Augmenting TV Viewing using Acoustically Transparent Auditory Headsets

Mark McGill, Florian Mathis, Mohamed Khamis, Julie Williamson
first.last@glasgow.ac.uk
School of Computing Science
University of Glasgow, Scotland, UK

ABSTRACT

This paper explores how acoustically transparent auditory headsets can improve TV viewing by intermixing headset and TV audio, facilitating personal, private auditory enhancements and augmentations of TV content whilst minimizing occlusion of the sounds of reality. We evaluate the impact of *synchronously mirroring* select audio channels from the 5.1 mix (dialogue, environmental sounds, and the full mix), and *selectively augmenting* TV viewing with additional speech (e.g. Audio Description, Directors Commentary, and Alternate Language). For TV content, auditory headsets enable better spatialization and more immersive, enjoyable viewing; the intermixing of TV and headset audio creates unique listening experiences; and private augmentations offer new ways to (re)watch content with others. Finally, we reflect on how these headsets might facilitate more immersive augmented TV viewing experiences within reach of consumers.

CCS CONCEPTS

• **Human-centered computing**;

KEYWORDS

TV; Mixed Reality; Augmented Reality; Audio;

ACM Reference Format:
Mark McGill, Florian Mathis, Mohamed Khamis, Julie Williamson. 2020. Augmenting TV Viewing using Acoustically Transparent Auditory Headsets. In *ACM International Conference on Interactive Media Experiences (IMX '20), June 17–19, 2020, Cornella, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3391614.3393650

1 INTRODUCTION

Mixed Reality (MR) headsets, be they Augmented or Virtual Reality, have been discussed at length with respect to their capability to facilitate new, visually-dominated media experiences, from shared mixed-reality 360°video [29], to hyper-personalized and immersive TV [16], to physically immersive media [52]. There is however a comparatively new form of Mixed Reality headset that has the potential to enhance or augment the TV viewing experience non-disruptively, working within the shared and social environment TV is often consumed within. *Acoustically transparent* Auditory Headsets [31] are effectively the auditory equivalent of a visually-oriented Augmented/Mixed Reality headset, in that they allow the (passive or active) intermixing of both virtual and real-world audio. As a consequence, these headsets allow for listening to occur in ways that are more social and inclusive of the surrounding real-world soundscape e.g. listening to music whilst retaining an awareness of the conversation or activities of others in the same space. Notably compared to more traditional visually-oriented MR headsets, acoustically transparent auditory headsets are comparatively affordable, come in form factors that are (arguably) fashionable, and are intended to be wearable all day long [31], such as the Bose Frames (see Figure 1), a pair of glasses with directional speakers integrated into the frames, such that the ear canal is not occluded and virtual audio intermixes with real-world audio. If such headsets see significant adoption, we might expect to again see something of a (more modest) revolution in terms of ubiquitous computing, providing users with a personal, private soundscape that can augment, supplement or supplant their existing aural experiences - a rich new design space for applications to operate within.

We posit that acoustically transparent auditory headsets could be exploited to personally and privately enhance the TV viewing experience for wearers - regardless of the existing TV audio output audible to others, and without requiring the use of occlusive headphones which cut users off from the affective sounds and conversation of other viewers e.g. when watching with family in a living room context. However, to do so we must first understand 1) what audio (if any) from the existing TV viewing experience should be mirrored on to these auditory headsets, and what impact this mirroring and, its intermixing with environmental audio, has on the perception of the TV content; 2) the acceptability and potential utility of personal, private auditory augmentations; and 3) user attitudes toward usage of auditory headsets by ourselves and others during TV viewing.

In a user study (n=12) we firstly explore how we can personally enhance aural perception of the displayed video content, through synchronous playback of select audio channels on the acoustically

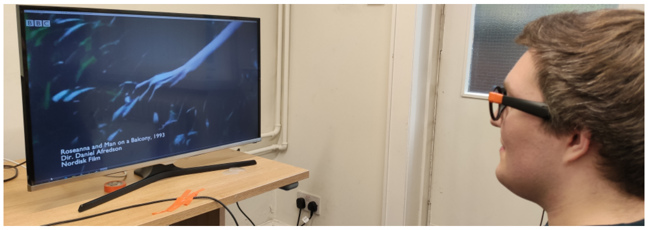

Figure 1: A TV viewer wears a pair of acoustically transparent Bose Frames, hearing the intermixing of both TV and Frames audio (shown lit, lighting was dimmed for study).

IMX '20, June 17–19, 2020, Cornella, Barcelona, Spain

transparent auditory headset, implicitly intermixed with the TV audio output. We look at whether we can enhance our transportation into the video content by playing the front left/right channels of 5.1 media (i.e. the non-speech channels) through the headset, creating an environmental "surround sound" experience, and our perception of dialogue by playing the front center channel (typically utilized predominantly for speech). Secondly, we look at personal augmentations of the underlying TV video content for which other viewers may not wish, or need, to attend to, specifically adding *Directors Commentary* e.g. when re-watching a film that others have not yet seen, *Audio Description* for those with visual impairments/differing needs, and *Alternate Languages* in multi-lingual settings – all without disrupting or altering the existing auditory experience of other non-headset viewers. Finally, we reflect on the disparate ways by which our TV viewing experiences might be additionally augmented in the future, given consumer adoption of such wearable personal audio. The use of acoustically transparent audio headsets meaningfully improves both the perception of the TV audio, and our ability to augment the TV experience through this private channel. We suggest that such headsets may see significant adoption in the future, and could be appropriated to pragmatically personally augment the TV viewing experience in engaging and immersive ways.

