Towards Audio Personalization for Accessible Digital Media

Dhruv Jain
Computer Science and Engineering
University of Michigan
Ann Arbor, Michigan, USA
profdj@umich.edu

Jason Miller
Computer Science and Engineering
University of Michigan
Ann Arbor, Michigan, USA
jmiller@umich.edu

Abstract

Digital media is increasingly audio-rich, yet much of its sound content remains inaccessible for deaf and hard-of-hearing (DHH) individuals. While prior work has focused on captioning and sound recognition, little research has explored how sound itself can be transformed to better align with hearing needs, preferences, and contexts for people with partial hearing. In this paper, we present findings from a formative study with 24 DHH participants that examines their experiences and unmet needs around digital media audio. Participants emphasized the importance of features such as selective speaker amplification, contextual sound control, semantic summarization, and adaptive personalization. Based on these insights, we introduce the ReMediaTion Framework, a layered model that articulates user goals, audio transformation dimensions, interaction strategies, contextual modulation, and expressive engagement. Our work provides a foundation for designing future audio accessibility systems that go beyond substitution, empowering DHH users to reshape how they experience and interpret sound.

CCS Concepts

• **Human-centered computing** → **Accessibility technologies**.

Keywords

Accessibility; audio personalization; generative AI; source separation; sound processing; music; dialogue; podcast; streaming; video.

ACM Reference Format:
Dhruv Jain and Jason Miller. 2025. Towards Audio Personalization for Accessible Digital Media. In *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025, Canberra, ACT, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3716553.3750774

1 Introduction

Digital media spanning podcasts, music, and video streaming has become a dominant form of information, storytelling, and entertainment. While Deaf and hard-of-hearing (DHH) individuals engage with media through diverse modalities—including captions, transcripts, sign language, and visual storytelling [21, 25]—sound itself often remains an underexplored accessible layer, particularly for individuals with partial hearing. This is especially true for non-speech audio such as background music, spatial cues, ambient sounds, and emotional tones. These sonic elements frequently carry narrative weight and shape the aesthetic and affective experience of media. Yet, current systems offer limited ways for users with residual hearing to access, personalize, or interpret these elements based on their individual hearing profiles or preferences.

Recent accessibility efforts have primarily focused on improving automatic captions [8, 16] or environmental sound recognition [13, 14]. While these tools provide functional access to spoken content or real-world audio events, they rely on sound-to-text substitution. Meanwhile, commercial media platforms have introduced early sound transformation features—such as "dialogue boost" [20]—but these offer generic modifications and are often insufficient for the nuanced needs of DHH individuals with partial hearing.

We argue that an untapped opportunity lies in sound-to-sound personalization: allowing DHH users with residual hearing to modify how audio is presented to better match their hearing preferences, media context, and narrative engagement goals. For instance, a user may wish to amplify a soft-spoken character, reduce ambient interference, highlight a specific instrument, or receive high-level sound summaries in lieu of complex acoustic layering. With emerging advances in generative audio, voice separation, and semantic audio analysis [3, 15, 19], such capabilities are increasingly feasible—and potentially transformative—for inclusive media design.

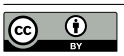
In this paper, we present findings from a formative study with 24 DHH individuals with partial hearing, aimed at uncovering their experiences and unmet needs around digital media sound. Participants described a wide range of desired features, including selective speaker enhancement, emotional annotation of background audio, adaptive personalization, and scene-level sound summaries.

Building on these insights, we propose the *ReMediaTion* framework—a layered conceptual model that articulates user goals, sound transformation dimensions, interaction strategies, and contextual factors for DHH-centered audio personalization. Rather than treating sound access as a static feature, the framework emphasizes adaptability, narrative relevance, and expressive control.

In summary, this short paper contributes: (1) empirical insights from interviews with 24 DHH participants with partial hearing on audio accessibility needs in digital media, and (2) the ReMediaTion Framework, a design model to guide the development of future AI tools for personalized sound transformations. Together, this work expands the accessibility research agenda beyond transcription and toward reshaping how sound itself can be experienced, interpreted, and personalized by DHH individuals with partial hearing.

2 Related Work

We review related work on DHH sound experiences, sound accessibility tools, and emerging audio personalization techniques.



