the... no, no need... volume included... speaker, music and background, it is in... it is already included in... in the program, because when a speaker is... speaking, and the music and the background... it chooses automatically*". In cases like this, a built-in setting shaped preferences and reduced the perceived need to adjust the mix on the tablet. Some participants preferred not to disturb their usual flow of watching. They resisted pausing to change settings while casting or moved to a simpler arrangement that avoided extra steps. As **P1** put it: "*So, I would just watch it on the tablet for the rest of the day.*". Here, maintaining continuity of viewing took precedence over fine-tuning features. Across these examples, the role of routine and existing equipment guided whether adjustments were made, with captions, audio mixing, and control placement fitting or clashing with established practice depending on the person and the context in which they watched.

Passive or Active Engagement. Participants engaged with the features in two main ways, passive use or active use. In the passive pattern, viewers set up a small number of adjustments that worked for them, then, as **P2** put it: "*just leave it*". This approach suited those who wanted minimal effort during leisure. **P4** described preferring to avoid further steps once viewing had begun: "***P4**: When, once you set it up and so you didn't have to go into it and make alterations, and then go and start watching it again*". Some kept helpful options enabled throughout rather than revisiting controls. **P5** explained a steady use of one feature: "***P5**: Way back when... Um... Two weeks [at the beginning of the deployment], so, at the first... But... I... So turning on the spotlight, I use once... But then turn it on and... Um... All the time*". Other participants took an active approach, adjusting settings in response to what was happening on screen, such as **P9**, who reported that he "*[changed settings] quite a lot*". **P1** described settling into the programme and then testing options to deal with specific barriers: "***P1**: settled into watching the program I thought, what happens if this, then I've done this*". These accounts show contrasting preferences, where some participants avoided interaction to preserve a simple flow, while others used the controls dynamically to manage pace, audio balance, or clarity in the moment.

Controls and Interaction Expectations

This theme examines participants' preferences for interacting with personalised content, as well as their expectations of such a system.

Default to Simple and Intuitive Interactions. Participants repeatedly called for simple, transparent controls that made purpose obvious at first glance. From the outset, the interface needed to show what each feature changed, so that operating it did not add extra effort during viewing. Perceived agency was tied to the control type, such as **P1** remarking: *"I suppose [the slider] gives me the impression that I can control it more*". The tactile motion and visible scale helped some participants form a clear link between action and effect. **P7** described this preference: "*Because then I can control what I'm seeing, but the actual feeling of doing a slider rather than a button [...] Because the slider, I can control what I hear [...] Because I can put it all the way down or all the way up or half"*.

Clarity at the point of use affected whether features were tried at all. Where interactions layered steps or concealed outcomes, participants hesitated, such as **P9**, who stated: *"No, it doesn't think [layered interaction] worked well for me, because I couldn't understand."* Simple, self-evident operations lowered the threshold for experimentation during a programme. Participants described moving towards controls that could be adjusted quickly without losing the thread of the content. Sliders, single toggles, and direct numeric steps were preferred over sequences that required navigating away from the video or recalling hidden states.

Defaults that behaved sensibly were seen as helpful during early encounters with the system. Participants wanted a baseline that felt familiar, with the option to refine later once a workable routine emerged. Presets that mapped to common situations reduced initial configuration while leaving room for personal choice. **P8** suggested a straightforward presentation for pace: "*Let's say... this is like normal speed, and then you have numbers that... just like... put like this and then you pick the number, rather than the writing going up and down."*.

Minimising Interaction Latency. Participants reported frustration with having to interrupt viewing to adjust features. This was a deliberate design choice that placed controls away from the video to keep the screen uncluttered, yet the separation introduced friction that broke immersion. While controls were judged usable, leaving the content to reach a settings page disrupted the flow, as **P1** described: "*It was a bit frustrating. You had to stop... and go back to the... the controls were all frustrating... purely because... you have to go to a separate screen... The controls themselves were okay... but... just having to go back... all the time*".

Participants also commented on the repetitive nature of these interruptions. Returning to the controls, then back to the video, created a cycle that felt cumbersome during ordinary viewing. **P3** captured this experience: "*Oh god. And then you go back. And then play. So, if you don't want... If you don't want, then click in. This is the questions [...] And then... Again. And then... press. And then pause. And then you can play.*". These delays functioned as an accessibility barrier in their own right. Breaking the flow to operate adjustments was frustrating even when the features improved comprehension. Participants asked for controls that could be used in real time without leaving the content. **P4** stated a clear preference for in-context operation: "*People find it more satisfying if they could actually do the changes as they're watching, rather than going into the screen and sort of having the screen somehow, you know, as a bit of a sound effect or something*".

Differing Levels of Control. As participants became familiar with the accessibility features, several asked for finer control over the ones they used most. **P1** proposed adding more granular adjustment to *spotlight*, stating: "*Um, although I think the highlight could have been more intensified. I suppose you gave it a slider on that as well*". Others focused on the presentation of text, with **P5** reflecting on *simplified captions*, suggesting controls over subtitles appearance: "*Uh, maybe, um... Big or small letters*". Some participants were unsure which additional options would help, but wanted the possibility to tailor settings if needed, as **P5** put it: "*More [control], but I don't know what they would be. Um... I mean... I don't know*". A different preference was to keep interaction simple. **P6** favoured fewer choices and judged the current set sufficient: "*This is enough. This is good [...] This is already plenty*". Concerns about complexity extended to audio, such as when **P2** warned that increasing the number of mixable tracks "*would make it more difficult*", while **P3** preferred a narrower set of tracks: "*Fewer [volume control tracks], I think speaker and music, that's it*". Participants also described expectations for straightforward setup and operation, seeking a default that worked without effort at the start, with simple steps to change settings only when required. **P4** summarised this view as "*I would say, you know, sort of plug and play*".

Enhancements and Constraints to Personalised Viewing

This theme examines how content adaptation affected participants' viewing experiences, alongside factors that limited their ability to adapt content.

Tailoring to Personal Needs. Participants valued being able to adapt the audiovisual mix to suit individual needs, chiefly to reduce distractions and prioritise speech clarity: *"**P2**: I could adjust it. It made things, the background, ground noise adjustable. And that was good [...] You know, the, um... things can distract you from hearing."*. For **P5**, lowering non-speech audio supported focus on dialogue; he set music near 30 and background near 25, then reported: *"Well... Music... I like it...* 30*? And background...* 25*? But, um... it enables me to listen to the speaker. Um... as opposed to music, which is not... that... not that... crucial*". Some participants were willing to invest setup time before a programme to reach workable configurations; **P6** described pausing for 10 minutes: *"**P6**: I had to pause for* 10 *minutes on the tablet just before starting a new program*". Fine-tuning often occurred at the start of a viewing session, with iterative adjustments to suit content and context, such as **P3** began at 55 and continued to tweak: *"**P3**:* 55*. Not bad. And then... you can... play... you can... set it up [further]... like this*". Through these adjustments, participants exercised control over the experience, making the content 'theirs', and treated configuration as a routine part of settling into a viewing session.

Enhanced Understandability. Participants reported that the accessibility features made programmes easier to follow, which increased confidence and comfort during viewing. **P1** reflected on this when asked about improved understanding: "*Yeah. I might not, but [chuckles] I could watch and listen easily*". For **P2**, the *multi-track volume controls "made... made my watching good... better"*, while **P3** found the *speed controls* made the video "*clear [...] it's clearer for me*". These adaptations also enabled activities that were previously difficult. **P4**, who typically struggled to keep up with captions, noted: *"read all the subtitles on when I was showing them, I got through all the words*". When watching Weakest Link, a quiz show with fast-paced questions, **P4** slowed the playback and reported: *"then I could understand what they was asking... and I tried to answer*". The *spotlight* further supported **P4** in tracking turn-taking: "*Oh, yes, because see, even though you can't understand them, you, you know who's talking*". Reducing distractions also supported concentration, as **P8** put it: "*The focus... focus to... to... to... to talk or and watch... without directing distractions... no distractions*". Across these accounts, adaptive features reduced cognitive load and supported a sense of agency, allowing participants to engage with the programme rather than expend effort managing accessibility barriers and, for several participants, audio and pace adjustments became routine parts of settling into a session.

Contextual Barriers and Breakdowns. While many participants found the accessibility features helpful, there were instances where the features did not behave as expected or failed to meet needs. **P9** noted confusion when adjusting audio layers and expressed uncertainty about the available options: "*I think there were, like, the background sound and background sound and the speaker*". In some sessions, features introduced secondary issues, such as for **P1**, who described how the speaker highlight, while useful, was not always synchronised due to occasional delay caused by the Chromecast[3]: "*That means that when I used the highlighter, it was good, but it wasn't always appearing at the right time [...] counterproductive*". Likewise, **P4** found that slowing down playback sometimes caused unexpected problems: "*The slower you made it, the worse some of the little quirks happened*".

Technical difficulties further compounded these frustrations and led to uncertainty about whether faults stemmed from the system or user actions. **P1** asked us if "*someone else had this issue*", and later added that the issues he experienced were his fault: "*Probably, probably means that I didn't set it properly*". Such moments affected confidence and disrupted routine viewing, particularly when errors occurred mid-programme or when recovery required repeated steps. For some, these experiences challenged persistence with the prototype. **P6** reported near abandonment after struggling to proceed without support: "*I nearly gave up. I was trying to phone up somebody... Not clear. Nearly gave up.*".

Emotional Highs and Lows. Participants' experiences with the accessibility features involved strong emotional responses, ranging from pride and satisfaction

[3]This issue was fixed for **P3** onwards

to frustration and near abandonment. Successful operation of the features brought a sense of accomplishment and confidence: "***P3**: I'm very, very... I was so surprised [...] It was nice. It was very nice. I'm proud of myself*". For **P4**, effective use of the features deepened engagement with the programme and supported a meaningful interaction with her grandson: "*And then [grandson] come in and sat with me and we were talking about what we was watching [...] Yeah, and I, he, he, uh... just sat there and listened to what I was saying, and that, that, that made me happy*".

When difficulties arose, participants reported frustration and uncertainty: "***P3**: when... first started, it wasn't getting to my head*". **P4** described being unable to resume use after a problem and becoming "*very frustrated, because we couldn't get back on it*". In some cases, repeated failures undermined persistence with the system: "***P6**: Not clear. I nearly gave up [...] Yes. Very frustrated... I couldn't get it worked.*". These accounts show how emotional reactions accompanied both successful and unsuccessful encounters with content adaptation, with confidence building when adjustments worked and frustration mounting when recovery from errors was unclear or slow. They also indicate that emotional responses were closely tied to the immediate viewing context, such as whether a viewer could continue without interruption or discuss the programme with a family member during use, and to the perceived effort required to operate or repair the features mid-session.

Social Negotiating in Shared Viewing

This theme describes how shared viewing contexts influenced the use of content adaptation, the negotiation of feature use, and the role of communication.

Dependency on Viewing Partner. Participants varied in how easily they operated the accessibility features and in whether they needed support. Several asked others to help with setup or with steps they found difficult. **P4** described a difference in confidence between everyday TV use and anything beyond that, and compared herself to family members who operated devices for her: "*I know how to turn the TV on, uh... because I know what button I press. But to, um... get... go... go to any... anything that's not on the normal TV, I can't do, because I don't know how to do it, how* *to do it [...] because they... they got the controller, and they know how to turn it on. Even my grandchildren know how to put all different things on the, uh... on the, um... the, what's it called? Um, channels*". In line with this, **P4** chose not to operate the system herself and asked her husband to handle it. He summarised this arrangement: "***P4's husband****: **P4** and technology, if she'd been on her own, she wouldn't have known what to do, or when to do it*". Others sought help at specific moments. **P3** recalled calling to resolve a problem when the picture changed: "*On Friday, last week, it came dark. And I said, oh my God, where can... What can I do? And then you told me? And then she... He told me*".

Independence was also evident once people became familiar with the steps. Some started with assisted setup, then moved to operating the system on their own after repeating the same actions during viewing. **P6** first relied on his wife to get the system ready so he could begin watching, then used it unaided on a later day. As his wife noted, "*The last time I wasn't even at home and I walked in and you had the program going. Um, so you had managed perfectly well on your own by then*".

Negotiating Collaborative Viewing Adaption. Watching with a partner shaped how participants used accessibility features, with settings discussed to meet one viewer's needs while keeping viewing comfortable for others. This was clear when partners differed on playback speed. **P4**'s husband expressed a preference for a higher value than the one set during viewing: "*I think if **P4** had been at* 85% *rather than* 65%, *then it would have been more comfortable for me as well. But, you know, as I say, this was an experiment to see what... What would suit **P4**, not for me*". Adjustments intended to support comprehension for one person could alter the experience for the co-viewer, so pairs often kept a single configuration through an episode rather than switching repeatedly.

Participants also inferred their partners' reactions without explicit discussion. **P1** treated a lack of comment as acceptance: "*She... She didn't make comments. I... I... Take it as good*". Others chose settings that minimised disruption to family routines. **P8** avoided text simplification when watching with his children to prevent confusion during conversation: "*With kids... wouldn't simplify [captions]... they're still learning... when interrupts them... causes confusion*". In such cases, viewers prioritised a stable configuration that everyone could follow, even if different settings might have supported the primary viewer more effectively when watching alone. Choices about pace, subtitles, or audio mix therefore reflected both individual access needs and the immediate social setting, with partners weighing comprehension against flow of conversation and comfort over the duration of the programme.

Communication Challenges Affecting Social Viewing. Negotiating accessibility feature use during shared viewing was complicated by communication challenges linked to aphasia, with **P5** stating that "*speaking is very hard*". Participants described needing extra time to express preferences so that others could follow what was being requested: "***P3**: You've got to be patient [...] because, me I... I can understand it you know... slowly by slowly or slowly... to... to... understand what... what you mean*". Viewers often paused discussion to allow for turn taking, or waited until a quiet moment before proposing a change to settings.

Social dynamics in group conversations added further difficulty. **P6** reported that noise and overlapping talk made it hard to participate, stating "*I need silence because people wouldn't do anything, I can't get a word in*". Communication during shared viewing could introduce stress when participants felt rushed or were talked over. **P8** described this frustration: "*I need... to be heard... time... to construct my thought and speak out... don't like people... talking over*". **P8** also noted pressures when watching with young children, where expectations to answer questions or entertain reduced focus on the content and made adjustments harder to discuss: "*The moment... what... with the kids, don't focus and too much questions*". These accounts show how timing, noise, and competing demands shaped the ability to articulate preferences and to operate features during shared viewing without losing the thread of conversation.

7.4 Discussion

Participants valued adaptable content, yet two challenges were evident. First, lowering the effort of making adjustments during viewing. Second, managing feature use when watching with others. These challenges distinguish *customisation*, where viewers manually adjust existing content, from *personalisation*, where viewer or system choices produce adapted versions of the programme. This discussion examines how systems can start with simple, low effort defaults and develop into richer personalisation without interrupting playback, as well as how adaptations are decided in shared viewing, with attention to negotiation with co-viewers. The focus is on control design, interaction latency, and the social conditions that shape whether adjustments persist during ordinary viewing.

7.4.1 From 'Plug-and-Play' to Personalisation

The first challenge concerns the effort required to personalise viewing. Participants wanted systems that were immediately usable, described by **P4** as plug-and-play, yet able to grow with the viewer over time. This tension between simplicity and flexibility suggests a route for personalisation that begins with effortless defaults, then offers simple automation, and later supports more advanced adaptation, where early interactions should feel easy and rewarding to encourage repeated use. This is supported by the EMA ratings showing a small upward trend, indicating that familiarity improved with continued use. Defaults that just work at first launch can reduce the cognitive load of initial encounters, particularly for viewers hesitant to engage with unfamiliar technology. Clear, context-aware defaults can increase the likelihood that personalisation is adopted without removing choice [49]. Defaults should be paired with straightforward opt-out so that viewers are not locked into settings they do not want. Automatically lowering less important audio during dialogue is one such default that provides immediate benefit while keeping control in the viewer's hands. As seen in our usage logs, the *multi-track volume controls* were active for 92.7% of the watch time with , which indicates that audio balance was a routine adjustment. Defaults act as a starting point rather than an end state. As viewers assess the accessibility features, they adjust settings to reach a comfortable configuration for their content and context. For some, this involves minor changes at the start of a session, while for others, adjustments occur in response to specific scenes. In both cases, defaults reduce early friction while leaving room for personal judgement. Therefore, where defaults do not fit, simple manual controls should allow for quick corrections without disrupting viewing, thus reducing the risk of reverting to familiar but inaccessible routines.