2 LITERATURE REVIEW

2.1 Acoustically Transparent Headwear

Personal, private audio has traditionally been enabled through the use of headphones - circumaural (over-the-ear) speakers, allowing for the rendering of high fidelity stereo/spatialized audio content. However, they do so at a cost, being the gatekeepers to auditory awareness of our surrounding reality. Predominantly, such headphones (either passively, or actively through noise cancellation technology) isolate us from our real-world acoustic surroundings, a sort of auditory cocoon [19] where external distractions can be blocked out. Consequently, the benefits of personal, private audio have yet to be entirely transposed to shared, social settings because the delivery mechanism typically prohibits continued speech/conversation-based social interactions and environmental awareness during viewing. In effect, we want to be available to the emotive, affective sounds of others, be it laughter, speech, groans and so on, as these sounds often enhance the viewing experience.

However, breakthroughs in battery technology, sensing and acoustics are resulting in new forms of personal audio wear which break free of prior limitations of circumaural headphones. *Acoustically transparent* Auditory Headsets [31] are auditory wearables that have the capability to, either actively or passively, intermix both real and virtual sounds, "not caus[ing] audible modification to the surrounding sound" [46]. Where Nomadic Radio [54] first explored this concept through wearable directional speakers, now we have a variety of circum- and supra-aural (on the ear) Hearables [8, 45] and Earables [25, 69] reaching consumer hands. Bone conductance headphones led the way in this regard, however their limited fidelity [67] inhibited uptake. More recently, passive acoustic transparency has been facilitated through wearable directional speakers integrated into glasses frames (e.g. Bose Frames [5], Amazon Echo Frames [17], Vue [68]) or wearables (e.g. Bose SoundWear [4], Amazon Echo Loop [17]). And battery technology is such that these headsets are now targeting wearable form factors such that they will have "the capability to infiltrate our everyday lives in ways that visually-oriented AR and VR headsets have yet to fully accomplish, be it for reasons of cost, technological capability, or social acceptability" [31]. Integrated Inertial Measurement Units (IMUs) have become somewhat commonplace, with many headsets utilizing an IMU for head orientation tracking, equivalent to 3DoF consumer VR, for personal spatialized exocentric rendering of sound.

McGill *et al.* [31] explored both the perception of acoustically transparent audio headwear indoors and outdoors, and their potential usage/adoption. Regarding perception, they found that "in select instances, participants noted that audio delivered via the acoustically transparent frames could appear indistinguishable from reality", with dramatic content noted as being more "part of the real world". Moreover, for speech content in particular it "helped in making speech appear more located in reality". Regarding usage and adoption, participants suggested that such headwear would make them more likely to listen to ambient audio, music, immersive spatialized and non-spatialized podcasts in particular. However, for general media/TV usage participants were largely neutral regarding their likelihood of usage, perhaps reflecting the lack of experience of such a use case. In addition to uses such as in collaboration proposed by Bauer *et al.* [1], McGill *et al.* [31] suggested that such headsets could be effectively integrated into a variety of day-to-day activities, particularly in TV where they might enable "private speech support for those with visual impairments; personal audio for multi-view TV; or even additional immersive spatial audio effects supplementing any display with surround sound capabilities".

2.2 Augmenting TV Audio

When discussing mixed reality augmentations of the TV content, previous research has focused predominantly on:

- The technical challenges of enacting **cross-device synchronization** during TV viewing [66] (e.g. for the *HbbTV* platform [10, 62]) and the **perception of synchronization** [64, 73] pioneered by Vinayagamoorthy *et al.*;
- Visually-oriented **synchronous experiences** across multiple devices [30, 40, 41, 66] and headsets [18];
- Visually-oriented **augmentations** e.g. enhancing accessibility for TV [65, 74] and immersive content [37], altering the surrounding environment for immersion [6], enhancing TV commercials [11], presenting other content [11, 27, 28, 63] and other augmentations around the display, such as rendering elements of the program in 3D, or rendering holograms of others [51];
- Augmentations delivered synchronously that are **not necessarily semantically linked** to the underlying TV content. For example, social TV [7] has repeatedly utilized synchronized TV experiences at-a-distance alongside additional audio communication channels to allow shared, at-a-distance TV viewing e.g. Harboe *et al.* [20] provided an open audio link between participants' homes, whilst McGill *et al.* [29] utilized synchronized Chromecasts with smartphone-based audio/visual communications, as well as MR presentations, to similar effect.

However, visually-oriented TV augmentations, primarily those reliant on AR headsets, have some caveats which suggest that we are as-yet some way from practical deployment and adoption. Visual consumer AR is currently costly, predominantly cumbersome (with some exceptions such as the NReal headset [42]) and of low-fidelity (e.g. in terms of field of view); the "fear of missing out" can be induced in non-AR equipped others in a shared setting [51]; and the very presence of visual AR arguably puts into question the necessity of the TV given content can be rendered virtually instead (although persuasive arguments can be made here regarding the superior luminosity and density of the TV, lower fatigue in attending to a physical display etc.).

Conversely, if we consider audio alone, there is a breadth of common aural augmentations that might be more practically facilitated in shared viewing environments through consumer acoustically transparent auditory headsets. The aural presentation is often recorded as a 5.1 mix, meaning there are 5 audio channels utilised to provide a spatialised audio experience [50] of the TV content: Front Center (FC), Front Left/Right (FL/FR), and Back Left/Right (BL/BR). In an idealised setup with 5 configured speakers (or appropriated devices as speakers [21]), the viewer in the sweet spot of the configuration will perceive the prescribed spatial audio experience. However, typically this audio will be downmixed to 2.1 and rendered on a soundbar or TV, with much of this spatial information lost in the process. A case could be made that this spatial sound could be individually personally rendered if the viewer is wearing an acoustically transparent auditory headset, providing a degree of spatial immersion whilst still allowing for social interactions, as well as enabling clear speech through the inverse of background ducking [60].