ICMI '25, Canberra, ACT, Australia

ACM ISBN 979-8-4007-1499-3/25/10
https://doi.org/10.1145/3716553.3750774

2.1 DHH Experiences with Sound in Media

Accessibility research has increasingly recognized that DHH individuals' relationships with sound go beyond audibility—encompassing narrative interpretation, emotional resonance, and cultural meaning. Studies by Alonzo et al. [1] and May et al. [17] highlight that conventional captions often neglect or flatten non-speech sounds, rendering them as generic (e.g., "[music]") or overly prescriptive, which DHH viewers found immersion-breaking. Salazar [24] further critiques how fixed captions, even when stylized (e.g., in *Stranger Things*), impose fixed interpretations that limit viewer autonomy.

Despite these insights, prior work has not directly explored how DHH individuals—particularly those with partial hearing—wish to experience, control, or personalize sound in digital media. This population includes a wide range of subgroups, such as hard-of-hearing, late-deafened, and culturally deaf individuals with residual hearing. Our work addresses this gap through qualitative interviews with 24 DHH participants with partial hearing, revealing diverse needs for sound-to-sound transformation across podcasts, music platforms, and streaming video.

2.2 Sound Accessibility Tools

Most accessibility tools for DHH users rely on substituting sound with visual or textual forms. Real-time captioning systems improve speech accessibility, while captioning standards (e.g., W3C [32]) advocate for including non-speech cues such as background sounds. Recent work has explored augmenting captions using color, font, and emoticons to convey tone or affect [8, 16], but these modifications affect caption display rather than how sound is experienced.

In real-world environments, systems such as SoundWatch [14] and ProtoSound [13] help users detect environmental events. Though valuable for safety and awareness, these systems translate sound into text notifications to support functional access rather than delivering richer or narrative experiences. Hearing technologies like hearing aids offer noise suppression and frequency tuning [29], but do not allow user-driven, selective control of sound sources.

In media contexts, Netflix's "Dialogue Boost" [20] allows speech amplification, and some video games offer selective audio controls (e.g., music vs. effects sliders) [4, 5, 31] but these are fixed, one-size-fits-all presets. In contrast, our work explores dynamic, personalized, and semantic sound transformations that DHH users can shape to match narrative context, emotion, and listening needs.

2.3 Emerging Audio Personalization Techniques

In the broader media and audio research community, recent advances have introduced personalized audio experiences. Tools like Sonarworks SoundID [27] and Mimi Hearing [30] let users customize EQ profiles of headphones to match their hearing abilities. Hearing aids also increasingly incorporate adaptive algorithms that adjust to environmental conditions [29].

In digital media, sound separation models like DeepFilterNet [26] and VoiceFilter [19] enhance speech in noisy environments, while personalized sound zones (e.g., [28]) allow spatial separation of audio streams for individuals sharing the same space. Meanwhile, semantic audio analysis [3, 15] have enabled recognition and labeling of complex acoustic scenes, offering new possibilities for context-aware audio control. Commercial platforms such as Dolby and Spotify have begun offering personalized audio profiles.

However, these personalization tools focus on global sound properties (e.g., loudness, frequency balance) or single-task enhancement (e.g., speech clarity). They do not yet support selective source transformations, narrative-aware adjustments, or emotionally meaningful reinterpretation across media types.

Our framework builds on this foundation but is novel in four key ways. First, it emphasizes fine-grained control: users can selectively adjust voices, instruments, ambient sounds, and more. Second, it proposes semantic transformation, enabling contextually enriched or emotion-driven reinterpretations of sounds (e.g., "anxious footsteps," "melancholic piano") beyond simple volume modifications. Third, it supports dynamic interaction through conversational interfaces, adaptive profiles, and context-aware behaviors rather than relying solely on static presets. Finally, it aligns with DHH conceptualizations of sound, building on prior work [12, 23] to advocate for approaches that address not only hearing loss but also the diverse ways DHH individuals experience sound.

3 Formative Study

To investigate the sound personalization needs of DHH individuals in digital media, we conducted a semi-structured interview.

Participants. We recruited 24 DHH participants (13 women, 9 men, 2 non-binary) with partial hearing loss through email lists, social media, and snowball sampling. Participants had an average age of 37.4 years (*SD*=13.8, range=21–64). Twelve reported severe hearing loss, seven profound, three moderate, and two moderate-to-severe. Ten participants identified as deaf, 9 as hard of hearing, and 5 as (capital 'D') Deaf [6, 18]. Eighteen participants used hearing devices (11 hearing aids, 7 cochlear implants). Regarding media accessibility practices, all participants reported using subtitles, and three reported using the dialogue boost feature.