Automation can reduce repeated work, yet participants preferred manual control when it provided high controllability. Automated adaptations were useful only when predictable, transparent, and easy to reverse, consistent with findings on controllability [258]. Suggestions or temporary adjustments help, but viewers need to correct or undo decisions without breaking their flow. Slowing a fast segment was helpful only if normal playback could be restored immediately once pace relaxed, otherwise automation risked worsening the viewing experience. This idea is supported by our participants preferring to use manual pace controls over the automated ones, as per the usage logs. As familiarity increased, some participants asked for finer-grained controls, while others preferred to keep interaction simple. Several of our participants requested additional control over the features, such as **P1** proposing a spotlight intensity slider and **P5** asking for larger or smaller caption text, which others had an opposite view. Showing the complexity of interaction too early risks alienating viewers who describe themselves as not technical [82, 114, 212, 229]. Thus, progressive disclosure provides a pathway through this tension [88, 156], starting with a small set of visible adjustments that map to common situations, and revealing more advanced options over time once basic control feels stable. This staged approach can prevent overload and build confidence, potentially reducing abandonment by demonstrating the payoff of personalisation as routines form [282]. Mixed-initiative prompts extend this by offering context-aware suggestions while keeping the final decision with the viewer. Prompts should be lightweight, clearly justified, and reversible, so trials feel safe and viewers can return to their prior state at any time. Metadata within flexible media formats can support such prompts by labelling the narrative importance of audio elements, enabling recommendations that fit common viewing aims without forcing change. Local models that learn from on-device behaviour can further tailor prompts while preserving privacy [197, 313].

Interaction latency was a recurring barrier. Having to pause and navigate away from content to reach controls disrupted viewing and discouraged adjustment in the moment. Our participants described friction from leaving the video to reach the controls page as being a significant challenge, such as **P1** reporting repeated navigation as frustrating, and **P3** outlining the back-and-forth steps of pausing, switching screens, and resuming. Scaffolding strategies address this by enabling lightweight, real-time adjustments at natural transition points [49]. Controls should remain available in-player rather than behind separate screens where feasible, with direct manipulation that is legible at a glance. Any automated suggestion must state why it is proposed and what it will do, with reversibility being essential. For people with aphasia, experiencing an adaptation is often more effective than imagining its impact [225, 320], so brief, reversible trials help viewers judge accessibility feature settings without committing to a change that may not work in context. We see this in practice, as in-session tweaks were common.

7.4.2 Socially Negotiated Personalisation

Accessibility in shared viewing operates as a relational and collective activity rather than an isolated one, with co-viewers producing access through everyday coordination. When features adapt the underlying content, the question is not only whether those features exist, but how they are taken up and sustained socially. Negotiation is therefore a central accessibility practice rather than a flaw to conceal, because personalisation is co-created and interdependent, and one participant's adjustment can reshape the experience for everyone else [28, 35, 270]. Some participants described deferring to settings that suited their co-viewers, since negotiated interdependence brings emotional demands that influence whether our participants try, keep, or discard adaptations with others present. Several participants reported feeling self-conscious about making changes and worrying about burdening co-viewers. These feelings could outweigh the benefit of using a feature, which led to quiet compromise even when that compromise reduced comprehension, viewing co-viewers silence as acceptance of the viewing environment. For some participants, successful use was associated with confidence and pride, whereas unclear recovery led others towards frustration and near abandonment. These emotional responses shape accessibility practices in domestic settings, because they influence persistence alongside usability.

Two factors intensified the burden of negotiation during ordinary viewing. First, aphasia complicated moment-to-moment communication – stating a preference, asking for a change, or repairing a misunderstanding needed time and patience, which is hard to secure while a programme runs. Our participants described difficulty speaking, asked others to slow the conversation to allow understanding and response, and reported that overlapping talk prevented participation. Second, adjustment carried a high interaction cost. Moving away from the video to a separate controls page, then returning, broke the flow of watching and promoted a passive default, with participants describing the repeated navigation as frustrating. Some participants chose to abandon helpful features when many or slow steps interrupted their viewing, which is consistent with evidence that high setup or adjustment costs suppress adoption and persistence [233].

Recognising negotiation as part of the viewing routine reframes design priorities around lowering both communication load and interaction load so that accessible choices do not become the demanding ones. Normalising accessibility when designing for shared viewing also makes it easier to state needs without stigma. Prior work reports that people with disabilities often avoid assistive technologies in social settings to avoid being perceived as different [272], with internalised ableist expectations fostering self-doubt, fear of burdening others, and withdrawal from seeking improved access [47, 140]. Concealment can further reinforce the idea that access is exceptional. A practical path forward is to place accessibility controls alongside standard media controls and refer to them using neutral language rather than special labels [271, 272]. This aligns with universal design approaches that build features for broad use rather than marking them as exceptional [235].

Work on aspirational framing demonstrates that presenting adjustments as desirable choices can reduce stigma, and that positioning devices as premium or fashionable rather than medical can make them more socially acceptable [66, 69, 235]. This framing, however, introduces a tension around visibility. If access becomes too implicit, the underlying need can become less visible to others [69]. Designs should therefore make the presence and effect of adaptations legible enough to support shared understanding without labelling the participant as different. Interfaces that show what is currently active can provide common ground for coordination since everyone starts from the same information [84]. Our participants favoured controls whose movement and effect were obvious, and mid-session tweaks to pace or audio were common when the link between action and outcome was easy to see. Reversible trials also help in negotiation. A short trial does not imply commitment, offering a clear path back if the change proves unhelpful, which reduces decision anxiety and supports quick repair [84].

Lowering interaction latency is central to sustaining adaptations during co-viewing. Participants wanted to adjust while watching, which requires fast, in-flow, easily reversible controls. Placing controls off-screen and across multiple steps led to abandonment, which aligns with evidence that difficulty reconfiguring on the fly reduces use [231]. Several participants described reluctance to leave the video to adjust the accessibility settings, asking to operate controls while viewing. Designs should provide direct manipulation in the player where feasible, keep mappings between action and effect obvious, and limit steps so that adjustments can fit natural breaks in conversation and content. Reducing latency also has a social payoff because slower adjustments consume group attention and increase pressure on the person requesting change.

Equity in shared viewing involves distributing effort fairly across participants. Groupware research warns that one person can end up doing most of the work for a group's gain [110], and assigning all interaction effort to a participant with aphasia mirrors the invisible labour many disabled people perform to participate in everyday life [36]. A more equitable practice begins from the access needs of the person most at risk of exclusion, then invites co-viewers to accept or soften specific elements of the adaptation so that the personal burden reduces. The system should invite this behaviour by design. Consensus prompts can propose rather than impose, allowing co-viewers to accept, defer, or suggest an alternative without undermining baseline personalisation [110]. Such brief proposals that explain a suggested change and offer a quick decision would fit our participants preferences.

Personal rather than global adjustments can contain disagreement when preferences diverge, and hardware can help here. Acoustically transparent bone-conducting headsets can provide personalised audio while preserving situational awareness and social inclusion for the wearer [188]. Since such devices do not alter room audio, they limit impact on co-viewers and reduce pressure to revert to a shared baseline that under-serves the primary participant. When co-viewers preferred different mixes or speeds, a personal audio channel would have reduced friction while conversation continued.

7.4.3 Limitations and Future Work

This study has several limitations that should guide future work. The cohort was small and not fully diverse, with only 2 women, an age profile tilted toward older adults, and participation drawn from so-called 'Western, Educated, Industrialised, Rich and Democratic' countries [172]. These characteristics reflect practical difficulties in recruiting people who have aphasia discussed in previous chapters, with the addition of being a two week in-home deployments. Screening further narrowed the pool because we excluded individuals who did not report difficulty watching video, which focused the work while omitting certain voices. Future studies should widen recruitment and include groups with other complex communication needs, as similar personalisation features may be relevant beyond aphasia. Additionally, the insights into shared viewing experiences came from a subset of participants who used the system with others. This points to the need for studies that foreground co-viewing and examine how adaptations are discussed and sustained in households with varied relationships.

The prototype was intentionally minimal and functioned as a probe rather than a complete product. Subsequent iterations should incorporate design directions identified in the discussion, including progressive disclosure, adjustments that can be made within the flow of watching, and transparency about active states, and should expand the feature set to support richer personalisation and appropriation over time. Following the deployment, the control navigation frictions were fixed in the prototype by allowing the viewer to have smaller video player visible while adjusting settings, thus allowing them to see the changes they are making in real time before entering the full screen viewing experience – see Figure 7.8. The 2-week deployment provided situated accounts of everyday use but could not capture long-term adoption or abandonment. Longer and less researcher intensive deployments, possibly through releasing the system as an open tool, would enable study of sustained engagement and evolving practices in domestic contexts.

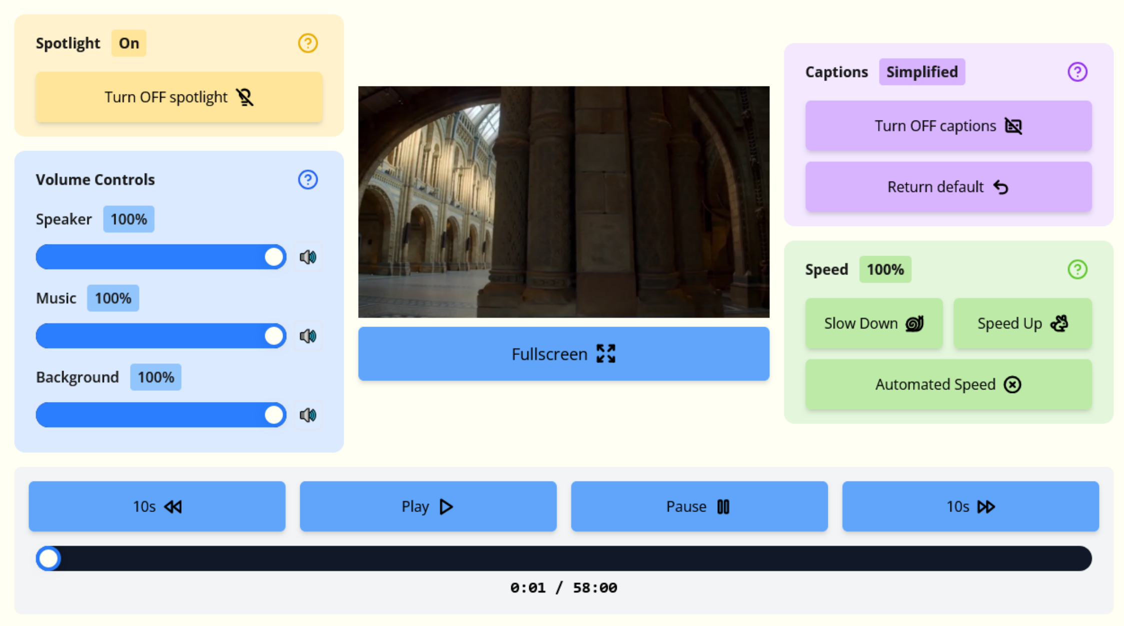

Figure 7.8: Screenshots of the updated deployment system. The video is now visible in the middle of the page, flanked on both side by feature controls.

7.5 Conclusion

This chapter examined how people with aphasia use personalised accessibility features during everyday viewing at home, focusing on what features they adopt, when those features help, and what makes use difficult. A two-week deployment with nine participants and a custom video player exposing four features provided situated evidence about usage patterns, interaction costs, and social influences on uptake. Rather than treating access as the mere presence of features, the analysis frames access as the ability to keep up with content under routine viewing, documenting where adjustments reduced effort and where they introduced new burdens.

Participants most valued the sound rebalancing, reducing competition between audio streams, with multi-track volume controls being the most used and rated as most helpful by participants. Viewers commonly lowered background and music levels to prioritise speech, reporting improved dialogue comprehension without needing constant interaction. Playback speed adjustment was used and sometimes appreciated, but was rarely essential. Simplified captions and speaker spotlight produced mixed responses and were often disabled, either because reading remained too effortful at pace or because the visual overlay distracted from the scene. These patterns place moment-to-moment intelligibility at the centre of effective support.

Adoption followed a recognisable sequence. Participants first explored each feature to learn what it did, often on a clip-by-clip basis, and then narrowed to a subset that fit their needs and habits. Some set stable configurations and left them unchanged; others adjusted frequently; a third group converged over time on settings that worked across programmes. Prior habits shaped these choices, with viewers who regularly used captions elsewhere tended to keep them enabled, while those who found on-screen reading difficult avoided captions and relied on audio changes. This curation by context shows that usefulness depends on both the content and established viewing practices.

Interaction latency emerged as a key barrier, because leaving the video to change settings made viewing feel interrupted and discouraged experimentation, even when adjustments were beneficial. Participants wanted controls that operate in place so changes can be made while watching. Simple layouts with immediate feedback encouraged exploration, whereas slow or unclear steps led people to abandon features that might otherwise help. Although confusion about controls decreased with practice, early sessions imposed a cognitive cost that some participants found discouraging. When settings reduced effort, viewers reported greater comfort and confidence, described comprehending sequences they would previously have missed. When behaviour felt unpredictable or controls were opaque, frustration rose quickly and some considered abandoning the system. Adoption therefore depended on early, low-effort successes and on visible, reversible changes that allow recovery when an adjustment does not help.

Use was also shaped by co-viewing. Family members often explained features or handled setup, reducing the need for some tools while increasing reliance on others for activation. Participants sometimes avoided pausing or changing settings to avoid imposing on others, even when doing so would have improved comprehension. This can be seen as requiring high communication load: asking for changes could be difficult when speech was effortful or in group settings. Personalisation in the home was therefore negotiated, with decisions shaped by habits, roles, and expectations as much as by immediate needs.

The results indicate where personalised adjustments help and what makes them usable at home. Reducing background and music levels relative to speech lowered effort across content types. Speed changes helped in specific moments, but were content dependent. We suggest that clear defaults could be accept quickly, with the option to refine later, since our participants wanted transparency about what a feature was doing and how to undo it. On the other hand, the deployment also clarifies where failures arise, with breakdowns caused by uncertainty about whether features are working or by noticing discrepancies. These breakdowns carried costs beyond the moment, because early struggles reinforced beliefs that new tools were 'not for them', especially when difficulties were attributed to personal ability rather than design.

Systems therefore need low-effort entry with sensible defaults that address common barriers such as speech masking, while still allowing fine control, without forcing it upfront. In-flow adjustment reduces the cost of trying changes and supports brief interventions at moments of high demand. Making system states and effects legible to all viewers supports co-viewing without adding communication burden. Where alternative versions are available, they should remain recognisable as the same work and be selectable when demand increases, rather than acting as persistent overlays that reshape every scene. Advances in flexible media and lightweight on-device machine-learning models make such adjustments more feasible, while keeping control local.

The contribution of this chapter is an account of personalised accessibility in the home, grounded in what people actually used and how they decided to keep or discard options. It shows that access for people with aphasia can be supported by small changes that alleviate cognitive load at the right moment. It also shows that personalisation carries costs when interaction is slow or when social coordination is difficult to initiate. Together, these findings identify concrete targets for reducing effort during viewing and for designing controls that are easy to try, easy to reverse, and acceptable in shared settings.

Chapter 8

Conclusion

This thesis has examined audiovisual accessibility through the lens of comprehension during viewing, shifting away from a focus on perceptibility or the presence of access services. Within research and practice, accessibility is often operationalised as the provision of features, with quality assessed through accuracy, completeness, or compliance with standards and regulatory requirements such as WCAG or broadcast guidelines [153, 301]. This framing has supported advances in provision, yet it often treats comprehension as a consequence of perceptual access, not as a design concern in its own right, particularly for content that unfolds over time. The work presented here departs from this orientation by treating comprehension as the outcome of interest. Across the thesis, difficulties in understanding arise even when standard services are available. These difficulties extend beyond missing dialogue or visual detail to include breakdowns in maintaining orientation during viewing. This aligns with critiques of service based accessibility that describe access as appended to media, with limited attention to how meaning is constructed through viewing experience [78, 79].

The thesis began by establishing a field-level baseline through a systematic review of technical accessibility interventions for digital audiovisual media. That review found that interventions and evaluation practices are concentrated on sensory access, which aligns with broader surveys of accessibility research that report persistent under-representation of several disabled communities and needs beyond sensory needs [181]. Against this baseline, the thesis treats comprehension during viewing as a first-order accessibility concern for audiences whose main barriers relate to language processing under time pressure. The thesis follows a research through design (RtD) approach [100, 331], using successive intervention artefacts to examine how changes to audiovisual presentation affect comprehension during viewing. This approach does not aim to validate a single intervention as a general solution. It develops a design space by moving between accounts of breakdown, prototype making, and evaluation in contexts of use. The studies function as probes into this space, attending to how comprehension difficulties are experienced, then exploring how changes to audiovisual presentation might address them in domestic viewing. Taken together, these studies support a reframing of accessibility as an ongoing, situated process that emerges through interaction between media, viewers, and viewing contexts, not as a fixed property of content or services alone.

8.1 Accessibility Through Comprehension

Treating comprehension as the primary accessibility concern shifts attention away from isolated failures of perception toward the conditions under which audiovisual media becomes difficult to follow. Difficulties in understanding arise even when dialogue and visual information are technically available. In this thesis, they are closely tied to how quickly information is presented, which is consistent with evidence that higher subtitle speed reduces opportunities to reread and is associated with lower comprehension during multimodal viewing [158]. Prior work on audiovisual media has noted that time constrained presentation introduces additional cognitive load beyond that associated with static media [51, 277]. Here, such demands recur as a source of breakdown for viewers with complex communication needs, particularly when content progresses faster than viewers can process and regain orientation.