Layered on top of this audio mix are augmentations which semantically modify or re-present the auditory stream, optionally presented Voice-over-Voice with the intention of masking/replacing the existing information being conveyed [58, 60]. **Audio Description** [15, 44, 49] has become a commonly supported feature for assisting those with visual impairments, having been provided by the BBC for example since 2000, verbalising "changes of location, actions, facial expressions, gestures and so on [to] give the context and set the scene" [2]. **Directors Commentary** and other such "commentary" tracks (e.g. [47]) have also become increasingly popular and available, geared particularly toward repeat viewings of movies, providing new insights and information, or simply enhancing the entertainment value through new perspectives. And **Alternate Language** tracks have been commonplace since the advent of the DVD, allowing for dubbed viewing in the language of the viewer's choosing. Any of these such additions could be auralized per-person, if said viewer is wearing an Auditory Headset, without necessitating that the TV experience be altered (e.g. through subtitles or audio description) unnecessarily for other viewers. As Torcoli *et al.* concluded "only the personalization offered by object-base audio technologies" can meet the aural preferences of viewers [60], however delivery of this personalized audio remains an open question.

3 STUDY: AUGMENTING TV AUDIO

Given video content playing on a TV with audio output on the TV speakers, and one or more viewers wearing acoustically transparent headsets, this leads us to envision scenarios where the TV audio could be augmented/intermixed with personal, private synchronized audio:

Mirrored Listening You privately listen to the 5.1 mix of the content downmixed to stereo for your headset. This provides you with a personal, optimally spatialised audio experience, regardless of where you are seated in the room and regardless of the speaker configuration of the TV.

Enhanced Presence You privately listen to the left/right channels of the 5.1 mix. This gives you a spatialised sense of presence and atmosphere, as these channels are typically dedicated to environmental sounds and soundtracks.

Enhancing Dialogue You privately listen to the speech [43] or center channel from a 5.1 mix, giving you the ability to control the speech/dialogue volume independent of the TV speaker(s) volume, resolving arguments over volume.

(Re-)Watching Movies A friend notes they have never seen *The Shawshank Redemption*. You offer to watch it with them, despite having already seen it multiple times. This time, however, you listen to the Directors Commentary, piped directly to you via your Auditory Headset, without unduly interrupting your friend's experience.

Visual Impairments You privately enable Audio Description, allowing you to better attend to events without altering their experience of the content.

Language Barriers You sit down to watch *Amélie* with a French friend. Not wanting to disrupt their experience, you privately listen to the dubbed English version that overlays the French version and eliminates the need for subtitles.

Such scenarios are now eminently feasible, and emphasize the potential for acoustically transparent headsets in facilitating private synchronized auditory augmentations of the TV content, intermixed with both the background audio from the TV, and other real-world audio such as conversation with others. However, the impact of these intermixed audio streams on the perception of the content is not yet understood. Moreover, we can say nothing regarding the potential adoption and usage of such features, in particular considering the social acceptability of use in shared viewing scenarios. Consequently, this study examines the usage of personal, wearable acoustically transparent audio for augmenting TV content, to try to understand where the clear strengths of such a combination lie, where potential impediments to adoption and usage might arise.

3.1 Prototype Implementation

Our requirements were to play video content back, and selectively output synchronized audio to both the TV and the Bose Frames. For video playback, the experience was built in the Unity gaming engine [61]/C#, which gave us per-frame control of the video. For audio playback, we utilized the NAudio library [36], which allows audio content to be selectively played back on different audio devices in Windows. Playback of both streams was started concurrently, and then an offset (manually determined prior to the study by manipulating the delay in samples until audio on the Frames and TV appeared to converge) was applied to account for differences in end-to-end latency when playing audio back on the TV and the Frames. The Frames were hard wired to the PC to minimize variance in latency (e.g. due to Bluetooth) and connected using the USB debugging feature of the Bose AR SDK [34]. The result of this was that audio and video playback was highly synchronized i.e. when listened to side by side, the audio streams appeared to converge into one source.

3.2 Experimental Design

Based on our literature review, we had three Research Questions (RQs):

RQ1 What channels, if any, should be mirrored on the auditory headset during standard TV viewing to potentially enhance the auditory experience of typical TV content?

RQ2 To what extent can personal speech augmentations (e.g. directors commentary, alternate languages, audio description) be discerned over the existing TV speaker output?

RQ3 How likely might users be to use features such as mirrored audio or personal speech augmentations in a shared TV viewing context?

To answer these, we proposed a Study in two parts. Note that for all parts, TV clips were played for 120 seconds with the standard TV audio track playing throughout on the TV speakers, with participants answering questionnaires after each clip. The apparatus throughout was the same, as described in the implementation, and both parts were completed in the same session within-subjects. The study took place in a controlled lab environment, with 12 participants recruited (age 27.3±6.5, 7 male, 5 female, 8 non-native English speakers) from University mailing lists, each paid £8 for taking part. The study was approved by our University ethics committee.

3.2.1 Part 1 - Perception of Mirrored Audio. Firstly, we would examine RQ1, iterating through the potential permutations of TV and auditory headset (hereafter referred to as Frames for brevity) audio, with four conditions defined:

TV Only The TV only plays the full 5.1 mix, downmixed to stereo. This is the control condition, representative of existing viewing with a standard TV.

TV + Frames FLR As "TV Only", with the front left/right channels also duplicated in stereo on the Frames.

TV + Frames FC As "TV Only", with the front center channel duplicated in mono on the Frames.

Mirrored As "TV Only", with the 5.1 channels downmixed to stereo and duplicated on the Frames.

Note that we omit hypothetical conditions such as listening to the Frames only, as we are targeting shared, mixed use scenarios where not everyone may have personal, private audio, necessitating the TV always provide a full audio stream from its speakers. These conditions allowed us to compare a control condition (TV Only) against mirroring the FLR channels (hypothesized to be more immersive), the FC channel (hypothesized to make it easier to attend to speech dialogue) and the full 5.1 mix (both immersive and easier to attend to dialogue). Participants would experience these conditions for three different media types, chosen as they exhibited quite different auditory experiences:

Sport A 5.1 clip from Formula 1. The LR channels here present spatial audio of the cars.