Study Preparation. Prior to the study, we compiled a set of potential audio personalization features based on three sources: (1) prior accessibility research, (2) the lived experiences of our interdisciplinary team (including audio engineers, DHH individuals, and accessibility researchers), and (3) informal reviews of 13 blog posts and 28 Reddit threads discussing media experiences of DHH users. The compiled list included features such as global sound adjustments (e.g., volume, speed), selective sound source modification (e.g., enhancing individual voices, suppressing background music), and sound summarization (e.g., scene-level audio summaries).

Procedure. Study sessions were conducted remotely over Zoom and lasted approximately one hour. Participants were offered accommodations of their choice: eight opted for a real-time captioner, three requested a sign language interpreter, and the remainder relied on Zoom's auto-captioning. The semi-structured interviews covered participants' current strategies and challenges for accessing sound in digital media, reactions to proposed sound personalization features, and future ideas for tailoring media audio to better match their hearing abilities, preferences, and goals. Participants were asked to consider multiple forms of digital media, including podcasts, music apps (e.g., Spotify), and video content (e.g., Netflix, broadcast TV). We used probes to explore both functional and experiential aspects of sound accessibility (e.g., cognitive fatigue,

emotional salience, narrative understanding). All sessions were audio-recorded, and participants were compensated 50 USD.

Data Analysis. We transcribed all interviews verbatim and conducted an applied thematic analysis [11]. One researcher initially skimmed the transcripts to develop a preliminary codebook, which was refined through iterative coding. The final codebook consisted of a two-level hierarchy with 6 first-level codes and 26 second-level codes. To ensure reliability, a second researcher independently applied the final codebook across all transcripts. Interrater reliability, measured using Cohen's Kappa [7], averaged 0.87, and raw agreement was 96.2%. Discrepancies were resolved through consensus.

3.1 Findings

Our analysis revealed tensions in how DHH individuals with partial hearing experience sound accessibility in digital media today, and what they envision for future personalization. Rather than discrete feature requests, participants' narratives reflect layered tensions between access and overload, control and automation, interpretation and autonomy. We organize our findings into seven themes, annotated where applicable by relevant media type (e.g., video, music, podcast).

3.1.1 ***Negotiating Access: Between Audibility and Overload.*** Participants consistently expressed frustration with current approaches to sound accessibility in media. While captions and transcripts offered basic access to speech, they rarely conveyed the emotional and narrative dimensions carried by soundscapes. P4 expressed: *"The words are there, but the feeling is missing. I miss the storm, the heavy footsteps, the music building up tension."*

Many participants noted that simply increasing the overall volume often overwhelmed or distorted important details, leading to sensory overload rather than meaningful access: *"Boosting everything just creates noise. I want richness, not a wall of sound."* (P7)

These frustrations echo prior critiques of captioning saturation [1, 17], but here they emerge around auditory density itself. Participants envisioned *adaptive richness*—systems that could modulate auditory complexity based on scene intensity, listener fatigue, or viewing mode. This request was particularly relevant for cinematic video and music streaming contexts, extending ideas of attention-aware accessibility [22] into sound design.

3.1.2 ***Craving Fine-Grained and Selective Sound Control.*** Participants emphasized a strong desire for fine-grained control over individual sound sources, rather than global audio adjustments. This was noted across all media types: podcasts, streaming video, and music. They envisioned selectively amplifying, suppressing, or modulating dialogue, music, ambient noise, or off-screen sounds based on situational needs: *"Sometimes I want to hear only the whispers, not the background chatter. Other times, I want the music to lead."* (P10)

This aspiration aligns with advances in sound separation technologies [19], but participants stressed that contextual importance, not just acoustic properties, should drive prioritization.

Several also pointed to the need for dynamic per-scene adaptation: more detail during intense scenes, minimalism during dialogue-heavy moments. For example, P12 said: *"Maybe I care about the ocean sounds in a romance movie, but in an action scene, just the dialogue matters. Let me pick."*

Importantly, participants did not just want to adjust volume—they wanted control over how sounds were layered in time. Many described a need for temporal rebalancing, such as automatically lowering music during dialogue, especially in podcasts and video. As P6 explained: *"When someone talks, I want the music to duck. Step back for a bit, then come back."*

However, these desires for control were tempered by concerns about interaction burden. Participants advocated for lightweight overrides—quick toggles, voice commands, or learned preferences—to maintain agency without constant micromanagement.