The survey and focus group studies provide an initial view into how these breakdowns are experienced in practice (see Chapter 4). Participants describe moments where they lose track of speakers or referents despite being able to access picture and sound. These accounts point to a gap between perceptual access and understanding, with commonplace interventions failing to guarantee that viewers can integrate information into a coherent account of events as content progresses. Similar concerns have been raised in accessibility scholarship that critiques compliance driven approaches for prioritising feature presence over experiential usability, particularly in dynamic media contexts [78, 164].

Putting comprehension first reframes accessibility problems as temporal rather than static. Understanding depends on maintaining orientation and regaining it when it has been lost, not on uninterrupted access to individual elements. This perspective aligns with critiques of retrofit models of accessibility, which position services as additions to a finished artefact, with limited capacity to affect how meaning is structured over time [251]. As discussed in the thesis, when accessibility is treated primarily as a service delivered after production, design attention is less likely to address pacing or information structure.

Across the empirical work, moments of confusion were organised into a taxonomy of barriers and facilitators that links comprehension difficulties to properties of audiovisual presentation and to recovery during viewing. This taxonomy bridges description of breakdowns with design exploration by specifying where changes to presentation could plausibly reduce effort at the point of use. If comprehension difficulties arise from how information is organised and presented, then interventions need to act on these properties, not only translate content across modalities.

The participatory design work with people with aphasia makes this shift explicit (see Chapter 5). Participants did not primarily request extra explanatory layers. They identified moments where the content itself introduced barriers, then proposed changes to presentation that would reduce effort in those moments. This led to design directions that focused on adapting how content is staged, including changes to timing and reductions in competing visual detail.

The bespoke design work extended these ideas by exploring how localised adaptations could support comprehension (see Chapter 6). Rather than introducing parallel explanatory content, the designs intervened in the presentation of existing audiovisual material in ways that matched the needs of the co-designers. These interventions were limited in scope and reversible, reflecting participants preferences for control over when support is applied. In the thesis, this emphasis on controllability responds to a recurring concern that changes are harder to use when the system behaviour is difficult to anticipate.

The in-home deployments further situate comprehension-oriented accessibility within everyday viewing practices (see Chapter 7). In domestic settings, comprehension support was shaped by routines and shared arrangements. Some adaptations were used at points of difficulty. Others were abandoned because they did not fit viewing practices, even when participants could describe potential benefits. These observations reinforce that comprehension support is negotiated within shared viewing, which is consistent with accounts of accessibility as something that emerges through coordination with other people, not only through individual interaction with a device [28]. Support that does not account for co-viewing negotiation or preferences for low disruption may be technically effective yet unused. Across these studies, accessibility through comprehension emerges as a matter of fitting audiovisual presentation to viewers processing needs within specific contexts, not ensuring access to all information at all times. This perspective does not reject existing services, but it exposes limits when understanding is treated as a downstream effect rather than a design objective.

8.2 Exploratory Design Space

The contribution of this thesis lies in mapping a design space for comprehension-first accessibility for people with aphasia through a sequence of design interventions. RtD provides the method for this mapping, with each intervention artefact used to make specific presentation changes concrete, then to study how those changes affect comprehension during viewing. The work does not converge on a single system or workflow. It investigates how modifying presentation can alter viewers' comprehension. This treats comprehension as a design target. It also makes aspects of the problem space visible, including constraints in interaction, social viewing, and production practice.

A central role of the early studies is to clarify the conditions under which viewers with aphasia identify breakdowns in understanding. Instead of establishing prevalence or isolating causal mechanisms, these studies surface the kinds of situations that participants recognise as problematic during everyday viewing. Within the thesis, this material informs later design exploration by grounding it in what participants notice and prioritise. Across stages, the thesis relies on a consistent set of comprehension-oriented outcomes suited to ongoing playback. These include perceived effort and the ability to regain orientation after missed information. In the in-home work, these outcomes are captured in-the-moment to reflect fluctuations in comprehension during domestic viewing, including co-viewing where accessibility choices can require negotiation.

The participatory design work translates reported difficulties into ideas for change by engaging people with aphasia in explaining how content might be altered to better support their own understanding. This work depended on methodological adaptations that support participation despite language impairment, drawing on approaches that use tangible materials to reduce reliance on speech during co-design activities [320]. These adjustments shaped what could be expressed and carried forward into prototypes.

The bespoke design work further develops this exploratory space by materialising selected ideas as working prototypes that modify audiovisual presentation. The purpose of these designs is not to demonstrate optimal solutions, but to examine how specific transformations affect comprehension as content progresses. Through this process, the thesis characterises relationships between presentation choices and the ability to remain oriented during viewing. The in-home deployments extend this exploration into routine contexts, where adaptations intersect with established viewing practices and shared arrangements. Responses to these interventions, including selective use and non-use, help define the contours of the design space, consistent with arguments in HCI that appropriation and non-use can be analytically productive when studying emerging technologies and practices [232, 248].

8.2.1 Production constraints on accessibility

Alongside this thesis, related work was conducted in which I contributed but did not act as lead author, reported in [32]. That study involved interviews and workshops with practitioners working within a national television broadcaster. The aim was to examine how accessibility is produced and constrained across commissioning, production, and post-production, with attention to remit boundaries, handovers, and economic pressures. The findings describe accessibility as infrastructural work that is fragmented across roles and introduced late in production, with limited scope to influence presentation once content is locked [32].

Introducing this production-focused perspective clarifies the status of the exploratory design space mapped in the thesis. The comprehension-oriented interventions explored here often involve presentation changes that are difficult to accommodate within existing workflows. Media professionals describe tensions around creative ownership, schedules, and the feasibility of altering content, while noting that earlier integration of accessibility considerations can enable more substantial change when authority and time allow. Placing these accounts alongside the thesis designs describes a gap between what viewers find helpful for understanding and what current production ecologies readily support. The thesis does not attempt to resolve this gap. It uses design exploration to articulate it by materialising viewer-driven ideas for changing presentation and connecting them to constraints in production practice. As argued in [32], addressing these tensions would require shifts in how accessibility is commissioned and how responsibility for understanding is distributed across the media lifecycle. The exploratory design space mapped in this thesis provides a basis for such discussions, without assuming that the interventions themselves are transferable at scale.

8.3 Implications of Comprehension-First Accessibility

Approaching accessibility through comprehension has implications for how accessibility problems are defined and evaluated. When understanding is treated as the outcome, accessibility concerns extend beyond the presence or accuracy of services to include whether viewers remain oriented as content progresses. These issues can occur even when subtitles or audio description meet formal quality criteria, which aligns with critiques that compliance does not guarantee a usable or meaningful viewing experience [79]. Accessibility quality is therefore tied to how media is experienced over time, not only whether information is technically available.

This perspective also affects how personalisation is understood in accessible media. In this thesis, comprehension support benefits from being controllable at the point of use, since perceived agency over media configuration affects how people engage with personalised systems [289]. Viewers engage with adaptations when they recognise a need and can anticipate what an action will do. Changes without clear indication can introduce uncertainty, which can undermine understanding. In shared viewing situations, this becomes more pronounced because accessibility choices can affect other viewers and require negotiation. These observations position personalisation as an interactional process that unfolds during viewing, shaped by domestic context, rather than as a background system that optimises presentation without user involvement.

The work also has implications for how audiovisual media is represented for adaptation. The interventions explored here assume an object-based representation in which media assets are assembled at playback using metadata, allowing variant presentations to be produced without relying on a single fixed rendition across contexts [17, 130]. Across the bespoke work and the in-home deployments, uptake depended on predictable behaviour and manageable interaction trade-offs, which matches evidence that real-time media adaptation requires structured interaction to support informed choices during viewing [105]. Participants used support when they could anticipate the effect of an action and reverse it quickly. Features that required repeated adjustment or introduced uncertainty were less likely to become part of routine viewing, even when participants described potential benefits.

A comprehension-first perspective also raises questions about where interventions can be situated. The designs explored in this thesis operate at playback, modifying presentation as it is experienced by viewers. This positioning allows adaptations to respond to difficulties during viewing without requiring changes to production.

At the same time, production-focused research with media practitioners highlights limits to this approach. Practitioners describe how accessibility is shaped by remit boundaries, schedules, and ownership of creative decisions, with constraints on altering content once it is locked [32]. From this viewpoint, comprehension-oriented changes to presentation can conflict with established production practices, even when they align with audience needs. Considering these production accounts alongside the thesis designs clarifies that comprehension-first accessibility can create tensions between viewer-centred adaptation and existing workflows. Practitioners identify opportunities for addressing such tensions when accessibility considerations are introduced earlier and responsibilities are assigned in ways that allow changes to be planned [32].

The implications of this thesis do not lie in prescribing how these tensions should be resolved. They lie in making explicit that supporting comprehension may require shifts in how accessibility is integrated, as well as how adaptation is treated as a legitimate response to audience needs. This approach does not displace existing services, instead adding a concern with understanding during viewing and with the conditions under which viewers can remain oriented through moments of difficulty.

8.4 Scope and Limitations

The work presented in this thesis is exploratory in scope and does not aim to produce generalisable findings. The studies focus on a small number of participants and settings to examine how comprehension difficulties are experienced and how they can be addressed through changes to audiovisual presentation. This reflects an emphasis on depth of engagement and situated understanding. The findings should therefore be read as characterising a design space and a set of concerns, not as establishing prevalence or universal patterns of use.

This stance is also reflected in how the thesis treats scalability. Work on non-scalability describes growth that preserves existing methods, which can leave little space for cases that require reworking assumptions about how media is authored and delivered [296]. Accounts of the politics of scaling describe how demands for scale can narrow problem definitions in ways that tend to sideline minority groups [232]. In other domains, appeals to scale have been used to justify designs that many people cannot use, including transport systems that prioritise a particular kind of mobility while treating other journeys as residual [217]. This thesis therefore treats bespoke outcomes as exemplars that specify recurring barriers and corresponding presentation changes, consistent with design-for-one accounts in which uniqueness is used to derive systematic requirements rather than averaged away [118]. It also treats early, participant-led prototypes as sites where defaults can be contested before they are stabilised by platform and workflow constraints [123]. The thesis does not claim that a single solution should scale across all viewers. It points to an approach in which platforms accumulate supported cases grounded in real viewing, each with defined limits of operation and a known effect on comprehension [232].

These choices sit alongside other boundaries that shape the thesis scope. The design work is grounded in collaboration with people with aphasia and focuses on recorded content viewed in domestic settings. The thesis does not address all forms of disability or all media formats. These boundaries reflect the intention to examine comprehension in depth within a specific context.

Finally, the thesis does not evaluate feasibility within existing production or regulatory frameworks. While related work with practitioners highlights tensions between comprehension-oriented adaptation and current workflows, the designs explored here are not presented as ready for integration into broadcast or film pipelines. The focus remains on how changes to presentation affect viewing experience, not on cost or compliance.

8.5 Directions for Future Work

This thesis suggests directions for future research that build on its exploratory character. One direction is to extend comprehension-first accessibility beyond the specific population and viewing contexts examined here by working with other disabled communities for whom following audiovisual content is difficult for reasons tied to language processing, cognition, or fluctuating ability. Such work would support a clearer account of how comprehension challenges vary across viewers and situations, including how they are described in relation to everyday media use.

Future work can also broaden the set of stakeholders involved in design exploration by including practitioners in production and distribution. Working with these groups can clarify how comprehension difficulties are recognised from different positions, how responsibility for addressing them is allocated, and how proposed adaptations intersect with creative constraints. This would also support analysis of which kinds of presentation change are treated as acceptable within current workflows and which would require changes to commissioning, documentation, or tool support.

A further direction is to develop a more systematic account of possible interventions by relating recurring comprehension difficulties to families of adaptation strategies and documenting when each family is useful. This could take the form of a repository grounded in viewer experience and design practice, intended to support comparative analysis across genres and contexts.

Finally, future work can examine how comprehension-first adaptations change over time through longer deployments that capture learning, routine formation, and changes in need. Comparative work across platforms and household arrangements can further clarify how context shapes uptake and non-use, particularly where viewing is shared and accessibility choices remain subject to negotiation.

Bibliography

[10] Tania Acosta, Patricia Acosta-Vargas, Jose Zambrano-Miranda, and Sergio Lujan-Mora. 2020. Web Accessibility Evaluation of Videos Published on YouTube by Worldwide Top-Ranking Universities. *IEEE Access* 8 (2020), 110994–111011. https://doi.org/10.1109/access.2020.3002175

[11] Diana Affi, Joel Dumoulin, Marco Bertini, Elena Mugellini, Omar Abou Khaled, and Alberto Del Bimbo. 2015. SensiTV: Smart EmotioNal System for Impaired People's TV. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Brussels, Belgium) *(TVX '15)*. Association for Computing Machinery, New York, NY, USA, 125–130. https://doi.org/10.1145/2745197.2755512

[12] Chanchal Agrawal and Roshan L Peiris. 2021. I See What You're Saying: A Literature Review of Eye Tracking Research in Communication of Deaf or Hard of Hearing Users. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 41, 13 pages. https://doi.org/10.1145/3441852.3471209

[13] Areen AL-Zoubi. 2024. *Digital Technology and Changes in Media Consumption: A Case Study of Smartphone and App Usage*. Springer Nature Switzerland, Switzerland, 433–444. https://doi.org/10.1007/978-3-031-54019-6_39

[14] Meryl Alper. 2021. Critical media access studies: Deconstructing power, visibility, and marginality in mediated space. *International Journal of Communication* 15 (2021), 22.

[15] Anwar Alsalamah. 2020. Using Captioning Services With Deaf and Hard of Hearing Students in Higher Education: A Systematic Review. *American Annals of the Deaf* 165, 1 (2020), 114–127. https://doi.org/10.1353/aad.2020.0012

[16] Akher Al Amin, Joseph Mendis, Raja Kushalnagar, Christian Vogler, Sooyeon Lee, and Matt Huenerfauth. 2022. Deaf and Hard of Hearing Viewers' Preference for Speaker Identifier Type in Live TV Programming. In *Universal Access in Human-Computer Interaction. Novel Design Approaches and Technologies*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, Switzerland, 200–211. https://doi.org/10.1007/978-3-031-05028-2_13

[17] M. Armstrong, A. Churnside, M.E.F. Melchior, M. Shotton, and M. Brooks. 2014. Object-Based Broadcasting - Curation, Responsiveness and User Experience. In *International Broadcasting Convention (IBC) 2014 Conference*. Institution of Engineering and Technology, Salford, United Kingdom, 8 pages. https://doi.org/10.1049/ib.2014.0038

[18] Michael Armstrong and Maxine Glancy. 2023. *Frameworks for understanding personalisation*. Technical Report WHP 404. BBC, UK. 51 pages. https://www.bbc.co.uk/rd/publications/research-personalisation-recommendation-discovery

[19] Ali Selman Aydin, Shirin Feiz, Vikas Ashok, and IV Ramakrishnan. 2020. Towards Making Videos Accessible for Low Vision Screen Magnifier Users. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 10–21. https://doi.org/10.1145/3377325.3377494

[20] Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Enhancing independence and safety for blind and deaf-blind public transit riders. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI*

'11). Association for Computing Machinery, New York, NY, USA, 3247–3256. https://doi.org/10.1145/1978942.1979424

[21] Liam Bannon, Jeffrey Bardzell, and Susanne Bodker. 2018. Introduction: Reimagining Participatory Design – Emerging Voices. *ACM Transactions on Computer-Human Interaction* 25, 1 (feb 2018), 1–8. https://doi.org/10.1145/3177794

[22] Cynthia L. Bartlett and Gail V. Pashek. 1994. Taxonomic theory and practical implications in aphasia classification. *Aphasiology* 8, 2 (mar 1994), 103–126. https://doi.org/10.1080/02687039408248645

[23] Gokce Elif Baykal, Maarten Van Mechelen, and Eva Eriksson. 2020. Collaborative Technologies for Children with Special Needs: A Systematic Literature Review. 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, 1–13. https://doi.org/10.1145/3313831.3376291

[24] BBC. 2024. How do I create subtitles? https://www.bbc.co.uk/academy-guides/how-do-i-create-subtitles

[25] BBC. 2025. Performance against public commitments. , 48 pages. https://www.bbc.co.uk/aboutthebbc/documents/ara-24-25-data-packs-and-additional-information.pdf

[26] Bernd Benecke. 2004. Audio-Description. *Meta* 49, 1 (Sept. 2004), 78–80. https://doi.org/10.7202/009022ar

[27] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics. 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, 1–13. https://doi.org/10.1145/3313831.3376344

[28] 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. https://doi.org/10.1145/3234695.3236348

[29] Karianne Berg, Jytte Isaksen, Sarah J. Wallace, Madeline Cruice, Nina Simmons-Mackie, and Linda Worrall. 2020. Establishing consensus on a definition of aphasia: an e-Delphi study of international aphasia researchers. *Aphasiology* 36, 4 (dec 2020), 385–400. https://doi.org/10.1080/02687038.2020.1852003

[30] Alexy Bhowmick and Shyamanta M. Hazarika. 2017. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces* 11, 2 (jan 2017), 149–172. https://doi.org/10.1007/s12193-016-0235-6

[31] Filip Bircanin, Margot Brereton, Laurianne Sitbon, Bernd Ploderer, Andrew Azaabanye Bayor, and Stewart Koplick. 2021. Including Adults with Severe Intellectual Disabilities in Co-Design through Active Support. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 486, 12 pages. https://doi.org/10.1145/3411764.3445057

[32] Filip Bircanin, Alexandre Nevsky, Madeline Cruice, and Timothy Neate. 2026. "You meet the audience at where they're at, not where you want them to be": Rethinking Media Accessibility and Design with Media Practitioners. , 18 pages. Awaiting peer-review..