Film A 5.1 clip from "Avengers: Endgame". The LR channels here present sounds of burning and other atmospheric effects, as well as elements of the soundtrack.

Documentary A 5.1 clip from "Seven Worlds, One Planet". The LR channels here present sounds of nature, the footsteps of animals etc.

For each clip, channels were extracted using `ffmpeg` [35] to separate mp3s: for FC using `-af "pan=mono|c0=FC"`, for FLR using `-af "pan=stereo|c0=FL|c1=FR"` and for full using `-vn -ac 2`. Order of the clips was counter-balanced, with condition order counter-balanced per-user, but shared across the three media types. After each combination of Condition and Media, participants answered 7-point Likert-type questions regarding:

Dialogue: "It was easy to pay attention to the speech/dialogue"

Enjoyment: "I enjoyed listening to the audio using the given headwear"

Clarity: "The audio was clear and of good quality"

Spatial Realism: From Begault [3], "Please rate the realism of the spatial reproduction / your sensation of sounds coming from different directions."

Transportation: The 5 questions from the transportation subscale of the Film IEQ [48].

Synchronization: "The TV audio was well synchronized with the Frames audio" (mirrored audio conditions only).

For these results, a repeated measures two-way *anovaBF* on $Condition * Content$ was performed using the *BayesFactor* package [38] following [39], with Bayes Factors reported, see [23] for interpretation. Bayes Factors between 3 to 20 constitute weak evidence between levels of the independent variables; 20+ constitutes strong evidence, a conservative interpretation [24]. All plots show 95% confidence intervals (red error bars) in line with [9]. Participants were also asked to rank the conditions in order of preference for each media type at the end of Part 1.

3.2.2 Part 2 - Perception of Speech Augmentations. Users trialled the following augmentations in turn, selected as they the following audio tracks optionally available:

Audio Description A clip from "The Art of Scandinavia"

Directors Commentary A clip from "Thor: Ragnarok" narrated by Director Taika Waititi

Alternate Language A clip from "Alien", presented in French on the TV, and English on the Frames.

The additional audio channels were extracted using `ffmpeg` to separate mp3s as previously discussed. For each clip, participants were asked Likert-type questions (Strongly Disagree to Strongly Agree) regarding:

Dialogue "It was easy to pay attention to the speech/dialogue"

Synchronization "The TV audio was well synchronized with the Frames audio"

3.2.3 End Survey. At the end of the study, participants then answered Likert-type questions for each audio scenario across the two parts, as well as an additional scenario regarding listening to non-TV content (e.g. podcasts/other unrelated media) regarding:

Acceptability How acceptable you would find it for those wearing Frames to privately augment their experience in this way, if others did not have this capaiblity?

Likelihood How likely would you be to use Frames to listen in this way?

Conversation How easy do you think it would be to hold a conversation with others and hear their responses?

Semi structured interviews were also conducted, regarding preferences and other anticipated uses of acoustically transparent audio during TV viewing.

4 RESULTS

4.1 Perception of Mirrored Audio

As can be seen in Table 1 there was predominantly strong evidence regarding the effect of *Condition*, with evidence to the contrary for *Media Type* and the *Media Type* * *Condition* interaction. In Figure 2, the effect of Condition across our measures was evident predominantly in the difference between TV Only (and less frequently TV + Frames FLR) and the other Conditions. Regarding **spatial realism** we can see that the TV Only condition, perhaps unsurprisingly, performed poorly, with the stereo speaker output of the TV failing to spatialise audio compared to the stereo headset. For **sound quality**, results reflect both the perception of better quality sound from the Frames, and also (given the similarity to *attending to dialogue*, and findings to be discussed from our interviews) that, despite the acoustic transparency, the audio of the FLR channels impaired hearing dialogue originating from the TV. It is also clear that the use of the Frames significantly improved the viewing experience, reflected best in the user **rankings**. **Synchronization** was largely rated positively (see Figure 3).

	Bayes Factor		
Measure	Media Type	Condition	*Interaction*
Spatial Realism	0.09	>100	0.12
Clarity	0.22	>100	0.07
Distinguishable	0.09	4.36	0.08
Dialogue	0.12	>100	0.06
Enjoyment	0.23	>100	0.17
Transportation	0.19	>100	0.20
Rank	0.07	>100	0.39

Table 1: Bayes Factors for Part 1. Green indicates a $BF_{10} > 20$ (strong evidence of effect). All questions were on a 7-point scale (0 to 6) apart from the Film IEQ Transportation factor, which was a sum score of five 7-point scales.

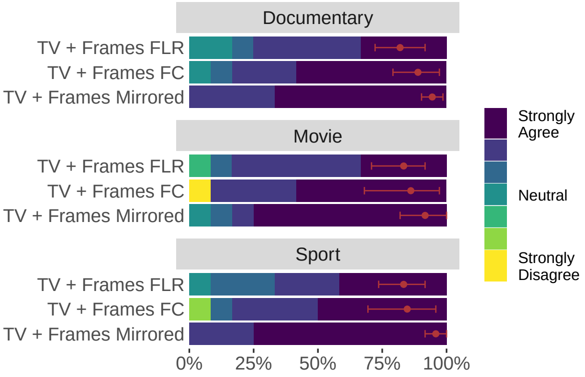

Figure 3: Responses to "The TV audio was well synchronized with the Frames audio"

Given our implementation, we are uncertain what caused the occasional ratings of poor synchronization - potentially glitches in the playback software, but given the responses predominantly