3.1.3 ***Situated Sound Preferences: Genre, Context, and Listening States.*** Participants' sound personalization needs varied substantially by genre and listening context. For video content, P15 said: *"In documentaries, just give me facts. In horror movies, I want the whole storm, the creaking floorboards, the creepy music."*

For music or ambient media, emotional immersion and background modulation were more relevant. P10 added: *"Friday night? Give me full cinematic experience. Tuesday after work? Bare minimum, please."*

These findings extend context-aware media personalization frameworks [10], highlighting the need for fluid, dynamic adaptation even within the same user. Many participants also envisioned adaptive profiles that could learn and evolve across devices: *"learn that I always want whispers boosted in thrillers but not in sitcoms."* (P7)

While automation was welcomed, participants emphasized that manual overrides must remain effortless to protect user agency.

3.1.4 ***From Sound Recognition to Emotional Resonance.*** For video and music platforms, a major insight was that participants did not simply want to hear sounds—they wanted to feel and interpret them meaningfully, in order to catch emotional tone, narrative shifts, and atmospheric cues. P5 observed: *"It's not the thunder itself. It's how the thunder changes the mood—how it says something bad is coming."*

Several participants longed for semantic overlays—subtle audio transformations that convey emotional intent ("tense strings rising", "melancholic piano fade") rather than literal labels ("rain", "piano").

At the same time, not everyone welcomed emotional reinterpretations. Some participants worried about over-interpretation by AI systems, fearing that emotional labeling might intrude upon personal meaning-making. P3 cautioned: *"Don't tell me what to feel. Give me clues, but let me experience it my way."*

This tension between emotional guidance and viewer autonomy surfaced repeatedly, echoing broader accessibility debates about preserving interpretive freedom [2], especially in narrative video genres like thrillers or dramas. This suggest that customizable emotional granularity—where users can choose how much interpretive help they receive—is crucial for future systems.

3.1.5 ***Trust, Transparency, and Collaborative Personalization.*** Concerns about system opacity and misinterpretation spanned all media types. For example, P9 cautioned: *"I don't want the system deciding that some music isn't important and muting it. Maybe that music was the whole emotional point."*

More broadly, participants demanded transparency and collaborative control. They proposed features such as: (1) previewing planned transformations, (2) allowing reversible changes, (3) querying the rationale behind adjustments, and (4) showing "confidence indicators" for emotional tagging. These requests resonate with emerging frameworks in explainable AI (XAI) [9] and collaborative AI [33], but here grounded specifically in preserving interpretive agency during media consumption.

3.1.6 ***Lack of Representation and Cultural Sensitivity.*** Participants expressed broader concerns about the lack of DHH perspectives in designing current media accessibility systems. Several felt that sound modifications often reflected hearing-centric assumptions about what mattered: *"They think boosting dialogue is enough. But sometimes I want to hear the music, not the words."* (P14)

This concern was particularly noted in TV and film soundtracks. Participants called for culturally sensitive systems that support emotional and spatial understandings of sound often overlooked in mainstream accessibility [12, 23].

3.1.7 ***Emerging Aspirations: Playfulness, Texture Transformation, Sound Summarization, and Conversation Interfaces.*** Beyond functional accessibility, participants also imagined future systems that go beyond functional access to enable playful and expressive engagement with sound.

For example, they envisioned thematic sound modes such as a "detective mode" or "cinematic mode" for films and series, which could selectively enhance suspenseful cues or dramatic flair. Others proposed texture filters in music applications to reduce sensory discomfort—such as softening harsh metallic sounds—or rebalancing chaotic environments (e.g., cafés) into calming ambient textures. Some participants described whimsical "sound personas" (e.g., robotic, cinematic, poetic) that could transform the auditory tone to match mood or aesthetic preferences. In both podcasts and narrative video, participants expressed interest in scene-level *sound summarization* , where key auditory events and emotional arcs (e.g., "a storm builds while a woman sobs quietly") could be replayed or skimmed. Finally, many advocated for real-time conversational control, describing natural language commands like "focus on the flute" (P21, music) or "mute laughter but keep dialogue" (P14, video) as intuitive ways to reframe the audio experience. These ideas reflect a broader aspiration for dynamic, genre-aware, and multimodal personalization of sound.