[33] Filip Bircanin, Alexandre Nevsky, Himaya Perera, Vaasvi Agarwal, Eunyeol Song, Madeline Cruice, and Timothy Neate. 2025. Sounds Accessible: Envisioning Accessible Audio Media Futures with People with Aphasia. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama,

Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3706598.3714000

[34] Eva Brandt, Erling Bjorgvinsson, Per-Anders Hillgren, Viktor Bergqvist, and Marcus Emilson. 2002. PDA's, Barcodes and Video-Films for Continuous Learning at an Intensive Care Unit. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction* (Aarhus, Denmark) *(NordiCHI '02)*. Association for Computing Machinery, New York, NY, USA, 293–294. https://doi.org/10.1145/572020.572070

[35] 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. https://doi.org/10.1145/2702123.2702511

[36] Stacy M. Branham and Shaun K. Kane. 2015. The Invisible Work of Accessibility: How Blind Employees Manage Accessibility in Mixed-Ability Workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 163–171. https://doi.org/10.1145/2700648.2809864

[37] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, USA, 57–71. https://doi.org/10.1037/13620-004

[38] Virginia Braun, Victoria Clarke, Nikki Hayfield, Louise Davey, and Elizabeth Jenkinson. 2022. *Doing Reflexive Thematic Analysis*. Springer International Publishing, New York, NY, USA, 19–38. https://doi.org/10.1007/978-3-031-13942-0_2

[39] Barbara B. Braverman and Melody Hertzog. 1980. The Effects of Caption Rate and Language Level on Comprehension of a Captioned Video Presentation.

American Annals of the Deaf 125, 7 (Oct. 1980), 943–948. https://doi.org/10.1353/aad.2012.1118

[40] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Brussels, Belgium) *(TVX '15)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2745197.2745204

[41] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. 2017. Subtitles in 360-Degree Video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands) *(TVX '17 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 3–8. https://doi.org/10.1145/3084289.3089915

[42] Jessica A. Brown, Kelly Knollman-Porter, Karen Hux, Sarah E. Wallace, and Camille Deville. 2020. Effect of digital highlighting on reading comprehension given text-to-speech technology for people with aphasia. *Aphasiology* 35, 2 (June 2020), 200–221. https://doi.org/10.1080/02687038.2020.1787728

[43] Emeline Brule, Brianna J. Tomlinson, Oussama Metatla, Christophe Jouffrais, and Marcos Serrano. 2020. Review of Quantitative Empirical Evaluations of Technology for People with Visual Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376749

[44] Shelagh M. Brumfitt and Paschal Sheeran. 1999. The development and validation of the Visual Analogue Self-Esteem Scale (VASES). *British Journal of Clinical Psychology* 38, 4 (nov 1999), 387–400. https://doi.org/10.1348/014466599162980

[45] Janine Butler. 2018. Integral Captions and Subtitles: Designing a Space for

Embodied Rhetorics and Visual Access. *Rhetoric Review* 37, 3 (may 2018), 286–299. https://doi.org/10.1080/07350198.2018.1463500

[46] Susanne BÃ¸dker, Christian Dindler, Ole S. Iversen, and Rachel C. Smith. 2022. *Participatory Design*. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-031-02235-7

[47] Fiona A. Kumari Campbell. 2008. Exploring internalized ableism using critical race theory. *Disability and Society* 23, 2 (mar 2008), 151–162. https://doi.org/10.1080/09687590701841190

[48] Virginia Pinto Campos, Luiz Marcos G. Goncalves, and Tiago Maritan U. de Araujo. 2017. Applying audio description for context understanding of surveillance videos by people with visual impairments. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, New York, NY, USA, 1–5. https://doi.org/10.1109/avss.2017.8078530

[49] Ana Caraban, Evangelos Karapanos, Daniel Goncalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300733

[50] Patrick Carrington, Amy Hurst, and Shaun K. Kane. 2014. Wearables and chairables: inclusive design of mobile input and output techniques for power wheelchair users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3103–3112. https://doi.org/10.1145/2556288.2557237

[51] Jade Cartwright and Kym A. E. Elliott. 2009. Promoting strategic television viewing in the context of progressive language impairment. *Aphasiology* 23, 2 (2009), 266–285. https://doi.org/10.1080/02687030801942932

[52] Avner Caspi, Paul Gorsky, and Meira Privman. 2005. Viewing comprehension: Students? learning preferences and strategies when studying from video. *Instructional Science* 33, 1 (jan 2005), 31–47. https://doi.org/10.1007/s11251-004-2576-x

[53] Teresa Chambel, Sergio Neves, Celso Sousa, and Rafael Francisco. 2010. Synesthetic Video: Hearing Colors, Seeing Sounds. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments* (Tampere, Finland) *(MindTrek '10)*. Association for Computing Machinery, New York, NY, USA, 130–133. https://doi.org/10.1145/1930488.1930515

[54] Deborah Chambers. 2021. Emerging temporalities in the multiscreen home. *Media, Culture and Society* 43 (2021), 1180–1196. Issue 7. https://doi.org/10.1177/0163443719867851

[55] Frederic Chaume. 2018. Is audiovisual translation putting the concept of translation up against the ropes? *The Journal of Specialised Translation* 30 (jul 2018), 84–104. https://jostrans.soap2.ch/issue30/art_chaume.php

[56] Frederic Chaume. 2019. *Audiovisual translation in the age of digital transformation* (1st ed.). John Benjamins Publishing Company, Amsterdam, Netherlands, 103–124. https://doi.org/10.1075/btl.148.05cha

[57] Frederic Chaume. 2019. *Localizing media contents: Technological shifts, global and social differences and activism in audiovisual translation* (1st edition ed.). Routledge, New York, NY, USA, 320–331. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315192468-32/localizing-media-contents-frederic-chaume

[58] Craig Cieciura, Maxine Glancy, and Philip J.B. Jackson. 2023. Producing Personalised Object-Based Audio-Visual Experiences: an Ethnographic Study. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (Nantes, France) *(IMX '23)*. Association for Computing Machinery, New York, NY, USA, 71–82. https://doi.org/10.1145/3573381.3596156

[59] Amanda S. Clossen. 2014. Beyond the Letter of the Law: Accessibility, Universal Design, and Human-Centered Design in Video Tutorials. *Pennsylvania Libraries: Research and Practice* 2, 1 (may 2014), 27–37. https://doi.org/10.5195/palrap.2014.43

[60] Shauna Concannon, Natasha Rajan, Parthiv Shah, Davy Smith, Marian Ursu, and Jonathan Hook. 2020. Brooke Leave Home: Designing a Personalized Film to Support Public Engagement with Open Data. 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, 1–14. https://doi.org/10.1145/3313831.3376462

[61] Enrique Costa-Montenegro, Fatima M. Garcia-Doval, Jonathan Juncal-Martinez, and Belen Barragans-Martinez. 2015. SubTitleMe, subtitles in cinemas in mobile devices. *Universal Access in the Information Society* 15, 3 (jun 2015), 461–472. https://doi.org/10.1007/s10209-015-0420-5

[62] Francisco M. Costela, Stephanie M. Reeves, and Russell L. Woods. 2021. The Effect of Zoom Magnification and Large Display on Video Comprehension in Individuals With Central Vision Loss. *Translational Vision Science and Technology* 10, 8 (jul 2021), 30. https://doi.org/10.1167/tvst.10.8.30

[63] Jasmine Cox, Rhianne Jones, Chris Northwood, Jonathan Tutcher, and Ben Robinson. 2017. Object-Based Production. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, New York, NY, USA, 79–80. https://doi.org/10.1145/3084289.3089912

[64] Michael Crabb, Rhianne Jones, Mike Armstrong, and Chris J. Hughes. 2015. Online News Videos: The UX of Subtitle Position. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 215–222. https://doi.org/10.1145/2700648.2809866

[65] Arthur Cropley. 2006. In Praise of Convergent Thinking. *Creativity Research Journal* 18, 3 (2006), 391–404. https://doi.org/10.1207/s15326934crj1803_13

[66] Humphrey Curtis, Adam D G Jenkins, Seray B Ibrahim, and Timothy Neate. 2024. Looking Past Screens: Exploring Mixed Reality and Discreet AAC Devices. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 42, 22 pages. https://doi.org/10.1145/3663548.3675655

[67] Humphrey Curtis and Timothy Neate. 2024. Beyond Repairing with Electronic Speech: Towards Embodied Communication and Assistive Technology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 59, 12 pages. https://doi.org/10.1145/3613904.3642274

[68] Humphrey Curtis, Timothy Neate, and Carlota Vazquez Gonzalez. 2022. State of the Art in AAC: A Systematic Review and Taxonomy. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 1–22. https://doi.org/10.1145/3517428.3544810

[69] Humphrey Curtis, Zihao You, William Deary, Miruna-Ioana Tudoreanu, and Timothy Neate. 2023. Envisioning the (In)Visibility of Discreet and Wearable AAC Devices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 52, 19 pages. https://doi.org/10.1145/3544548.3580936

[70] Ruth Dalemans, Derick T Wade, Wim JA Van den Heuvel, and Luc P De Witte. 2009. Facilitating the participation of people with aphasia in research: a description of strategies. *Clinical rehabilitation* 23, 10 (2009), 948–959. https://doi.org/10.1177/0269215509337197

[71] Ruth J. P. Dalemans, Luc de Witte, Derick Wade, and Wim van den Heuvel. 2010. Social participation through the eyes of people with aphasia. *International Journal of Language and Communication Disorders* 45, 5 (sep 2010), 537–550. https://doi.org/10.3109/13682820903223633

[72] Calua de Lacerda Pataca, Saad Hassan, Lloyd May, Michelle M Olson, Toni D'aurio, Roshan L Peiris, and Matt Huenerfauth. 2025. Tactile Emotions: Multimodal Affective Captioning with Haptics Improves Narrative Engagement for d/Deaf and Hard-of-Hearing Viewers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 68, 17 pages. https://doi.org/10.1145/3706598.3713304

[73] Calua de Lacerda Pataca, Matthew Watkins, Roshan Peiris, Sooyeon Lee, and Matt Huenerfauth. 2023. Visualization of Speech Prosody and Emotion in Captions: Accessibility for Deaf and Hard-of-Hearing Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 831, 15 pages. https://doi.org/10.1145/3544548.3581511

[74] Herie B. de Vries and Todd I. Lubart. 2017. Scientific Creativity: Divergent and Convergent Thinking and the Impact of Culture. *The Journal of Creative Behavior* 53, 2 (dec 2017), 145–155. https://doi.org/10.1002/jocb.184

[75] Willemijn J. Doedens and Lotte Meteyard. 2022. What is Functional Communication? A Theoretical Framework for Real-World Communication Applied to Aphasia Rehabilitation. *Neuropsychology Review* 32, 4 (Jan. 2022), 937–973. https://doi.org/10.1007/s11065-021-09531-2

[76] Patricio Domingues, Ruben Nogueira, Jose Carlos Francisco, and Miguel Frade. 2020. Post-mortem digital forensic artifacts of TikTok Android App. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3407023.3409203

[77] Pelle Ehn. 2008. Participation in design things. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008* (Bloomington, Indiana) *(PDC '08)*. Indiana University, USA, 92–101.

[78] Elizabeth Ellcessor. 2010. Bridging Disability Divides. *Information, Communication and Society* 13, 3 (apr 2010), 289–308. https://doi.org/10.1080/13691180903456546

[79] Elizabeth Ellcessor. 2016. *Restricted Access: Media, Disability, and the Politics of Participation*. NYU Press, New York, NY, USA. http://www.jstor.org/stable/j.ctt18040rg

[80] Charles Ellis, Rose Y. Hardy, Richard C. Lindrooth, and Richard K. Peach. 2017. Rate of aphasia among stroke patients discharged from hospitals in the United States. *Aphasiology* 32, 9 (oct 2017), 1075–1086. https://doi.org/10.1080/02687038.2017.1385052

[81] Katie Ellis. 2019. *Disability and Digital Television Cultures: Representation, Access, and Reception*. Routledge, UK. https://doi.org/10.4324/9781315755663

[82] Kirsten Ellis, Emily Dao, Osian Smith, Stephen Lindsay, and Patrick Olivier. 2021. TapeBlocks: A Making Toolkit for People Living with Intellectual Disabilities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 280, 12 pages. https://doi.org/10.1145/3411764.3445647

[83] Fatemeh Erfanian, Robab Latifnejad Roudsari, Abbas Heydari, and Mohsen Noghani Dokht Bahmani. 2020. A Narrative on Using Vignettes: Its Advantages and Drawbacks. *Journal of Midwifery and Reproductive Health* 8, 2 (2020), 2134–2145. https://doi.org/10.22038/jmrh.2020.41650.1472

[84] Thomas Erickson and Wendy A. Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (mar 2000), 59–83. https://doi.org/10.1145/344949.345004

[85] Sarah Eskens, Natali Helberger, and Judith Moeller. 2017. Challenged by news personalisation: five perspectives on the right to receive information. *Journal of Media Law* 9, 2 (jul 2017), 259–284. https://doi.org/10.1080/17577632.2017.1387353

[86] FCC. 1973. Section 508 of the Rehabilitation Act - 29 U.S.C. 798.

[87] Deborah I. Fels, John Patrick Udo, Jonas E. Diamond, and Jeremy I. Diamond. 2006. A Comparison of Alternative Narrative Approaches to Video Description for Animated Comedy. *Journal of Visual Impairment and Blindness* 100, 5 (may 2006), 295–305. https://doi.org/10.1177/0145482x0610000507

[88] Leah Findlater and Joanna McGrenere. 2004. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 89–96. https://doi.org/10.1145/985692.985704

[89] Leah Findlater, Karyn Moffatt, Jon E. Froehlich, Meethu Malu, and Joan Zhang. 2017. Comparing Touchscreen and Mouse Input Performance by People With and Without Upper Body Motor Impairments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6056–6061. https://doi.org/10.1145/3025453.3025603

[90] John Fiske and John Hartley. 2003. *Bardic Television* (2nd ed.). Routledge, UK, 64–77.

[91] Joseph L. Fleiss and Jacob Cohen. 1973. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement* 33, 3 (oct 1973), 613–619. https://doi.org/10.1177/001316447303300309

[92] Christopher Frauenberger, Judith Good, and Wendy Keay-Bright. 2011. Designing technology for children with special needs: bridging perspectives through participatory design. *CoDesign* 7, 1 (March 2011), 1–28. https://doi.org/10.1080/15710882.2011.587013

[93] Christopher Frauenberger, Julia Makhaeva, and Katta Spiel. 2017. Blending Methods: Developing Participatory Design Sessions for Autistic Children. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 39–49. https://doi.org/10.1145/3078072.3079727

[94] Liam Funnell, Isabel Garriock, Ben Shirley, and Tracey Williamson. 2019. Dementia-friendly design of television news broadcasts. *Journal of Enabling Technologies* 13, 3 (sep 2019), 137–149. https://doi.org/10.1108/jet-02-2018-0009

[95] Julia Galliers, Stephanie Wilson, Abi Roper, Naomi Cocks, Jane Marshall, Sam Muscroft, and Tim Pring. 2012. Words are not enough: empowering people with aphasia in the design process. In *Proceedings of the 12th Participatory Design Conference: Research Papers - Volume 1* (Roskilde, Denmark) *(PDC '12)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/2347635.2347643

[96] Walter Gantz and Lawrence A. Wenner. 1995. Fanship and the Television Sports Viewing Experience. *Sociology of Sport Journal* 12, 1 (mar 1995), 56–74. https://doi.org/10.1123/ssj.12.1.56

[97] Angel Garcia-Crespo, Jose Luis Lopez-Cuadrado, and Israel Gonzalez-Carrasco. 2016. Accesibility on VoD Platforms via Mobile Devices. In *Applications and Usability of Interactive TV*, Maria Jose Abasolo, Francisco J. Perales, and Antoni Bibiloni (Eds.). Springer International Publishing, Cham, Switzerland, 149–160.