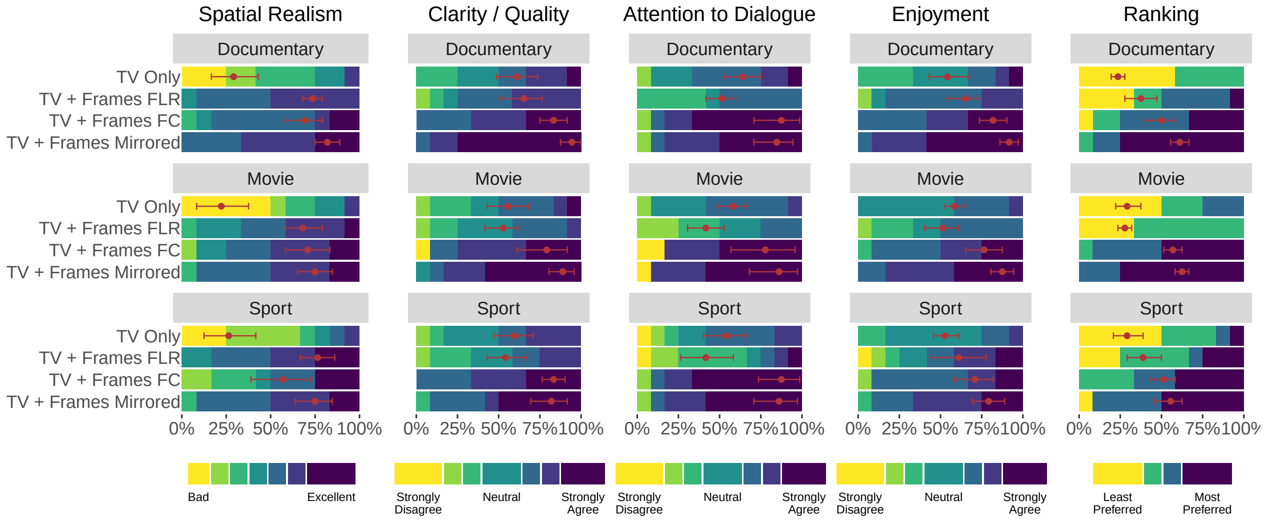

Figure 2: Responses to (in order): *Spatial realism* "Please rate the realism of the spatial reproduction / your sensation of sounds coming from different directions"; *Clarity/Quality* "The audio was clear/of good quality"; *Attention to Dialogue* "It was easy to pay attention to the speech/dialogue"; *Enjoyment* "I enjoyed the media experience"; *Rank Preferences.*

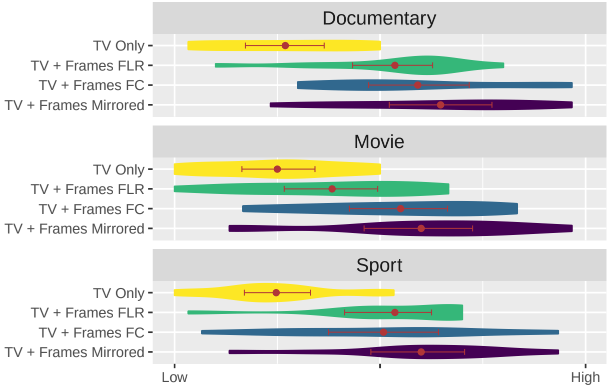

Figure 4: Score for the Film IEQ "Transportation" subscale using a Violin plots [22], displaying a rotated kernel density plot on either side of a box plot. Kernel density plots are "a variation of a Histogram that uses kernel smoothing to plot values, allowing for smoother distributions by smoothing out the noise. The peaks of a Density Plot help display where values are concentrated over the interval" [59], allowing for density estimation.

change with the FLR audio, this suggests that perhaps these environmental effects and music tracks are more difficult to relate to the content if they obscure the dialogue. Finally, the Film IEQ **Transportation** subscale results (see Figure 4) suggest that the more of the audio mix is added to the Frames, the more real the experience was perceived.

4.2 Perception of Speech Augmentations

Participants broadly found that the additional speech augmentations appeared well in-sync with the TV audio, and were able to attend to the additional speech over the TV content (see Figure 5).

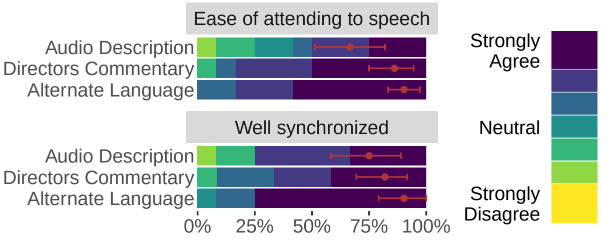

Figure 5: Responses to "It was easy to pay attention to the speech/dialogue" and "The TV audio was well synchronized with the Frames audio" for the speech augmentations.

4.3 Acceptability and Adoption

Examining the audio features we have tested regarding the acceptability of use by others in a shared viewing setting, the likelihood of personal usage, and the compatibility with conversation (see Figure 6), we see that firstly regarding acceptability there were no major impediments to adoption if the audio was augmenting the TV experience - with the debatable exception of the Director's Commentary, which in interviews was suggested as semantically changing the content, and thus undermining the shared viewing experience. Listening to non-TV content in a shared viewing environment was however broadly unacceptable.

Regarding the likelihood of personal usage, Directors Commentary and Audio Description saw less positive responses, for the former reflecting the lack of need (none of our participants were visually impaired) and for the latter reflecting the perceived acceptability. Non-TV content was rated as particularly unacceptable, suggesting an interesting tension if such headsets do see everyday adoption and use in the near future. Finally, regarding the ability to converse with others, the spatial environment and music sound was rated highly, with participants split for Language Change and Audio Description. Use for Louder Dialogue, Directors Commentary, and Mirrored Audio had users being more negative however - given their prior exposure to intermixed listening of the TV and Frames audio, as more of the audio mix is added to the headset participants perceive that their ability to hear external sound is diminished.