4 The ReMediaTion Framework

Grounded in our findings, we propose *ReMediaTion*—a framework to guide future sound-to-sound personalization systems for DHH users. Rather than treating sound accessibility as a static adjustment (e.g., global amplification), ReMediaTion conceptualizes personalization as a dynamic, layered process that responds to both narrative meaning and situational needs.

In contrast to prior media accessibility work for DHH users [8, 16], which focuses on sensory substitution, ReMediaTion introduces audio personalization to address diverse user needs. In contrast to prior audio research techniques, such as source separation [19, 26] and semantic tagging [15], which focus on technical accuracy, ReMediaTion reorients these techniques to support DHH-centered narration, emotions, and situated meaning.

The framework comprises of six interdependent layers derived through thematic coding of participant interviews (see Section 3). Each layer synthesizes recurring needs expressed by participants and maps to user tensions identified in our analysis.

4.1 User-Centered Goals

Future systems should support diverse motivations for audio personalization beyond simple audibility:

- **Comprehension:** Understand speech, music, and ambient meaning.
- **Focus:** Elevate relevant sounds, suppress distractions.
- **Comfort:** Manage cognitive and sensory load dynamically.
- **Inclusion:** Participate fully in narrative and emotional arcs.
- **Exploration:** Engage playfully or reflectively with sounds.
- **Control:** Maintain agency over how sound behaves and evolves.

4.2 Sound Transformation Dimensions

Support fine-grained, situational control across multiple auditory properties:

- **Volume targeting:** Adjust loudness for specific sources (e.g., boost whispers in drama scenes or podcasts).
- **Frequency/EQ:** Tune frequencies for individual hearing profiles (e.g., enhance high-frequency instruments in music).
- **Layer separation:** Differentiate dialogue, ambient noise, and music streams (especially useful for video).
- **Texture rewriting:** Modify the emotional quality of sounds (e.g., soften harsh metallic tones in music).
- **Narrative salience:** Amplify sounds critical to plot (e.g., tension-building audio in video scenes).
- **Temporal flow:** Dynamically sequence overlapping sounds (e.g., duck music during podcast narration or movie dialogue).
- **Semantic overlays:** Add emotional or descriptive sound summaries (e.g., "ominous rumble" for narrative video or podcasts).

4.3 Interaction Modalities

Support multimodal and lightweight control mechanisms:

- **Direct manipulation:** Sliders, presets, and toggles (e.g., quickly mute laugh tracks in sitcoms).
- **Conversational queries:** Natural language interactions (e.g., "focus on the flute" in music or "mute crowd noise" in sports).
- **Automatic adaptation:** Systems learn behaviors over time across different media formats.
- **Playful interfaces:** Optional whimsical modes (e.g., detective or cinematic filters for video content).

4.4 Contextual Modulation

Support shifting preferences with media genre, scene dynamics, user state, and environment:

- **Genre:** Minimalism for news or podcasts, richness for horror or fantasy video.
- **Scene type:** Dialogue vs. action vs. ambient sequences (especially relevant in film and TV).

- **User state:** Fatigue, focus, emotional mood fluctuations (applies across all media types).
- **Listening environment:** Noise levels, device type (headphones, TV, mobile).
- **Cultural context:** Different conceptualizations of sound across DHH identities (e.g., spatial emphasis in video, emotional resonance in music).

4.5 Personalization and Learning

Learn preferences and evolve collaboratively with users:

- **Hearing profiles:** Configure based on audiograms.
- **Preference memory:** Learn recurrent actions (e.g., always mute laugh tracks in video).
- **Scenario presets:** Focus mode, relaxation mode.
- **Intersectional adaptation:** Support users with co-occurring sensory, cognitive, or processing needs.

4.6 Expressive and Reflective Layers

Support not just accessibility, but also self-expression:

- **Sound personas:** Thematic rewrites ("detective" or "dreamy").
- **Narrative summarization:** Story-driven scene recaps.
- **Curated soundscapes:** Transform busy environments into calming ambience.

4.7 Expressive and Reflective Layers

Support not just accessibility, but also self-expression:

- **Sound personas:** Thematic rewrites for video and narrative audio (e.g., "detective" or "dreamy").
- **Narrative summarization:** Story-driven scene recaps for podcasts and video (e.g., "storm builds while a woman sobs").
- **Curated soundscapes:** Transform busy environments into calming ambience (especially relevant for music and podcasts).