[98] Angel Garcia-Crespo, Mariuxi Montes-Chunga, Carlos Alberto Matheus-Chacin, and Ines Garcia-Encabo. 2018. Increasing the autonomy of deafblind individuals through direct access to content broadcasted on digital terrestrial television. *Assistive Technology* 32, 5 (dec 2018), 268–276. https://doi.org/10.1080/10400435.2018.1543219

[99] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural probes. *Interactions* 6, 1 (jan 1999), 21–29. https://doi.org/10.1145/291224.291235

[100] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 937–946. https://doi.org/10.1145/2207676.2208538

[101] Olivia Gerber-Moron, Olga Soler-Vilageliu, and Judit Castella. 2021. The effects of screen size on subtitle layout preferences and comprehension across devices. *Hermeneus. Revista de traduccion e interpretacion* 1, 22 (Feb. 2021), 157–182. https://doi.org/10.24197/her.22.2020.157-182

[102] Olivia Gerber-Moron, Agnieszka Szarkowska, and Bencie Woll. 2018. The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research* 11, 4 (jun 2018), 1–18. https://doi.org/10.16910/11.4.2

[103] Maxine Glancy, Lauren Ward, Nick Hanson, Andy Brown, and Michael Armstrong. 2020. Object-Based Media: An Overview of the User Experience. , 16 pages. https://www.bbc.co.uk/rd/publications/whitepaper390

[104] Gerard Goggin, Meryl Alper, and Joshua St Pierre. 2024. Disability athwart communication. *Journal of Communication* 74, 2 (02 2024), 177–182. https://doi.org/10.1093/joc/jqae005

[105] Benjamin M. Gorman, Michael Crabb, and Michael Armstrong. 2021. Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 733, 11 pages. https://doi.org/10.1145/3411764.3445509

[106] Gian Maria Greco. 2016. On Accessibility as a Human Right, with an Application to Media Accessibility. In *Researching Audio Description*. Palgrave Macmillan UK, London, UK, 11–33. https://doi.org/10.1057/978-1-137-56917-2_2

[107] Trisha Greenhalgh and Richard Peacock. 2005. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary

sources. *BMJ* 331, 7524 (oct 2005), 1064–1065. https://doi.org/10.1136/bmj.38636.593461.68

[108] Carole-Ann Greig, Renee Harper, Tanya Hirst, Tami Howe, and Bronwyn Davidson. 2008. Barriers and Facilitators to Mobile Phone Use for People with Aphasia. *Topics in Stroke Rehabilitation* 15, 4 (jul 2008), 307–324. https://doi.org/10.1310/tsr1504-307

[109] Brian Grellmann, Timothy Neate, Abi Roper, Stephanie Wilson, and Jane Marshall. 2018. Investigating Mobile Accessibility Guidance for People with Aphasia. 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, 410–413. https://doi.org/10.1145/3234695.3241011

[110] Jonathan Grudin. 1994. Groupware and social dynamics: eight challenges for developers. *Commun. ACM* 37, 1 (jan 1994), 92–105. https://doi.org/10.1145/175222.175230

[111] Nicholas-Brie Guarriello. 2019. Never give up, never surrender: Game live streaming, neoliberal work, and personalized media economies. *New Media and Society* 21, 8 (mar 2019), 1750—1769. https://doi.org/10.1177/1461444819831653

[112] Darren Guinness, Annika Muehlbradt, Daniel Szafir, and Shaun K. Kane. 2018. The Haptic Video Player: Using Mobile Robots to Create Tangible Video Annotations. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18)*. Association for Computing Machinery, New York, NY, USA, 203–211. https://doi.org/10.1145/3279778.3279805

[113] S.R. Gulliver and G. Ghinea. 2003. How level and type of deafness affect user perception of multimedia video clips. *Universal Access in the Information Society* 2, 4 (nov 2003), 374–386. https://doi.org/10.1007/s10209-003-0067-5

[114] Laxmi Gunupudi, Maryam Bandukda, Giulia Barbareschi, Tigmanshu Bhatnagar, Aanchal Singh, Satish Mishra, Amit Prakash, and Catherine Holloway. 2024. Scaffolding Digital Literacy Through Digital Skills Training for Disabled People in the Global South. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 29, 14 pages. https://doi.org/10.1145/3663548.3675666

[115] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1315–1324. https://doi.org/10.1145/2556288.2557048

[116] Noor Hammad, Erik Harpstead, and Jessica Hammer. 2023. GameAware Streaming Interfaces. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Stratford, ON, Canada) *(CHI PLAY Companion '23)*. Association for Computing Machinery, New York, NY, USA, 248–253. https://doi.org/10.1145/3573382.3616041

[117] Aimi Hamraie. 2017. *Sloped Technoscience: Curb Cuts, Critical Frictions, and Disability (Maker) Cultures*. University of Minnesota Press, Minneapolis, 95–130.

[118] Simon Harper. 2007. Is there design-for-all? *Universal Access in the Information Society* 6 (5 2007), 111–113. Issue 1. https://doi.org/10.1007/s10209-007-0071-2

[119] Niels Hendriks, Karin Slegers, and Pieter Duysburgh. 2015. Codesign with People Living with Cognitive or Sensory Impairments: A Case for Method Stories and Uniqueness. *CoDesign* 11, 1 (jan 2015), 70–82. https://doi.org/10.1080/15710882.2015.1020316

[120] Ruth Herbert, Caroline Haw, and Catherine Brown. 2012. Accessible information guidelines: Making information accessible for people with

aphasia. https://www.stroke.org.uk/sites/default/files/accessible_information_guidelines.pdf1_.pdf

[121] Mohamad Hassan Fadi Hijab, Dena Al-Thani, Joselia Neves, Nahwan Al Aswadi, and Shaza Khatab. 2023. Toward a Toolkit for Co-designing Collaborative Play Tool with and for Autistic Children. In *HCI in Games: 5th International Conference, HCI-Games 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23-28, 2023, Proceedings Part I* (Copenhagen, Denmark). Springer-Verlag, Berlin, Heidelberg, 114–132. https://doi.org/10.1007/978-3-031-35930-9_9

[122] Mohamad Hassan Fadi Hijab, Bilikis Banire, Joselia Neves, Marwa Qaraqe, Achraf Othman, and Dena Al-Thani. 2023. Co-design of Technology Involving Autistic Children: A Systematic Literature Review. *International Journal of Human-Computer Interaction* 40, 22 (Oct. 2023), 7498–7516. https://doi.org/10.1080/10447318.2023.2266248

[123] Stephen Hilgartner. 2015. *Capturing the Imaginary: Vanguards, visions and the synthetic biology revolution* (1st ed.). Routledge, London, UK, 33–55. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203564370-3/capturing-imaginary-stephen-hilgartner

[124] Alexis Hiniker, Jon E. Froehlich, Mingrui Zhang, and Erin Beneteau. 2019. Anchored Audio Sampling: A Seamless Method for Exploring Children's Thoughts During Deployment Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300238

[125] Dennis E. Hinkle, William Wiersma, and Stephen G. Jurs. 2003. *Applied Statistics for the Behavioral Sciences* (5 ed.). Houghton Mifflin Company, Boston, MA.

[126] Anthony Hornof, Haley Whitman, Marah Sutherland, Samuel Gerendasy, and Joanna McGrenere. 2017. Designing for the "Universe of One": Personalized Interactive Media Systems for People with the Severe Cognitive Impairment

Associated with Rett Syndrome. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2137–2148. https://doi.org/10.1145/3025453.3025904

[127] Lara Houston, Steven J. Jackson, Daniela K. Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. 2016. Values in Repair. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1403–1414. https://doi.org/10.1145/2858036.2858470

[128] Tami J. Howe, Linda E. Worrall, and Louise M. H. Hickson. 2008. Interviews with people with aphasia: Environmental factors that influence their community participation. *Aphasiology* 22, 10 (oct 2008), 1092–1120. https://doi.org/10.1080/02687030701640941

[129] Tami J. Howe, Linda E. Worrall, and Louise M. H. Hickson. 2008. Observing people with aphasia: Environmental factors that influence their community participation. *Aphasiology* 22, 6 (jun 2008), 618–643. https://doi.org/10.1080/02687030701536024

[130] Elfed Howells and David Jackson. 2021. Object-Based Media Report. https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/technology-research/2021/object-based-media-report.pdf?v=322207

[131] Chris Hughes, Mario Montagud Climent, and Peter tho Pesch. 2019. Disruptive Approaches for Subtitling in Immersive Environments. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) *(TVX '19)*. Association for Computing Machinery, New York, NY, USA, 216–229. https://doi.org/10.1145/3317697.3325123

[132] Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. 2015. Responsive design for personalised subtitles. In *Proceedings of the 12th International Web for All Conference* (Florence, Italy) *(W4A '15)*. Association

for Computing Machinery, New York, NY, USA, Article 8, 4 pages. https://doi.org/10.1145/2745555.2746650

[133] Lily Hughes, Niko Kargas, Maximilian Wilhelm, Hauke S. Meyerhoff, and Julia Focker. 2023. The Impact of Audio-Visual, Visual and Auditory Cues on Multiple Object Tracking Performance in Children with Autism. *Perceptual and Motor Skills* 130, 5 (jul 2023), 2047–2068. https://doi.org/10.1177/00315125231187984

[134] Karen Hux, Jessica A. Brown, Sarah Wallace, Kelly Knollman-Porter, Anna Saylor, and Erica Lapp. 2020. Effect of Text-to-Speech Rate on Reading Comprehension by Adults With Aphasia. *American Journal of Speech-Language Pathology* 29, 1 (Feb. 2020), 168–184. https://doi.org/10.1044/2019_ajslp-19-00047

[135] Albrecht W. Inhoff, Andrew Kim, and Ralph Radach. 2019. Regressions during Reading. *Vision* 3, 3 (jul 2019), 35. https://doi.org/10.3390/vision3030035

[136] Jan Ivarsson. 2009. *The History of Subtitles in Europe*. The Chinese University Press, Hong Kong, Chapter 1, 3–12.

[137] Juliane Jarke, Gabriela Molina Leon, Irina Zakharova, Hendrik Heuer, and Ulrike Gerhard. 2021. Beyond Participation: A Review of Co-Creation in Computing. https://doi.org/10.48550/ARXIV.2111.04524

[138] Lucy Jiang, Crescentia Jung, Mahika Phutane, Abigale Stangl, and Shiri Azenkot. 2024. "It's Kind of Context Dependent": Understanding Blind and Low Vision People's Video Accessibility Preferences Across Viewing Scenarios. 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 897, 20 pages. https://doi.org/10.1145/3613904.3642238

[139] Lucy Jiang, Woojin Ko, Shirley Yuan, Tanisha Shende, and Shiri Azenkot. 2025. Shifting the Focus: Exploring Video Accessibility Strategies and Challenges

for People with ADHD. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 561, 16 pages. https://doi.org/10.1145/3706598.3713637

[140] Asta Johannsdottir, Snaefriour Pora Egilson, and Freyja Haraldsdottir. 2022. Implications of internalised ableism for the health and wellbeing of disabled young people. *Sociology of Health and Illness* 44, 2 (jan 2022), 360–376. https://doi.org/10.1111/1467-9566.13425

[141] Rosie Jones, Hamed Zamani, Markus Schedl, Ching-Wei Chen, Sravana Reddy, Ann Clifton, Jussi Karlgren, Helia Hashemi, Aasish Pappu, Zahra Nazari, Longqi Yang, Oguz Semerci, Hugues Bouchard, and Ben Carterette. 2021. Current Challenges and Future Directions in Podcast Information Access. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 1554–1565. https://doi.org/10.1145/3404835.3462805

[142] Sushant Kafle, Becca Dingman, and Matt Huenerfauth. 2021. Deaf and Hard-of-Hearing Users Evaluating Designs for Highlighting Key Words in Educational Lecture Videos. *ACM Trans. Access. Comput.* 14, 4, Article 20 (oct 2021), 24 pages. https://doi.org/10.1145/3470651

[143] Aura Kagan. 1998. Supported conversation for adults with aphasia: methods and resources for training conversation partners. *Aphasiology* 12, 9 (sep 1998), 816–830. https://doi.org/10.1080/02687039808249575

[144] A. Kagan, N. Simmons-Mackie, and E. Shumway. 2018. Revised rating anchors and scoring procedures for measure of skill and measure of participation in conversation between adults with aphasia and their conversation partners. https://www.aphasia.ca/VF_MSCMPC. Toronto, ON, Canada. Aphasia Institute.

[145] Hyunjin Kang and S. Shyam Sundar. 2016. When Self Is the Source: Effects of

Media Customization on Message Processing. *Media Psychology* 19, 4 (2016), 561–588. https://doi.org/10.1080/15213269.2015.1121829

[146] Herminder Kaur and Paula Saukko. 2022. Social access: role of digital media in social relations of young people with disabilities. *New Media and Society* 24, 2 (feb 2022), 420–436. https://doi.org/10.1177/14614448211063177

[147] Claire Kearney-Volpe, Shannon Holloway, and Amy Hurst. 2019. Entertainment for All: Understanding Media Streaming Accessibility. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312882

[148] JP Kelly. 2020. "This Title Is No Longer Available": Preserving Television in the Streaming Age. *Television and New Media* 23, 1 (jun 2020), 3–21. https://doi.org/10.1177/1527476420928480

[149] Patrick J Kennedy and Andrea Prat. 2019. Where do people get their news? *Economic Policy* 34 (1 2019), 5–47. Issue 97. https://doi.org/10.1093/epolic/eiy016

[150] Daniel Killough and Amy Pavel. 2023. Exploring Community-Driven Descriptions for Making Livestreams Accessible. 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 42, 13 pages. https://doi.org/10.1145/3597638.3608425

[151] Nanum Kim, Sangsu Jang, Hansol Kim, Dayoung Shin, and Young-Woo Park. 2025. Lino: An Interactive System for Daily Mood Recordings Supporting Meaning-Making through Single Stroke Drawing Approach. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference* (Madeira, Portugal) *(DIS '25)*. Association for Computing Machinery, New York, NY, USA, 2255–2269. https://doi.org/10.1145/3715336.3735675

[152] J. Peter Kincaid, James A. Aagard, John W. O'Hara, and Larry K. Cottrell. 1981. Computer readability editing system. *IEEE Transactions on Professional Communication* PC-24, 1 (mar 1981), 38–42. https://doi.org/10.1109/tpc.1981.6447821

[153] Andrew Kirkpatrick, Joshue O Connor, Alastair Campbell, and Michael Cooper. 2021. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/

[154] Kelly Knollman-Porter, Jessica A. Brown, Karen Hux, Sarah E. Wallace, and Allison Crittenden. 2022. Reading Comprehension and Processing Time When People With Aphasia Use Text-to-Speech Technology With Personalized Supports and Features. *American Journal of Speech-Language Pathology* 31, 1 (jan 2022), 342–358. https://doi.org/10.1044/2021_ajslp-21-00182

[155] Kelly Knollman-Porter, Sarah E. Wallace, Karen Hux, Jessica Brown, and Candace Long. 2015. Reading experiences and use of supports by people with chronic aphasia. *Aphasiology* 29, 12 (apr 2015), 1448–1472. https://doi.org/10.1080/02687038.2015.1041093

[156] Junhan Kong, Jacob O. Wobbrock, Tianyuan Cai, and Zoya Bylinskii. 2025. Supporting Mobile Reading While Walking with Automatic and Customized Font Size Adaptations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokahama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1007, 13 pages. https://doi.org/10.1145/3706598.3713367

[157] Dimitrios Konstantinidis, Kosmas Dimitropoulos, Kiriakos Stefanidis, Thanassis Kalvourtzis, Salim Gannoum, Nikolaos Kaklanis, Konstantinos Votis, Petros Daras, Sara Rovira-Esteva, Pilar Orero, Silvia Uribe, Francisco Moreno, Alvaro Llorente, Pablo Calleja, Maria Poveda-Villalon, Pasquale Andriani, Giuseppe Vitolo, Giuseppa Caruso, Nicolamaria Manes, Fabrizio Giacomelli, Jordi Fabregat, Francesc Mas, Jordi Mata, Stavros Skourtis, Chrysostomos Bourlis, Giuliano Frittelli, Emilio Ferreiro Lago, and Federico Alvarez. 2020. Developing Accessibility Multimedia Services: The Case of EasyTV. In *Pro-*

ceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (Corfu, Greece) *(PETRA '20)*. Association for Computing Machinery, New York, NY, USA, Article 38, 8 pages. https://doi.org/10.1145/3389189.3397973

[158] Jan-Louis Kruger, Natalia Wisniewska, and Sixin Liao. 2021. Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics* 43, 1 (dec 2021), 211–236. https://doi.org/10.1017/s0142716421000503

[159] Kuno Kurzhals, Emine Cetinkaya, Yongtao Hu, Wenping Wang, and Daniel Weiskopf. 2017. Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6559–6568. https://doi.org/10.1145/3025453.3025772

[160] Kuno Kurzhals, Fabian Gobel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. 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. https://doi.org/10.1145/3313831.3376266