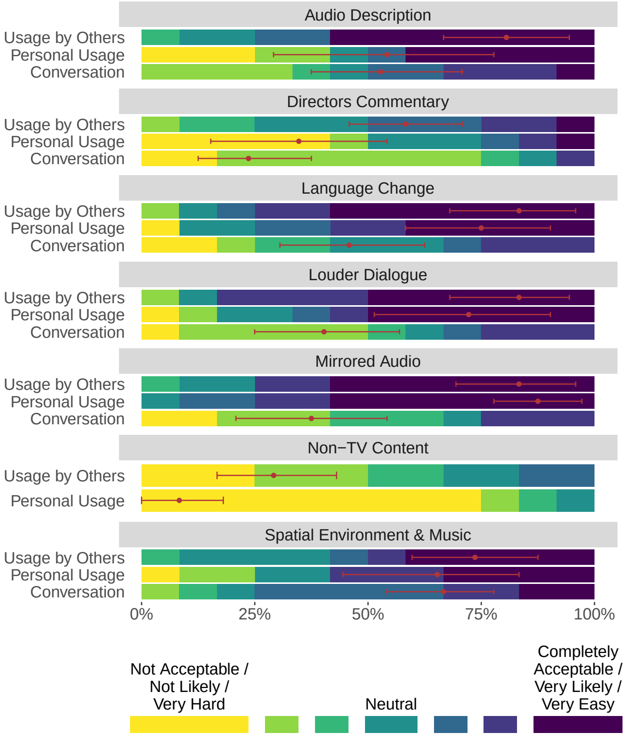

Figure 6: Responses to "how acceptable you would find it for those wearing Frames to privately augment their experience", "how likely would you be to use them to listen to...?" and "how easy would you think it would be to hold a conversation with others and hear their responses?" across the listening scenarios tested in the paper, and also for non-TV content (e.g. podcasts) for the first two questions.

4.4 Interviews

Semi-structured interviews were conducted at the end, regarding listening preferences across media types, the social acceptability of augmenting the TV audio, and potential use of the frames in other media-oriented scenarios (e.g. when mobile). Interviews were coded with Initial Coding using QCAmap [26], then grouped and sorted based on frequency and interest (see [53]), with representative excerpts quoted.

4.4.1 Acceptability of Private Augmentations. Participants were broadly in favour of private auditory augmentations of the TV content. For alternate languages, five participants noted how it might better facilitate enjoyable shared movie experiences:

> *P4:* I really like to watch movies in English and I my family doesn't understand English. So it would be, like, awesome, to watch it in English and they watch it with me but in a different language.
> *P6:* A lot people find subtitles quite tiring reading like, for example, a film that's 2 or 3 hours... [this] makes such a huge difference.
> *P8:* If I have friends over that speak don't speak English as well as I do, it's quite nice to know that they can listen in their most comfortable language.

For directors commentary, interview feedback was more muted, with only 2 participants expressing a strong preference for this as a feature "where if you were watching a movie that you had seen before and wanted your friends to watch it with you, then you could enjoy it to a different level" (P3), with two participants noting that such content might lead to a breakdown of the shared viewing experience, feeling "less communal... I'd feel it was a bit rude" (P10). The response to audio description largely centered around the mutual benefit to those that required AD being able to privately listen without unduly altering the shared viewing experience, noted by seven participants e.g. "I wouldn't mind, but I would prefer when I would not hear it" (P4), "if it's purely because of their impairment, I wouldn't need to listen to it, if it will deliver the same experience, I wouldn't want to [hear it]" (P5). However, the delivery of AD on the Frames without the mirrored audio of the TV content was noted to be jarring to two participants, as the AD voice would unpredictably be heard through the Frames, with one participant "feel[ing] uneasy, like when this was going to show up again" (P9).

4.4.2 Intermixing TV and Frames Audio. Participants were near-universally in favour of using the Frames for some form of synchronous audio delivery for the TV content:

> *P4:* I was sitting here and it felt like I was in the cinema, and the sounds around me coming from different directions.
> *P6:* I would love to get a set of those glasses. I thought the whole experience was incredible.
> *P8:* I could definitely see myself watching it in the living room, watching a television like with this.
> *P10:* It kind of had a headphones feeling... so you had the audio quality of headphones where you heard stuff better and it was next to your ears, but it was kind of fuller... it wasn't as restrictive as wearing headphones.

However, it must be noted that the TV audio quality was perceived by three participants as being poor in comparison. The reasons for these preferences lay in the combination of enhanced perception of speech, greater feelings of immersion in the content, and the contrast in the perceived audio quality of the TV versus the Frames. They enhanced immersion and perception of spatial audio for 9 participants:

> *P1:* For the sport, it was interesting because when you have the sound of the commentators through the Frames I felt like they were sitting with me on the couch... when I have the engine sound in my brain, I felt more like, you know, I was in the audience around the track.
> *P3:* I felt a lot more immersed and I really enjoyed noticing sound effects and especially music in the movie, I wouldn't have heard some of those like smaller details at all if it had just come from the TV.
> *P4:* It was the best, it was like being in the cinema, I could hear the movement from one ear to the other
> *P6:* You heard amazing sounds like the footsteps of the kind of ostrich emu thing [in the documentary]
> *P12:* The sound became more real when wearing [the] glasses... because they can simulate the sound from different directions.

Moreover, the intermixing of the TV audio and the FLR channels on the Frames lead to enhanced sensations of presence specifically for the Sport content for 3 participants:

> *P3:* I actually preferred it when the dialogue wasn't through the [Frames]... It felt a lot more immersive because it almost felt like I was actually in the stadium because the dialogue would have been distant, if you heard it through speakers or something, and I felt like the calls were coming more towards me... which I really enjoyed.
> *P6:* When you could hear the track sounds it was good, you felt like you were there.
> *P10:* It actually made the commentary initially hard to hear, but it almost made me feel like I was out in the grandstands listening to the sounds of the cars first and then some some piped in commentary from like a tannoy or something. And in that scenario it made the the cars and track feel way more real... Even though it was a bit less clear, it was way more immersive.