5 Limitations and Future Work

This paper focused on DHH individuals with partial hearing, who may benefit most directly from sound-to-sound personalization. While our findings provide meaningful insights into this population's needs, we recognize that the DHH community is diverse. Experiences of profoundly Deaf individuals—especially those who primarily use sign language—may differ. Future work should explore how layered sound personalization can complement existing visual accessibility tools, such as captioning or sign language interpretation, and investigate hybrid approaches co-designed with Deaf signers to support richer, multimodal engagement.

Second, while we proposed the ReMediaTion framework grounded in qualitative findings, we did not implement or evaluate a working prototype. Our goal was to first establish a conceptual and user-centered foundation. Nevertheless, implementation and evaluation are critical next steps. Many components of the framework are technically feasible today using emerging tools in machine listening and generative audio. For example, selective voice amplification and ambient suppression can be achieved through source separation models such as DeepFilterNet [26] and VoiceFilter [19]; semantic overlays and emotional labeling can be supported by pretrained audio tagging models such as PANNs [15]; and temporal rebalancing of music and dialogue may be accomplished with dynamic signal processing algorithms. Future prototypes should explore how these tools can be integrated to support dynamic personalization while preserving user agency and minimizing cognitive load.

As these systems evolve, it will also be important to consider transformation risks such as emotional misinterpretation, model hallucinations, or distortion of narrative meaning. Future implementations should incorporate safeguards such as user-controlled previews, confidence estimation, and transparent feedback to support responsible audio transformations.

Third, while our sample size of 24 participants may appear modest, it is relatively large for qualitative HCI research, especially studies involving marginalized groups. The sample included diverse gender identities, hearing profiles (moderate to profound), and cultural identities (deaf, Deaf, hard of hearing). While statistical generalization is outside the scope of this study, we observed qualitative trends: participants with more severe hearing loss often prioritized emotional sound summaries and texture transformation, while those with moderate hearing loss emphasized fine-grained control and clarity. These insights can inform future work on adaptive profiling and personalization based on hearing level, listening environment, and media genre.

Finally, while our study included multiple forms of digital media—such as video, podcasts, and music—most examples discussed by participants centered on video content. Future work should examine media-specific use cases more systematically, and prototype systems across these formats to evaluate differences in interaction design, content structure, and listening goals.

Safe and Responsible Innovation Statement. This study was approved by our Institutional Review Board (IRB), and all data collection and consent procedures followed established guidelines.

6 Conclusion

We presented the ReMediaTion Framework, a user-centered design space for sound-to-sound personalization, grounded in the needs of DHH media consumers. ReMediaTion was informed by our study with 24 DHH participants that revealed opportunities for selective tuning, emotional sound interpretation, and adaptive profiles. By reframing accessibility beyond mere comprehension, and toward dynamic sound personalization, our work informs the development of future systems that empower DHH individuals to experience, navigate, and reshape media sounds in richer, more personal ways.

Acknowledgments

We thank Jeremy Zhengqi Huang for his contributions to the initial problem formulation.

References

[1] Oliver Alonzo, Hijung Valentina Shin, and Dingzeyu Li. 2022. Beyond subtitles: captioning and visualizing non-speech sounds to improve accessibility of user-generated videos. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[2] Meryl Alper. 2017. *Giving voice: Mobile communication, disability, and inequality*. MIT Press.

[3] Daniele Barchiesi, Dimitrios Giannoulis, Dan Stowell, and Mark D Plumbley. 2015. Acoustic scene classification: Classifying environments from the sounds they produce. *IEEE Signal Processing Magazine* 32, 3 (2015), 16–34.

[4] Ben Bayliss. 2022. Fortnite Accessibility — Menu Deep Dive. Can I Play That? https://caniplaythat.com/2022/04/26/fortnite-accessibility-menu-deep-dive Accessed: 2024-02-18.

[5] Xinyun Cao and Dhruv Jain. 2024. SoundModVR: Sound Modifications in Virtual Reality to Support People who are Deaf and Hard of Hearing. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.

[6] Anna Cavender and Richard E Ladner. 2008. Hearing impairments. *Web accessibility: A foundation for research* (2008), 25–35.

[7] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[8] Caluã de Lacerda Pataca, Saad Hassan, Nathan Tinker, Roshan Lalintha Peiris, and Matt Huenerfauth. 2024. Caption royale: Exploring the design space of affective captions from the perspective of deaf and hard-of-hearing individuals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[9] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O Riedl. 2019. Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In *Proceedings of the 24th international conference on intelligent user interfaces*. 263–274.