[161] Raja Kushalnagar, Matthew Seita, and Abraham Glasser. 2017. Closed ASL Interpreting for Online Videos. In *Proceedings of the 14th International Web for All Conference*. ACM, New York, NY, USA, 1–4. https://doi.org/10.1145/3058555.3058578

[162] Walter S. Lasecki, Raja Kushalnagar, and Jeffrey P. Bigham. 2014. Helping Students Keep up with Real-Time Captions by Pausing and Highlighting. In *Proceedings of the 11th Web for All Conference* (Seoul, Korea) *(W4A '14)*. Association for Computing Machinery, New York, NY, USA, Article 39, 8 pages. https://doi.org/10.1145/2596695.2596701

[163] Dawon Lee, Jongwoo Choi, and Junyong Noh. 2025. OptiSub: Optimizing Video Subtitle Presentation for Varied Display and Font Sizes via Speech Pause-

Driven Chunking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 799, 12 pages. https://doi.org/10.1145/3706598.3714199

[164] Sarah Lewthwaite. 2014. Web accessibility standards and disability: developing critical perspectives on accessibility. *Disability and Rehabilitation* 36, 16 (July 2014), 1375–1383. https://doi.org/10.3109/09638288.2014.938178

[165] Chu Li, Katrina Oi Yau Ma, Michael Saugstad, Kie Fujii, Molly Delaney, Yochai Eisenberg, Delphine Labbe, Judy L Shanley, Devon Snyder, Florian P P Thomas, and Jon E. Froehlich. 2024. "I never realized sidewalks were a big deal": A Case Study of a Community-Driven Sidewalk Accessibility Assessment using Project Sidewalk. 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 969, 18 pages. https://doi.org/10.1145/3613904.3642003

[166] Franklin Mingzhe Li, Cheng Lu, Zhicong Lu, Patrick Carrington, and Khai N. Truong. 2022. An Exploration of Captioning Practices and Challenges of Individual Content Creators on YouTube for People with Hearing Impairments. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 75 (apr 2022), 26 pages. https://doi.org/10.1145/3512922

[167] Xiangbo Li, Mohsen Amini Salehi, and Magdy Bayoumi. 2016. High Performance On-demand Video Transcoding Using Cloud Services. In *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*. IEEE Computer Society, New York, NY, USA, 600–603. https://doi.org/10.1109/CCGrid.2016.50

[168] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. 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, 1–14. https://doi.org/10.1145/3313831.3376258

[169] Martin Valdemar Anker Lindrup, Arjun Rajendran Menon, and Aksel Biorn-Hansen. 2023. Carbon Scales: Collective Sense-making of Carbon Emissions from Food Production through Physical Data Representation. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1515–1530. https://doi.org/10.1145/3563657.3596043

[170] Stephen Lindsay, Katie Brittain, Daniel Jackson, Cassim Ladha, Karim Ladha, and Patrick Olivier. 2012. Empathy, Participatory Design and People with Dementia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 521–530. https://doi.org/10.1145/2207676.2207749

[171] Stephen Lindsay, Daniel Jackson, Guy Schofield, and Patrick Olivier. 2012. Engaging Older People Using Participatory Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1199–1208. https://doi.org/10.1145/2207676.2208570

[172] Sebastian Linxen, Christian Sturm, Florian Bruhlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 143, 14 pages. https://doi.org/10.1145/3411764.3445488

[173] Alexandra List and Eric E. Ballenger. 2019. Comprehension across mediums: the case of text and video. *Journal of Computing in Higher Education* 31, 3 (feb 2019), 514–535. https://doi.org/10.1007/s12528-018-09204-9

[174] Xingyu Liu, Patrick Carrington, Xiang 'Anthony' Chen, and Amy Pavel. 2021. What Makes Videos Accessible to Blind and Visually Impaired People?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*

(Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 272, 14 pages. https://doi.org/10.1145/3411764.3445233

[175] Xingyu "Bruce" Liu, Ruolin Wang, Dingzeyu Li, Xiang Anthony Chen, and Amy Pavel. 2022. CrossA11y: Identifying Video Accessibility Issues via Cross-Modal Grounding. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 43, 14 pages. https://doi.org/10.1145/3526113.3545703

[176] Yixin Liu, Kai Zhang, Yuan Li, Zhiling Yan, Chujie Gao, Ruoxi Chen, Zhengqing Yuan, Yue Huang, Hanchi Sun, Jianfeng Gao, Lifang He, and Lichao Sun. 2024. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. https://doi.org/10.48550/arxiv.2402.17177

[177] Peter Ljunglof, Sandra Derbring, and Maria Olsson. 2012. A Free and Open-Source Tool That Reads Movie Subtitles Aloud. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies* (Montreal, Canada) *(SLPAT '12)*. Association for Computational Linguistics, USA, 1–4.

[178] Mariana Lopez, Gavin Kearney, and Krisztian Hofstadter. 2021. Enhancing Audio Description: Inclusive Cinematic Experiences Through Sound Design. *Journal of Audiovisual Translation* 4, 1 (oct 2021), 157–182. https://doi.org/10.47476/jat.v4i1.2021.154

[179] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. 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. https://doi.org/10.1145/3313831.3376311

[180] Thomas W MacFarland and Jan M Yates. 2016. Wilcoxon matched-pairs signed-ranks test. In *Introduction to Nonparametric Statistics for the Biological Sciences Using R*. Springer, Cham, Switzerland, 133–175. https://doi.org/10.1007/978-3-319-30634-6_5

[181] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by 'Accessibility Research'? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 371, 18 pages. https://doi.org/10.1145/3411764.3445412

[182] Jytte Isaksen Madeline Cruice, Monica Blom Johansson and Simon Horton. 2018. Reporting interventions in communication partner training: a critical review and narrative synthesis of the literature. *Aphasiology* 32, 10 (2018), 1135–1166. https://doi.org/10.1080/02687038.2018.1482406

[183] Jennifer Mankoff, Gillian R. Hayes, and Devva Kasnitz. 2010. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (Orlando, Florida, USA) *(ASSETS '10)*. Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/1878803.1878807

[184] Alejandro Martin, Israel Gonzalez-Carrasco, Victor Rodriguez-Fernandez, Monica Souto-Rico, David Camacho, and Belen Ruiz-Mezcua. 2021. DeepSync: A novel deep learning-based tool for semantic-aware subtitling synchronisation. *Neural Computing and Applications* 33, 1 (feb 2021), 1–15. https://doi.org/10.1007/s00521-021-05751-y

[185] Jindrich Matousek, Zdenek Hanzlicek, Daniel Tihelka, and Martin Mener. 2010. Automatic dubbing of TV programmes for the hearing impaired. In *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*. IEEE,

New York, NY, USA, 589–592. https://doi.org/10.1109/icosp.2010.5655861

[186] Robert W. McChesney. 2003. *Public Broadcasting: Past, Present, and Future*. Routledge, London, UK, Chapter 2, 10–24. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315290690-3/public-broadcasting-robert-mcchesney

[187] Emma J McDonnell, Tessa Eagle, Pitch Sinlapanuntakul, Soo Hyun Moon, Kathryn E. Ringland, Jon E. Froehlich, and Leah Findlater. 2024. "Caption It in an Accessible Way That Is Also Enjoyable": Characterizing User-Driven Captioning Practices on TikTok. 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 492, 16 pages. https://doi.org/10.1145/3613904.3642177

[188] Mark McGill, Stephen Brewster, David McGookin, and Graham Wilson. 2020. Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. 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, 1–16. https://doi.org/10.1145/3313831.3376702

[189] Mark McGill, Florian Mathis, Mohamed Khamis, and Julie Williamson. 2020. Augmenting TV Viewing using Acoustically Transparent Auditory Headsets. In *Proceedings of the 2020 ACM International Conference on Interactive Media Experiences* (Cornella, Barcelona, Spain) *(IMX '20)*. Association for Computing Machinery, New York, NY, USA, 34–44. https://doi.org/10.1145/3391614.3393650

[190] Mark McGill, John H. Williamson, and Stephen Brewster. 2016. Examining The Role of Smart TVs and VR HMDs in Synchronous At-a-Distance Media Consumption. *ACM Transactions on Computer-Human Interaction* 23, 5 (nov 2016), 1–57. https://doi.org/10.1145/2983530

[191] Bogdan Mocanu and Ruxandra Tapu. 2021. Automatic Subtitle Placement Through Active Speaker Identification in Multimedia Documents. In *2021 International Conference on e-Health and Bioengineering (EHB)*. IEEE, New York, NY, USA, 1–4. https://doi.org/10.1109/ehb52898.2021.9657604

[192] Mario Montagud, Fernando Boronat, Javier Pastor, and Dani Marfil. 2020. Web-based platform for a customizable and synchronized presentation of subtitles in single- and multi-screen scenarios. *Multimedia Tools and Applications* 79, 29-30 (may 2020), 21889–21923. https://doi.org/10.1007/s11042-020-08955-x

[193] Mario Montagud, Issac Fraile, Juan A. Nunez, and Sergi Fernandez. 2018. ImAc: Enabling Immersive, Accessible and Personalized Media Experiences. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video* (SEOUL, Republic of Korea) *(TVX '18)*. Association for Computing Machinery, New York, NY, USA, 245–250. https://doi.org/10.1145/3210825.3213570

[194] Karla Kmetz Morris, Casey Frechette, Lyman Dukes III, Nicole Stowell, Nicole Emert Topping, and David Brodosi. 2016. Closed Captioning Matters: Examining the Value of Closed Captions for" All" Students. *Journal of Postsecondary Education and Disability* 29, 3 (2016), 231–238.

[195] Reg Morris, Alicia Eccles, Brooke Ryan, and Ian I. Kneebone. 2017. Prevalence of anxiety in people with aphasia after stroke. *Aphasiology* 31, 12 (mar 2017), 1410–1415. https://doi.org/10.1080/02687038.2017.1304633

[196] Martez Mott, John Tang, Shaun Kane, Edward Cutrell, and Meredith Ringel Morris. 2020. "I just went into it assuming that I wouldn't be able to have the full experience": Understanding the Accessibility of Virtual Reality for People with Limited Mobility. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3373625.3416998

[197] Kalyan Nakka, Jimmy Dani, and Nitesh Saxena. 2024. Is On-Device AI Broken

and Exploitable? Assessing the Trust and Ethics in Small Language Models. https://doi.org/10.48550/arxiv.2406.05364

[198] Rosiana Natalie, Ruei-Che Chang, Smitha Sheshadri, Anhong Guo, and Kotaro Hara. 2024. Audio Description Customization. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 39, 19 pages. https://doi.org/10.1145/3663548.3675617

[199] Rosiana Natalie, Jolene Loh, Huei Suen Tan, Joshua Tseng, Ian Luke Yi-Ren Chan, Ebrima H Jarjue, Hernisa Kacorri, and Kotaro Hara. 2021. The Efficacy of Collaborative Authoring of Video Scene Descriptions. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 17, 15 pages. https://doi.org/10.1145/3441852.3471201

[200] Timothy Neate, Aikaterini Bourazeri, Abi Roper, Simone Stumpf, and Stephanie Wilson. 2019. Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300880

[201] Timothy Neate, Michael Evans, and Matt Jones. 2016. Designing Visual Complexity for Dual-screen Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 475–486. https://doi.org/10.1145/2858036.2858112

[202] Timothy Neate, Abi Roper, Stephanie Wilson, Jane Marshall, and Madeline Cruice. 2020. CreaTable Content and Tangible Interaction in Aphasia. 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, 1–14. https://doi.org/10.1145/3313831.3376490

[203] David Nemer. 2022. *Repairing the Broken City*. The MIT Press, Cambridge, Massachusetts, United States, 33–54. https://doi.org/10.7551/mitpress/14122.003.0004

[204] Horace M. Newcomb and Paul M. Hirsch. 1983. Television as a cultural forum: Implications for research. *Quarterly Review of Film Studies* 8 (6 1983), 45–55. Issue 3. https://doi.org/10.1080/10509208309361170

[205] A. F. Newell and P. Gregor. 1999. Extra-Ordinary Human-Machine Interaction: What can be Learned from People with Disabilities? *Cognition, Technology and Work* 1, 2 (sep 1999), 78–85. https://doi.org/10.1007/s101110050034

[206] ngrid Burkett. 2012. An Introduction to Co-design. https://www.yacwa.org.au/wp-content/uploads/2016/09/An-Introduction-to-Co-Design-by-Ingrid-Burkett.pdf

[207] Simon Nielsen and Hans-Heinrich Bothe. 2007. SUBPAL: A Device for Reading Aloud Subtitles from Television and Cinema. In *Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments: Assistive Technology for All Ages (CVHI-2007), Granada, Spain, 28th - 31th August, 2007 (CEUR Workshop Proceedings, Vol. 415)*, Marion A. Hersh and James Ohene-Djan (Eds.). CEUR-WS.org, Leibniz, Germany, 1–7. https://ceur-ws.org/Vol-415/paper17.pdf

[208] Don Norman. 2013. *The design of everyday things*. Basic Books, London, England.

[209] Sarah Northcott, Alan Simpson, Becky Moss, Nafiso Ahmed, and Katerina Hilari. 2018. Supporting people with aphasia to 'settle into a new way to be': speech and language therapists' views on providing psychosocial support. *International Journal of Language and Communication Disorders* 53, 1 (2018), 16–29. https://doi.org/10.1111/1460-6984.12323

[210] Gwen C. Nugent. 1983. Deaf Students' Learning From Captioned Instruction: the Relationship Between the Visual and Caption Display. *The Journal of Special Education* 17, 2 (July 1983), 227–234. https://doi.org/10.1177/002246698301700214

[211] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. 2021. Designing AACs for People with Aphasia Dining in Restaurants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 496, 14 pages. https://doi.org/10.1145/3411764.3445280

[212] OFCOM. 2020. Internet users' experience of potential online harms: summary of survey research. https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-harms/2020/concerns-and-experiences-online-harms-2020-chart-pack-accessible.pdf?v=324903

[213] J. Ohene-Djan and R. Shipsey. 2006. E- Subtitles: Emotional Subtitles as a Technology to Assist the Deaf and Hearing-Impaired when Learning from Television and Film.. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT '06)*. IEEE, New York, NY, USA, 464–466. https://doi.org/10.1109/icalt.2006.1652472

[214] Keita Ohshiro and Mark Cartwright. 2024. Audio Engineering by People Who Are deaf and Hard of Hearing: Balancing Confidence and Limitations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 35, 13 pages. https://doi.org/10.1145/3613904.3642454

[215] Rita Oliveira, Jorge Ferraz de Abreu, and Ana Margarida Almeida. 2014. Promoting IPTV Accessibility for Visually Impaired Users: Implementation of an Adapted Service. *Procedia Computer Science* 27 (2014), 113–122. https://doi.org/10.1016/j.procs.2014.02.014 5th International Confer-

ence on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013.

[216] Michael Oliver. 1996. *Understanding Disability*. Macmillan Education UK, UK. https://doi.org/10.1007/978-1-349-24269-6

[217] Michael J. Oliver. 1993. What's so wonderful about walking. , 28 pages. https://disability-studies.leeds.ac.uk/wp-content/uploads/sites/40/library/Oliver-PROFLEC.pdf

[218] Pilar Orero, Marta Brescia-Zapata, and Chris Hughes. 2021. Evaluating Subtitle Readability in Media Immersive Environments. In *9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion* (Online, Portugal) *(DSAI 2020)*. Association for Computing Machinery, New York, NY, USA, 51–54. https://doi.org/10.1145/3439231.3440602

[219] Carol O'Sullivan and Jean-Francois Cornu. 2018. *History of Audiovisual Translation* (1st ed.). Routledge, UK, Chapter 2, 15–30.

[220] Patrick B O'Sullivan and Caleb T Carr. 2017. Mass personal communication: A model bridging the mass-interpersonal divide. *New Media and Society* 20, 3 (jan 2017), 1161–1180. https://doi.org/10.1177/1461444816686104

[221] Cristhian Pachacama and Diego Villamarin. 2020. Implementation of a Brain Computer Interface System to Allow TV Remote Control for People with Physical Disabilities. In *Communications in Computer and Information Science*. Springer International Publishing, Cham, Switzerland, 119–133. https://doi.org/10.1007/978-3-030-56574-9_8

[222] Jaclyn Packer, Katie Vizenor, and Joshua A. Miele. 2015. An Overview of Video Description: History, Benefits, and Guidelines. *Journal of Visual Impairment and Blindness* 109, 2 (mar 2015), 83–93. https://doi.org/10.1177/0145482x1510900204

[223] Rowan Page and Leah Heiss. 2023. *Method case study - Making design thinking tactile: unlocking meaning and experiences with tactile tools and generative*

prototypes. Edward Elgar Publishing, UK, 70–78. https://doi.org/10.4337/9781802203134.00010

[224] Cristian Pamparau and Radu-Daniel Vatavu. 2022. The User Experience of Journeys in the Realm of Augmented Reality Television. In *ACM International Conference on Interactive Media Experiences*. ACM, New York, NY, USA, 161—174. https://doi.org/10.1145/3505284.3529969

[225] Carolyn Pang, Zhiqin Collin Wang, Joanna McGrenere, Rock Leung, Jiamin Dai, and Karyn Moffatt. 2021. Technology Adoption and Learning Preferences for Older Adults: Evolving Perceptions, Ongoing Challenges, and Emerging Design Opportunities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 490, 13 pages. https://doi.org/10.1145/3411764.3445702

[226] Phil Parette and Marcia Scherer. 2004. Assistive Technology Use and Stigma. *Education and Training in Developmental Disabilities* 39 (9 2004), 217–226. Issue 3. https://www.jstor.org/stable/23880164

[227] Jimin Park, Chaerin Lee, Eunbin Cho, and Uran Oh. 2024. Enhancing the Podcast Browsing Experience through Topic Segmentation and Visualization with Generative AI. In *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences* (Stockholm, Sweden) *(IMX '24)*. Association for Computing Machinery, New York, NY, USA, 117–128. https://doi.org/10.1145/3639701.3656324

[228] Elisa Perego. 2014. *Film language and tool*. John Benjamins Publishing Company, Amsterdam, Netherlands, Chapter 5, 81–102.