However, the intermixing also emphasized that acoustic transparency also depends on the respective volumes of the TV and Frames content, and that as a consequence dialogue could be harder to hear with the TV + Frames FLR combination for 6 participants:

> *P3:* [I] found it quite hard to follow the dialogue when it was on the TV and the sound effects were happening
> *P6:* But the worst was it was not TV, it was the front/left right channels, because environmental effects meant you didn't really hear David Attenborough's commentary because you heard amazing sounds like the footsteps of the kind of ostrich emu thing. But if you don't have Attenboroughs commentary, it's kind of worthless.
> *P10:* Even though I'm theoretically sitting next to characters [in the movie], I actually can't hear them over this magical music that's playing in my ears, it's like you're a crazy person in the universe or something.

One participant offered that this was in-part because the LR channels combined both diegetic environmental effects *and* music, which led to an overpowering mix of audio, and that our replication of select 5.1 channels did not take into account the suitability of presenting non-diegetic (e.g. commentary, music) and diegetic (e.g. dialogue, environmental effects) audio:

> *P10:* So the thing that I should be hearing on the glasses in my mind is the thing that would be around me. So in F1, for example, I would be hearing the sound of the car engines. So it didn't bother me when those were kind of going over the commentators, that felt more immersive. But in the documentary and the film, when the effects are all around you... there's loads of music and the

> music doesn't feel realistic.... And you can't hear what the people who are theoretically right next to you in the film are saying.

Finally, mirroring the audio was seen as preferable, evidenced by the previous rankings, as it both provided spatialised audio and enhanced dialogue:

> *P5:* It was louder and louder and maybe helped me to focus on the movie. I could hear it more clearly, so I understood it better.
> *P8:* I felt like the dialogue, having the enhanced dialogue could be quite good. If, for example, there's not good volume on the television.

4.5 Limitations

There are some caveats to this study that temper interpretation of our findings to be considered.

4.5.1 Headphones. Some of the use cases here could have previously been enabled with standard occlusive headphones, and indeed some of the findings (e.g. regarding spatialization) may likely be enhanced by the higher fidelity audio of such headphones. However, we re-affirm that this is a new, and different, proposition to prior headphone usage. If viewing TV content with others, headphones are not a viable means of listening, being occlusive to the sounds of reality, inhibiting our ability to interact with others and be aware of our environment. Acoustically transparent audio offers enhanced, personal audio whilst still maintaining this connection to reality.

4.5.2 Fidelity. The point of comparison (a standard low-end consumer TV) is not representative of the gamut of TV audio experiences. Had we tested against a TV with a Dolby Atmos sound system for example, our findings regarding enjoyment, realism etc. may well have been significantly different. However, such a standard TV audio setup is representative of what is available in the homes of many consumers, with consumers predominantly relying on the in-built TV audio or soundbars which will struggle to recreate spatial experiences [70]. Further study would however be beneficial here. Moreover, even the most sophisticated consumer sound system cannot deliver per-person audio - devices such as the Frames enable us to personalize the audio based on impairment, need or preference.

4.5.3 Intermixing. If such auditory headsets see mass adoption, we may find that the TV audio may not even be necessary, if everyone in the room is wearing such headsets. However, we reasoned that this possibility is far off, choosing to examine the practical impact in mixed viewing settings.

4.5.4 Effect on media type. Our analysis by media type is cautious at best, primarily because whilst we do have documentary, film and sport content, the genres of content themselves are not entirely encapsulated by what we selected (e.g. an action movie versus a horror movie).

4.5.5 Ecological validity. Participants were in a dimmed lab environment rather than a home/living-room context, watching an unfamiliar TV whilst wearing a wired auditory headset for a short duration, and this will have undoubtedly affected both their comfort, and the acoustics/immersion of the experience.

4.5.6 Collocated use. Whilst this paper is geared toward the shared viewing experience, we did not evaluate the perception of these audio experiences with groups of viewers. This was a deliberate choice - whilst we do wish to follow-up with intimacy groups of viewers, introducing the social element diminishes our control over the perception of the audio, and we prioritised better understanding the impact of intermixing personal and public audio.

5 DISCUSSION

From this study, it is clear that the use of acoustically transparent audio headsets meaningfully improves both the perception of the TV audio, and our ability to augment the TV experience through this private channel.

RQ1: The TV content was more immersive and more enjoyable when using the Bose Frames to deliver synchronized audio, with users for the most part preferring that the TV audio be completely mirrored on the Frames. There were some exceptions where the intermixing of TV and Frame audio led to more immersive experiences, specifically with respect to the delivery of sports commentary, where the Frames being used for environmental sounds led to a feeling of "being in the grandstand". However for the most part the Frames were preferred as a means to amplify the dialogue and enhance the perception of spatial sounds.

RQ2: For augmentations, those that did not meaningfully change the semantics of the content were broadly favoured and accepted (e.g. audio description, alternate languages) and easily discerned over the TV audio, indeed being broadly praised for their ability to bring viewers together in contexts where barriers (visual impairments, languages) might have previously made viewing together more difficult, or led to a diminished experience for some.

RQ3: Our findings suggest that if wearable, personal, acoustically transparent audio is to see meaningful adoption by consumers, this private channel can be readily exploited to the benefit of shared viewing experiences, with the majority of features tested likely to see usage according to our participants. Where augmentations meaningfully altered the semantics of the experience (directors commentary), there was a fear amongst some participants that this would lead to a more disconnected viewing experience - however, some participants were openly enthused with the prospect of making repeat viewings of films more palatable in contexts where others may not have previously seen the given content.