[10] Krzysztof Z Gajos, Daniel S Weld, and Jacob O Wobbrock. 2010. Automatically generating personalized user interfaces with Supple. *Artificial intelligence* 174, 12-13 (2010), 910–950.

[11] Greg Guest, Kathleen M MacQueen, and Emily E Namey. 2011. *Applied thematic analysis*. sage publications.

[12] Jeremy Zhengqi Huang, Reyna Wood, Hriday Chhabria, and Dhruv Jain. 2024. A Human-AI Collaborative Approach for Designing Sound Awareness Systems. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–11.

[13] Dhruv Jain, Khoa Huynh Anh Nguyen, Steven M. Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehlich. 2022. Protosound: A personalized and scalable sound recognition system for deaf and hard-of-hearing users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[14] Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. SoundWatch: Exploring smartwatch-based deep learning approaches to support sound awareness for deaf and hard of hearing users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[15] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D Plumbley. 2020. Panns: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), 2880–2894.

[16] Daniel G Lee, Deborah I Fels, and John Patrick Udo. 2007. Emotive captioning. *Computers in Entertainment (CIE)* 5, 2 (2007), 11.

[17] Lloyd May, So Yeon Park, and Jonathan Berger. 2023. Enhancing Non-Speech Information Communicated in Closed Captioning Through Critical Design. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[18] Matthew S Moore and Linda Levitan. 2016. *For hearing people only*. Deaf Life Press.

[19] Hannah Raphaelle Muckenhirn, Ignacio Lopez Moreno, John Hershey, Kevin Wilson, Prashant Sridhar, Quan Wang, Rif A Saurous, Ron Weiss, Ye Jia, and Zelin Wu. 2019. Voicefilter: Targeted voice separation by speaker-conditioned spectrogram masking. In *conference of the international speech communication association*.

[20] Netflix. 2023. Dialogue Boost Feature on Netflix. https://www.netflix.com.

[21] Aline Remael. 2013. Media accessibility. In *Handbook of Translation Studies: Volume 3*. John Benjamins Publishing Company, 95–101.

[22] Claudia Roda and Julie Thomas. 2006. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior* 22, 4 (2006), 557–587.

[23] Russell S Rosen. 2007. Representations of sound in American deaf literature. *Journal of deaf studies and deaf education* 12, 4 (2007), 552–565.

[24] Savannah Salazar. 2022. *Wet Writhing, Eldritch Gurgling: A Chat With the Stranger Things Subtitles Team.* https://www.vulture.com/2022/07/stranger-things-subtitles-captions-team-interview.html

[25] Andreja Samčović. 2022. Accessibility of services in digital television for hearing impaired consumers. *Assistive Technology* 34, 2 (2022), 232–241.

[26] Hendrik Schroter, Alberto N Escalante-B, Tobias Rosenkranz, and Andreas Maier. 2022. DeepFilterNet: A low complexity speech enhancement framework for full-band audio based on deep filtering. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 7407–7411.

[27] Sonarworks. 2022. SoundID: Personalized Audio for Your Ears. https://www.sonarworks.com.

[28] NTT Sonority. 2023. Personalized Sound Zone (PSZ) Technology: Creating Private Acoustic Spaces. https://www.rd.ntt/e/research/JN202404_25732.html Accessed: 2025-05-01.

[29] Nafisa Zarrin Tasnim, Aoxin Ni, Edward Lobarinas, and Nasser Kehtarnavaz. 2024. A review of machine learning approaches for the personalization of amplification in hearing aids. *Sensors* 24, 5 (2024), 1546.

[30] Mimi Hearing Technologies. 2021. Mimi Hearing: A Customized Hearing Experience. https://www.mimi.io.

[31] TS4 Sound Tool. 2024. Sims 4 Modding Wiki. https://sims-4-modding.fandom.com/wiki/TS4_Sound_Tool Accessed: 2024-03-21.

[32] Web Accessibility Initiative (WAI). 2024. Captions/Subtitles | Web Accessibility Initiative (WAI) | W3C. https://www.w3.org/WAI/media/av/captions/.

[33] Zijie J Wang, Dongjin Choi, Shenyu Xu, and Diyi Yang. 2021. Putting humans in the natural language processing loop: A survey. *arXiv preprint arXiv:2103.04044* (2021).