[229] Andrew Perrin and Sara Atske. 2021. Americans with disabilities less likely than those without to own some digital devices. https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-harms/2020/concerns-and-experiences-online-harms-2020-chart-pack-accessible.pdf?v=324903

[230] Helen Petrie. 1997. User-Centred Design and Evaluation of Adaptive and Assistive Technology for Disabled and Elderly Users. *Information Technology* 39 (2 1997), 7–12. Issue 2. https://doi.org/10.1524/itit.1997.39.2.7

[231] Helen Petrie, Stefan Carmien, and Andrew Lewis. 2018. *Assistive Technology Abandonment: Research Realities and Potentials*. Springer International Publishing, New York, NY, USA, 532–540. https://doi.org/10.1007/978-3-319-94274-2_77

[232] Sebastian Pfotenhauer, Brice Laurent, Kyriaki Papageorgiou, and Jack Stilgoe. 2022. The politics of scaling. *Social Studies of Science* 52 (2 2022), 3–34. Issue 1. https://doi.org/10.1177/03063127211048945

[233] Betsy Phillips and Hongxin Zhao. 1993. Predictors of Assistive Technology Abandonment. *Assistive Technology* 5, 1 (jun 1993), 36–45. https://doi.org/10.1080/10400435.1993.10132205

[234] Mark Priestley, Martha Stickings, Ema Loja, Stefanos Grammenos, Anna Lawson, Lisa Waddington, and Bjarney Fridriksdottir. 2016. The political participation of disabled people in Europe: Rights, accessibility and activism. *Electoral Studies* 42 (6 2016), 1–9. https://doi.org/10.1016/j.electstud.2016.01.009

[235] Graham Pullin. 2009. *Design Meets Disability*. MIT Press, London, England. https://mitpress.mit.edu/9780262516747/design-meets-disability/

[236] Cynthia Putnam, Christina Hanschke, Jennifer Todd, Jonathan Gemmell, and Mia Kollia. 2019. Interactive Technologies Designed for Children with Autism. *ACM Transactions on Accessible Computing* 12, 3 (sep 2019), 1–37. https://doi.org/10.1145/3342285

[237] Gulnar Rakhmetulla and Ahmed Sabbir Arif. 2023. Crownboard: A One-Finger Crown-Based Smartwatch Keyboard for Users with Limited Dexterity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing*

Systems (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 46, 22 pages. https://doi.org/10.1145/3544548.3580770

[238] Parthasarathy Ranganathan, Sarita Adve, and Norman P. Jouppi. 1999. Performance of Image and Video Processing with General-Purpose Processors and Media ISA Extensions. In *Proceedings of the 26th Annual International Symposium on Computer Architecture* (Atlanta, Georgia, USA) *(ISCA '99)*. IEEE Computer Society, USA, 124–135. https://doi.org/10.1145/300979.300990

[239] Raisa Rashid, Jonathan Aitken, and Deborah I. Fels. 2006. Expressing Emotions Using Animated Text Captions. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 24–31. https://doi.org/10.1007/11788713_5

[240] Melissa L. Rethlefsen, Shona Kirtley, Siw Waffenschmidt, Ana Patricia Ayala, David Moher, Matthew J. Page, and Jonathan B. Koffel. 2021. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews* 10, 1 (jan 2021), 1–19. https://doi.org/10.1186/s13643-020-01542-z

[241] Jacob M. Rigby, Duncan P. Brumby, Anna L. Cox, and Sandy J. J. Gould. 2016. Watching movies on netflix. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM, New York, NY, USA, 714–721. https://doi.org/10.1145/2957265.2961843

[242] Valeria Righi, Sergio Sayago, and Josep Blat. 2017. When we talk about older people in HCI, who are we talking about? Towards a turn to community in the design of technologies for a growing ageing population. *Int. J. Hum.-Comput. Stud.* 108, C (Dec. 2017), 15–31. https://doi.org/10.1016/j.ijhcs.2017.06.005

[243] Kathryn E. Ringland, Jennifer Nicholas, Rachel Kornfield, Emily G. Lattie, David C. Mohr, and Madhu Reddy. 2019. Understanding Mental Ill-health

as Psychosocial Disability: Implications for Assistive Technology. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 156–170. https://doi.org/10.1145/3308561.3353785

[244] RNIB. 2018. Audio description on BBC. https://www.rnib.org.uk/news/audio-description-on-bbc/

[245] Toni Robertson and Jesper Simonsen. 2012. Challenges and Opportunities in Contemporary Participatory Design. *Design Issues* 28, 3 (July 2012), 3–9. https://doi.org/10.1162/desi_a_00157

[246] Sunil Rodger, Dan Jackson, John Vines, Janice McLaughlin, and Peter Wright. 2019. JourneyCam: Exploring Experiences of Accessibility and Mobility among Powered Wheelchair Users through Video and Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300860

[247] Aitor Rodriguez-Alsina, Guillermo Talavera, Pilar Orero, and Jordi Carrabina. 2012. Subtitle Synchronization across Multiple Screens and Devices. *Sensors* 12, 7 (jun 2012), 8710–8731. https://doi.org/10.3390/s120708710

[248] Yvonne Rogers, Youn-kyung Lim, William R. Hazlewood, and Paul Marshall. 2009. Equal Opportunities: Do Shareable Interfaces Promote More Group Participation Than Single User Displays? *Human-Computer Interaction* 24, 1–2 (apr 2009), 79–116. https://doi.org/10.1080/07370020902739379

[249] Dagfinn Romen and Dag Svanaes. 2008. Evaluating web site accessibility: validating the WAI guidelines through usability testing with disabled users. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges* (Lund, Sweden) *(NordiCHI '08)*. Association for Computing Machinery, New York, NY, USA, 535–538. https://doi.org/10.1145/1463160.1463238

[250] Dagfinn Romen and Dag Svanaes. 2012. Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users. *Univers. Access Inf. Soc.* 11, 4 (Nov. 2012), 375–385. https://doi.org/10.1007/s10209-011-0259-3

[251] Pablo Romero-Fresco. 2019. *Accessible Filmmaking: Integrating translation and accessibility into the filmmaking process*. Routledge, London, UK. https://doi.org/10.4324/9780429053771

[252] John Rooksby, Mattias Rost, Alistair Morrison, Marek Bell, Matthew Chalmers, and Timothy Smith. 2014. Practices of parallel media: Using mobile devices when watching television. , 5 pages.

[253] Abi Roper, Ian Davey, Stephanie Wilson, Timothy Neate, Jane Marshall, and Brian Grellmann. 2018. Usability Testing - An Aphasia Perspective. 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, 102–106. https://doi.org/10.1145/3234695.3241481

[254] Tanya A. Rose, Linda E. Worrall, Louise M. Hickson, and Tammy C. Hoffmann. 2012. Guiding principles for printed education materials: Design preferences of people with aphasia. *International Journal of Speech-Language Pathology* 14, 1 (dec 2012), 11–23. https://doi.org/10.3109/17549507.2011.631583

[255] Robert Rosenthal, Harris Cooper, Larry Hedges, et al. 1994. Parametric measures of effect size. *The handbook of research synthesis* 621, 2 (1994), 231–244. https://psycnet.apa.org/record/1993-99100-015

[256] Sylvia Rothe, Kim Tran, and Heinrich Hussmann. 2018. Dynamic Subtitles in Cinematic Virtual Reality. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video* (SEOUL, Republic of Korea) *(TVX '18)*. Association for Computing Machinery, New York, NY, USA, 209–214. https://doi.org/10.1145/3210825.3213556

[257] Sylvia Rothe, Kim Tran, and Heinrich Hussmann. 2018. Positioning of Subtitles in Cinematic Virtual Reality. In *ICAT-EGVE 2018 - International Conference*

on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, Gerd Bruder, Shunsuke Yoshimoto, and Sue Cobb (Eds.). The Eurographics Association, Germany, 1–8. https://doi.org/10.2312/egve.20181307

[258] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation Accuracy Is Good, but High Controllability May Be Better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300750

[259] Andrew Salway and Mike Graham. 2003. Extracting Information about Emotions in Films. In *Proceedings of the Eleventh ACM International Conference on Multimedia* (Berkeley, CA, USA) *(MULTIMEDIA '03)*. Association for Computing Machinery, New York, NY, USA, 299–302. https://doi.org/10.1145/957013.957076

[260] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the New Landscapes of Design. *CoDesign* 4, 1 (mar 2008), 5–18. https://doi.org/10.1080/15710880701875068

[261] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2020. *Progressive Transformers for End-to-End Sign Language Production*. Springer International Publishing, Switzerland, 687–705. https://doi.org/10.1007/978-3-030-58621-8_40

[262] Steven Schirra, Huan Sun, and Frank Bentley. 2014. Together alone: motivations for live-tweeting a television series. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2441–2450. https://doi.org/10.1145/2556288.2557070

[263] Tim Schlippe, Shaimaa Alessai, Ghanimeh El-Taweel, Matthias Wolfel, and Wajdi Zaghouani. 2020. Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic. In *2020 International Conference on Cyberworlds*

(CW). IEEE, New York, NY, USA, 196–203. https://doi.org/10.1109/CW49994.2020.00039

[264] Niels Seidel. 2015. Making Web Video Accessible: Interaction Design Patterns for Assistive Video Learning Environments. In *Proceedings of the 20th European Conference on Pattern Languages of Programs* (Kaufbeuren, Germany) *(EuroPLoP '15)*. Association for Computing Machinery, New York, NY, USA, Article 17, 16 pages. https://doi.org/10.1145/2855321.2855339

[265] Kyoungwon Seo, Samuel Dodson, Negar M. Harandi, Nathan Roberson, Sidney Fels, and Ido Roll. 2021. Active learning with online video: The impact of learning context on engagement. *Computers and Education* 165 (may 2021), 16 pages. https://doi.org/10.1016/j.compedu.2021.104132

[266] Tom Shakespeare. 2006. The Social Model of Disability. *The Disability Studies Reader* 2 (2006), 197–204.

[267] Devyani Sharma, Erez Levon, and Yang Ye. 2022. 50 years of British accent bias: Stability and lifespan change in attitudes to accents. *English World-Wide* 43, 2 (2022), 135–166. https://doi.org/10.1075/eww.20010.sha

[268] Shannon M. Sheppard and Rajani Sebastian. 2020. Diagnosing and managing post-stroke aphasia. *Expert Review of Neurotherapeutics* 21, 2 (dec 2020), 221–234. https://doi.org/10.1080/14737175.2020.1855976

[269] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological Momentary Assessment. *Annual Review of Clinical Psychology* 4, 1 (apr 2008), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[270] Kristen Shinohara, Cynthia L. Bennett, Wanda Pratt, and Jacob O. Wobbrock. 2018. Tenets for Social Accessibility: Towards Humanizing Disabled People in Design. *ACM Trans. Access. Comput.* 11, 1, Article 6 (mar 2018), 31 pages. https://doi.org/10.1145/3178855

[271] Kristen Shinohara and Josh Tenenberg. 2009. A blind person's interactions

with technology. *Commun. ACM* 52, 8 (aug 2009), 58–66. https://doi.org/10.1145/1536616.1536636

[272] Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 705–714. https://doi.org/10.1145/1978942.1979044

[273] Ben Shirley, Melissa Meadows, Fadi Malak, James Woodcock, and Ash Tidball. 2017. Personalized Object-Based Audio for Hearing Impaired TV Viewers. *Journal of the Audio Engineering Society* 65, 4 (apr 2017), 293–303. https://doi.org/10.17743/jaes.2017.0005

[274] Andy P. Siddaway, Alex M. Wood, and Larry V. Hedges. 2019. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology* 70, 1 (jan 2019), 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

[275] Rodrigo L. S. Silva and Francila Weidt Neiva. 2016. Systematic Literature Review in Computer Science - A Practical Guide. , 8 pages. https://doi.org/10.13140/RG.2.2.35453.87524

[276] Ellen Simpson, Samantha Dalal, and Bryan Semaan. 2023. "Hey, Can You Add Captions?": The Critical Infrastructuring Practices of Neurodiverse People on TikTok. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 57 (apr 2023), 27 pages. https://doi.org/10.1145/3579490

[277] Laurianne Sitbon, Ross Brown, and Lauren Fell. 2019. Turning Heads: Designing Engaging Immersive Video Experiences to Support People with Intellectual Disability when Learning Everyday Living Skills. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 171–182. https://doi.org/10.1145/3308561.3353787

[278] Than Htut Soe, Frode Guribye, and Marija Slavkovik. 2021. Evaluating AI assisted subtitling. In *Proceedings of the 2021 ACM International Conference*

on Interactive Media Experiences (Virtual Event, USA) *(IMX '21)*. Association for Computing Machinery, New York, NY, USA, 96–107. https://doi.org/10.1145/3452918.3458792

[279] Katta Spiel, Emeline Brulé, Christopher Frauenberger, Gilles Bailly, and Geraldine Fitzpatrick. 2018. Micro-ethics for participatory design with marginalised children. In *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1* (Hasselt and Genk, Belgium) *(PDC '18)*. Association for Computing Machinery, New York, NY, USA, Article 17, 12 pages. https://doi.org/10.1145/3210586.3210603

[280] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. 2018. Fitter, Happier, More Productive? The Normative Ontology of Fitness Trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3170427.3188401

[281] Katta Spiel, Laura Malinverni, Judith Good, and Christopher Frauenberger. 2017. Participatory Evaluation with Autistic Children. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5755–5766. https://doi.org/10.1145/3025453.3025851

[282] Aaron Springer and Steve Whittaker. 2020. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Trans. Interact. Intell. Syst.* 10, 4, Article 29 (oct 2020), 32 pages. https://doi.org/10.1145/3374218

[283] Mark Springett, Richard N Griffiths, and Martin Mihajlov. 2017. Issues faced by vision-impaired users of interactive TV search facilities. In *Proceedings of the Australasian Computer Science Week Multiconference*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3014812.3014824

[284] Abigale Stangl, Shasta Ihorn, Yue-Ting Siu, Aditya Bodi, Mar Castanon, Lothar D Narins, and Ilmi Yoon. 2023. The Potential of a Visual Dialogue Agent In a

Tandem Automated Audio Description System for Videos. 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 32, 17 pages. https://doi.org/10.1145/3597638.3608402

[285] Abigale Stangl, Nitin Verma, Kenneth R. Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who are Blind or Have Low Vision. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 16, 15 pages. https://doi.org/10.1145/3441852.3471233

[286] Pieter Jan Stappers and Elisa Giaccardi. 2017. *Research through Design* (2nd ed.). The Interaction Design Foundation, US, 1–94. https://research.tudelft.nl/en/publications/research-through-design-2

[287] S. Shyam Sundar, Saraswathi Bellur, and Haiyan Jia. 2012. *Motivational Technologies: A Theoretical Framework for Designing Preventive Health Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 112–122. https://doi.org/10.1007/978-3-642-31037-9_10

[288] S. Shyam Sundar, Haiyan Jia, T. Franklin Waddell, and Yan Huang. 2015. *Toward a Theory of Interactive Media Effects (TIME)*. Wiley, UK, 47–86. https://doi.org/10.1002/9781118426456.ch3

[289] S. Shyam Sundar and Sampada S. Marathe. 2010. Personalization versus Customization: The Importance of Agency, Privacy, and Power Usage. *Human Communication Research* 36, 3 (jun 2010), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

[290] Sarit Felicia Anais Szpiro, Shafeka Hashash, Yuhang Zhao, and Shiri Azenkot. 2016. How People with Low Vision Access Computing Devices: Understanding Challenges and Opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA)

(ASSETS '16). Association for Computing Machinery, New York, NY, USA, 171–180. https://doi.org/10.1145/2982142.2982168

[291] Karen Tanenbaum and Theresa Jean Tanenbaum. 2010. Agency as commitment to meaning: communicative competence in games. *Digital Creativity* 21, 1 (mar 2010), 11–17. https://doi.org/10.1080/14626261003654509

[292] Mary Frances Theofanos and Janice (Ginny) Redish. 2003. Bridging the gap: between accessibility and usability. *Interactions* 10, 6 (Nov. 2003), 36–51. https://doi.org/10.1145/947226.947227

[293] Frank Thorn and Sondra Thorn. 1996. Television Captions for Hearing-Impaired People: A Study of Key Factors that Affect Reading Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38, 3 (sep 1996), 452–463. https://doi.org/10.1518/001872096778702006

[294] David Thue, Vadim Bulitko, Marcia Spetch, and Trevon Romanuik. 2010. *Player Agency and the Relevance of Decisions*. Springer-Verlag, Berlin, Heidelberg, 210–215. https://doi.org/10.1007/978-3-642-16638-9_26

[295] Tiffany Tong, Jonathan H. Chan, and Mark Chignell. 2017. Serious Games for Dementia. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Perth, Australia) *(WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1111–1115. https://doi.org/10.1145/3041021.3054930

[296] Anna Lowenhaupt Tsing. 2012. On Nonscalability: The Living World Is Not Amenable to Precision-Nested Scales. *Common Knowledge* 18 (8 2012), 505–524. Issue 3. https://doi.org/10.1215/0961754X-1630424

[297] Graeme Turner. 2019. Approaching the cultures of use: Netflix, disruption and the audience. *Critical Studies in Television: The International Journal of Television Studies* 14, 2 (may 2019), 222–232. https://doi.org/10.1177/1749602019834554

[298] Graeme Turner and Jinna Tay (Eds.). 2009. *Television Studies After TV*. Routledge, London, England. https://www.routledge.com/Television-Studies-After-TV-Understanding-Television-in-the-Post-B Turner-Tay/p/book/9780415477703

[299] J.P. Udo, B. Acevedo, and D.I. Fels. 2010. Horatio audio-describes Shakespeare's Hamlet. *British Journal of Visual Impairment* 28, 2 (may 2010), 139–156. https://doi.org/10.1177/0264619609359753

[300] David Unbehaun, Konstantin Aal, Daryoush Daniel Vaziri, Peter David Tolmie, Rainer Wieching, David Randall, and Volker Wulf. 2020. Social technology appropriation in dementia: Investigating the role of caregivers in engaging people with dementia with a videogame-based training system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376648

[301] European Union. 2019. Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services. *Official Journal of the European Union* 151, 70 (2019), 70–115.

[302] Marian Ursu, Davy Smith, Jonathan Hook, Shauna Concannon, and John Gray. 2020. Authoring Interactive Fictional Stories in Object-Based Media (OBM). In *ACM International Conference on Interactive Media Experiences*. ACM, New York, NY, USA, 127–137. https://doi.org/10.1145/3391614.3393654

[303] Jose van Dijck, David Nieborg, and Thomas Poell. 2019. Reframing platform power. *Internet Policy Review* 8 (6 2019), 1–18. Issue 2. https://doi.org/10.14763/2019.2.1414

[304] Rosemary A. Varley, Nicolai J. C. Klessinger, Charles A. J. Romanowski, and Michael Siegal. 2005. Agrammatic but Numerate. *Proceedings of the National Academy of Sciences* 102, 9 (feb 2005), 3519–3524. https://doi.org/10.1073/pnas.0407470102

[305] Radu-Daniel Vatavu. 2021. Accessibility of Interactive Television and Media Experiences: Users with Disabilities Have Been Little Voiced at IMX and TVX.

In *ACM International Conference on Interactive Media Experiences* (Virtual Event, USA) *(IMX '21)*. Association for Computing Machinery, New York, NY, USA, 218–222. https://doi.org/10.1145/3452918.3465485

[306] Johana M. Rosas Villena, Bruno C. Ramos, Renata P.M. Fortes, and Rudinei Goularte. 2014. An Accessible Video Player for Older People: Issues from a User Test. *Procedia Computer Science* 27 (2014), 168–175. https://doi.org/10.1016/j.procs.2014.02.020 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013.

[307] Vinoba Vinayagamoorthy, Maxine Glancy, Paul Debenham, Alastair Bruce, Christoph Ziegler, and Richard Schaffer. 2018. Personalising the TV Experience with Augmented Reality Technology: Synchronised Sign Language Interpretation. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video* (SEOUL, Republic of Korea) *(TVX '18)*. Association for Computing Machinery, New York, NY, USA, 179–184. https://doi.org/10.1145/3210825.3213562

[308] Vinoba Vinayagamoorthy, Maxine Glancy, Christoph Ziegler, and Richard Schaffer. 2019. Personalising the TV Experience Using Augmented Reality: An Exploratory Study on Delivering Synchronised Sign Language Interpretation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300762

[309] Lakshmie Narayan Viswanathan, Troy McDaniel, and Sethuraman Panchanathan. 2011. Audio-Haptic Description in Movies. In *HCI International 2011 – Posters' Extended Abstracts*, Constantine Stephanidis (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 414–418.

[310] Quoc V. Vy and Deborah I. Fels. 2009. Using Avatars for Improving Speaker Identification in Captioning. In *Human-Computer Interaction - INTERACT 2009*. Springer Berlin Heidelberg, Berlin, Heidelberg, 916–919. https://doi.org/10.1007/978-3-642-03658-3_110

[311] Agnieszka Walczak. 2017. Audio Description on Smartphones: Making Cinema Accessible for Visually Impaired Audiences. *Universal Access in the Information Society* 17, 4 (aug 2017), 833–840. https://doi.org/10.1007/s10209-017-0568-2

[312] Agnieszka Walczak and Louise Fryer. 2017. Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment* 35, 1 (2017), 6–17. https://doi.org/10.1177/0264619616661603 arXiv:https://doi.org/10.1177/0264619616661603

[313] Xubin Wang, Zhiqing Tang, Jianxiong Guo, Tianhui Meng, Chenhao Wang, Tian Wang, and Weijia Jia. 2025. Empowering Edge Intelligence: A Comprehensive Survey on On-Device AI Models. *ACM Comput. Surv.* 57, 9, Article 228 (apr 2025), 39 pages. https://doi.org/10.1145/3724420

[314] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward Automatic Audio Description Generation for Accessible Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 277, 12 pages. https://doi.org/10.1145/3411764.3445347

[315] Yu-Huan Wang, Tian-Jun Gu, and Shyang-Yuh Wang. 2019. Causes and Characteristics of Short Video Platform Internet Community Taking the TikTok Short Video Application as an Example. In *2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*. IEEE, New York, NY, USA, 1–2. https://doi.org/10.1109/icce-tw46550.2019.8992021

[316] Lauren Ward and Ben Shirley. 2019. Personalization in Object-based Audio for Accessibility: A Review of Advancements for Hearing Impaired Listeners. *Journal of the Audio Engineering Society* 67, 7/8 (aug 2019), 584–597. https://doi.org/10.17743/jaes.2019.0021

[317] Lauren Ward, Ben Shirley, and Jon Francombe. 2018. *Accessible Object-Based Audio Using Hierarchical Narrative Importance Metadata*. Techni-

cal Report. BBC. https://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP395.pdf

[318] Jennifer Wehrmeyer. 2014. Eye-tracking Deaf and hearing viewing of sign language interpreted news broadcasts. *Journal of Eye Movement Research* 7, 1 (mar 2014), 1–16. https://doi.org/10.16910/jemr.7.1.3

[319] Maarten Wijnants, Eva Geurts, Hendrik Lievens, Peter Quax, and Wim Lamotte. 2021. Stay Tuned! An Investigation of Content Substitution, the Listener as Curator and Other Innovations in Broadcast Radio. In *Proceedings of the 2021 ACM International Conference on Interactive Media Experiences* (Virtual Event, USA) *(IMX '21)*. Association for Computing Machinery, New York, NY, USA, 120–133. https://doi.org/10.1145/3452918.3458793

[320] Stephanie Wilson, Abi Roper, Jane Marshall, Julia Galliers, Niamh Devane, Tracey Booth, and Celia Woolf. 2015. Codesign for people with aphasia through tangible design languages. *CoDesign* 11, 1 (jan 2015), 21–34. https://doi.org/10.1080/15710882.2014.997744

[321] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* 23, 3 (apr 2016), 38–44. https://doi.org/10.1145/2907069

[322] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (London, England, United Kingdom) *(EASE '14)*. Association for Computing Machinery, New York, NY, USA, Article 38, 10 pages. https://doi.org/10.1145/2601248.2601268

[323] Donghee Yvette Wohn and Guo Freeman. 2020. Audience Management Practices of Live Streamers on Twitch. In *Proceedings of the 2020 ACM International Conference on Interactive Media Experiences* (Cornella, Barcelona, Spain) *(IMX '20)*. Association for Computing Machinery, New York, NY, USA, 106–116. https://doi.org/10.1145/3391614.3393653

[324] D. Yvette Wohn and Eun-Kyung Na. 2011. Tweeting about TV: Sharing Television Viewing Experiences via Social Media Message Streams. *First Monday* 6, 3 (feb 2011), 13 pages. https://doi.org/10.5210/fm.v16i3.3368

[325] Christine T. Wolf and Kathryn E. Ringland. 2020. Designing accessible, explainable AI (XAI) experiences. *SIGACCESS Access. Comput.* 1, 125, Article 6 (March 2020), 1 pages. https://doi.org/10.1145/3386296.3386302

[326] James S. Wolffsohn, Ditipriya Mukhopadhyay, and Martin Rubinstein. 2007. Image Enhancement of Real-Time Television to Benefit the Visually Impaired. *American Journal of Ophthalmology* 144, 3 (sep 2007), 436–440. https://doi.org/10.1016/j.ajo.2007.05.031

[327] Linda Worrall, Tanya Rose, Tami Howe, Kryss McKenna, and Louise Hickson. 2007. Developing an evidence-base for accessibility for people with aphasia. *Aphasiology* 21, 1 (2007), 124–136. https://doi.org/10.1080/02687030600798352

[328] Shuchang Xu, Xiaofu Jin, Huamin Qu, and Yukang Yan. 2025. DanmuA11y: Making Time-Synced On-Screen Video Comments (Danmu) Accessible to Blind and Low Vision Users via Multi-Viewer Audio Discussions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, 22 pages. https://doi.org/10.1145/3706598.3713496

[329] Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A. Miele, and Ilmi Yoon. 2020. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 47–60. https://doi.org/10.1145/3357236.3395433

[330] John Zimmerman and Jodi Forlizzi. 2014. *Research Through Design in HCI* (1 ed.). Springer New York, New York, NY, 167–189. https://doi.org/10.1007/978-1-4939-0378-8_8

[331] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704

[332] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. 2010. An analysis and critique of Research through Design: towards a formalization of a research approach. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (Aarhus, Denmark) *(DIS '10)*. Association for Computing Machinery, New York, NY, USA, 310–319. https://doi.org/10.1145/1858171.1858228

Chapter A

Contributing Publications

[P1] Alexandre Nevsky. 2023. Object-Based Access: Enhancing Accessibility with Data-Driven Media. In Proceedings of the 2023 ACM International Conference on Interactive Media Experiences (IMX '23). Association for Computing Machinery, New York, NY, USA, 402–406.

Abstract: Audiovisual media is an integral part of many people's everyday lives. People with accessibility needs, especially people with complex accessibility needs, however, may face challenges accessing this content. This doctoral work addresses this problem by investigating how complex accessibility needs can be met by content personalisation by leveraging data-driven methods. To this end, I will collaborate with people with aphasia, a complex language impairment, as an exemplar community of people with complex accessibility needs. To better understand the needs of people with aphasia, I will use collaborative design techniques to meet the needs of end users. This will involve them in the design, development and evaluation of systems that demonstrate the benefits of content personalisation as an accessibility intervention. This paper outlines the background and motivation to this PhD, the work that has already been completed, and current planned future work.

Author contribution: An extended abstract submitted to the IMX 2023 Doctoral Consortium. The work was done by myself, with feedback from my supervisors.

[P2] Alexandre Nevsky, Timothy Neate, Elena Simperl, and Radu-Daniel Vatavu. 2023. Accessibility Research in Digital Audiovisual Media: What Has Been Achieved and What Should Be Done Next? In Proceedings of the 2023 ACM International Conference on Interactive Media Experiences (IMX '23). Association for Computing Machinery, New York, NY, USA, 94–114.

Abstract: The consumption of digital audiovisual media is a mainstay of many people's lives. However, people with accessibility needs often have issues accessing this content. With a view to addressing this inequality, there exists a wide range of interventions that researchers have explored to bridge this accessibility gap. Despite this work, our understanding of the capability of these interventions is poor. In this paper, we address this through a systematic review of the literature, creating a dataset of and analysing $N = 181$ scientific papers. We have found that certain areas have accrued a disproportionate amount of attention from the research community – for example, blind and visually impaired and d/Deaf and hard of hearing people account for 93.9% of papers ($N = 170$). We describe challenges researchers have addressed, end-user communities of focus, and interventions examined. We conclude by evaluating gaps in the literature and areas that could use more focus on in the future.

Author contribution: I led this systematic review, with guidance from my supervisors and colleagues. I created the dataset, manually coded it, and created the codebook in discussion with my supervisors. I wrote the published paper, with feedback from my supervisors and colleagues.

[P3] Alexandre Nevsky, Timothy Neate, Elena Simperl, and Madeline N Cruice. 2024. Lights, Camera, Access: A Closeup on Audiovisual Media Accessibility and Aphasia. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 912, 1–17.

Abstract: The presence of audiovisual media is a mainstay in the lives of many, increasingly so with technological progress. Accessing video and audio content, however, can be challenging for people with diverse needs. Existing research has explored a wide range of accessibility challenges and worked with disabled communities to design technologies that help bridge the access gap. Despite this work, our understanding of the challenges faced by communities with complex communication needs (CCNs) remains poor. To address this shortcoming, we present the first study that investigates the viewing experience of people with the communication impairment aphasia through an online survey ($N = 41$) and two focus group sessions (N=10), with the aim of understanding their specific access challenges. We find that aphasia significantly impact viewing experience and present a taxonomy of access barriers and facilitators, with suggestions for future research.

Author contribution: I created the online survey and planned the focus group sessions, with advice from my supervisors. I conducted the transcription of focus group audio recordings and conducted the thematic analysis. I wrote the paper, with feedback from my supervisors and colleagues.

[P4] Alexandre Nevsky, Filip Bircanin, Madeline N Cruice, Stephanie Wilson, Elena Simperl, and Timothy Neate. 2024. "I Wish You Could Make the Camera Stand Still": Envisioning Media Accessibility Interventions with People with Aphasia. In Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 46, 1–17.

Abstract: Audiovisual media is integral to modern living, yet is not always accessible to all. Modern accessibility interventions, such as subtitles, support many, however, communities with complex communication needs are largely unconsidered. In this work, we envision future accessibility interventions from the ground up with one such community – people with aphasia. Over two workshops and a probe activity, we problematise the space of audiovisual consumption by people with aphasia, and co-envision directions for development in accessible audiovisual media. From low-fi diegetic prototypes to mid-fidelity solutions, we explore new visions of accessibility interventions for complex communication needs – notably enabling high levels of content manipulation and personalisation. Our findings raise open questions and set

directions for the research community in developing accessibility interventions for audiovisual media to support users with diverse needs in accessing audiovisual content.

Author contribution: I planned and ran the co-design sessions with the help of colleagues. I created all the prototypes based on the sessions, with feedback from colleagues. I led the transcription and data analysis, with feedback from colleagues. I led the paper writing, with feedback from colleagues.

[P5] Alexandre Nevsky, Filip Bircanin, Elena Simperl, Madeline N Cruice, and Timothy Neate. 2025. To Each Their Own: Exploring Highly Personalised Audiovisual Media Accessibility Interventions with People with Aphasia. In Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25). Association for Computing Machinery, New York, NY, USA, 1826–1843.

Abstract: Digital audiovisual media (e.g. TV, streamed video) is an essential aspect of our modern lives, yet it lacks accessibility – people living with disabilities can experience significant barriers. While accessibility interventions can improve the access to audiovisual media, people living with complex communication needs have been under-represented in research and are potentially left behind. Future visions of accessible digital audiovisual media posit highly personalised content that meets complex accessibility needs. We explore the impact of such a future by conducting bespoke co-design sessions with people with aphasia – a language impairment common post-stroke – creating four highly personal accessibility interventions that leverage audiovisual media personalisation. We then trialled these prototypes with 11 users with aphasia; examining the effects on shared social experiences, creative intent, interaction complexity, and feasibility for content producers. We conclude by critically reflecting on future implementations, raising open questions and suggesting future research directions.

Author contribution: I planned and ran the co-design sessions with the help of colleagues. I created all the prototypes based on the sessions, with feedback from colleagues. I led the transcription and data analysis, with feedback from colleagues. I led the paper writing, with feedback from colleagues.