5.1 The Benefits of Auditory Headsets for TV

Acoustically transparent headsets offer viewers the possibility to remain connected to their aural environment, whilst personally and selectively altering the perception of the TV audio content. The reasons behind these augmentations can vary, but we see key benefits in terms of:

- Added **spatial realism and immersion**, providing a better quality of audio than may be available from the TV
- Giving users **personal control over volume** and other aspects of the audio mix, such as emphasising the speech content, particularly beneficial for some of the non-native English speakers in our study

- Supporting **accessibility and language comprehension** for select individuals, without unduly altering the TV experience (e.g. through audio description or subtitling in other languages) for all viewers
- Providing ways of making **repeat viewings more engaging**, without impacting others' initial experience of the content (albeit at a potential cost in terms of having the same shared experience across the group).

5.2 Open Questions and Future Work

There are, however, significant questions and challenges to be addressed for such a system to become a consumer reality:

5.2.1 How to enable production and delivery of personal TV audio? Logistically, synchronizing personal audio with the TV audio output is problematic, but solvable, as evidenced by prior work by Vinayagamoorthy *et al.* [66] in particular. However different devices may have different end-to-end latencies to be accounted for. More generally, our efforts in this paper were limited to using available broadcast content, utilizing center channel mixes that were not clean with respect to dialogue. However, there is a push in production toward the adoption of "Clean" [33, 55] object-based audio [56] (e.g. using MPEG-H [57] or Dolby Atmos [32]), and research is being carried out examining how best to give users control over their personal mix e.g. for accessibility [71]. There are a multitude of questions here e.g. should diegetic/non-diegetic speech be spatialized [72], and should this prioritise immersion versus comprehension? Auditory headsets offer an ideal mechanism for spatialized personalization – but they also introduce new issues in determining the correct headset mix *set against* the existing TV audio output (if present) in a shared viewing setting, as well as the potential for headset-based ducking to allow viewers to attend to events in the real-world environment e.g. the speech of other viewers.

5.2.2 When is acoustic transparency actually acoustically transparent? Despite the physical properties of the Bose Frames, perception of reality appears at least in-part contingent on the volume of reality against the volume of content being emitted by the headset (i.e. the headset audio energetically masking [58] the TV and environment audio), evidenced by the at-times occlusive nature of some of the FLR presentations in this paper. To exploit intermixing, and better facilitate use in social settings, research is needed to model the thresholds upon which an acoustically transparent headset in effect begins to significantly occlude the sounds of reality, and what interventions we might provide to overcome this e.g. using in-built directional microphones to incorporate a degree of active acoustic transparency, or automatically varying the headset volume to enhance perception of reality.

5.2.3 How might we exploit the personal, private audio channel? As McGill *et al.* previously noted [31], these auditory headsets are in-effect Mixed Reality displays, and often come complete with head orientation tracking. Consequently, whilst our use cases were largely functional, arguably the underlying technology is capable of facilitating far more creative uses of this private audio channel to augment our viewing experiences. For example, a murder mystery programme might deliver separate hints and clues to individuals in the room. Or a horror movie might deliver different exocentric spatial effects to different people, enhancing their fear or introducing an element of unpredictability in repeat viewings. In-situ studies might reveal other benefits - we might imagine that a trip to the kitchen during a live sports even would be less problematic if the user can still hear the audio of the commentary clearly as they walk around their home. Other experiences might benefit too, from a private channel for personal feedback or communications in split-screen gaming, to spatializing the voices of distant others in at-a-distance co-viewing experiences [29].

5.2.4 What happens when everyone has wearable audio? If the TV audio output is no longer strictly necessary, then the TV itself becomes an additional audio output to be readily exploited. Perhaps the TV might only output the environmental sounds, or at a lower volume, allowing collocated individuals to still share in the experience of co-viewing without actively attending to it. If viewers have a shared spatial audio space, we might imagine more engaging experiences being designed to take advantage of this, appropriating the TV speakers much as with the BBC Vostok-K experience [12–14, 21]. Even prior work regarding active multi-view displays [27, 28] could be re-assessed, given we can now deliver both personal private visuals *and* audio.

6 CONCLUSIONS

This paper has explored how acoustically transparent audio headsets could be used to selectively enhance or augment an individuals auditory experience of TV whilst retaining an auditory connection to reality. We examined the impact of synchronously mirroring select audio channels from the 5.1 mix (dialogue, environmental sounds, and the full mix) onto the audio headset, finding that auditory headsets enabled better spatialization and more immersive, enjoyable viewing, with the intermixing of TV and headset audio creating unique listening experiences. We also explored augmenting TV viewing with additional speech content, namely Audio Description, Directors Commentary, and Alternate Language, finding such augmentations as both broadly acceptable and offering new ways to (re)watch content with others, whilst potentially diminishing barriers to co-viewing such as visual impairments and language comprehension. Compared to visually-oriented AR, such auditory headsets offer an affordable and feasible route toward augmenting reality, enabling practical and immersive private augmentations of TV content, and are thus well suited to being used in shared viewing contexts such as the living room.

ACKNOWLEDGMENTS

This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (#835197, *ViAjeRo*).

REFERENCES

[1] Valentin Bauer, Anna Nagele, Chris Baume, Tim Cowlishaw, Henry Cooke, Chris Pike, and Patrick G. T. Healey. 2019. Designing an Interactive and Collaborative Experience in Audio Augmented Reality. 305–311. https://doi.org/10.1007/978-3-030-31908-3_20

[2] BBC. 2019. Access services at the BBC. https://www.bbc.com/aboutthebbc/whatwedo/publicservices/accessservices{#}audiodescription

[3] D. R. Begault, E. M. Wenzel, and M. R. Anderson. 2001. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer

functions on the spatial perception of a virtual speech source. *AES: Journal of the Audio Engineering Society* 49, 10 (2001), 904–916. http://www.aes.org/e-lib/browse.cfm?elib=10175

[4] Bose. 2019. SoundWear Companion Wearable Speaker. https://www.bose.co.uk/en_gb/products/speakers/portable_speakers/soundwear-companion.html

[5] Bose. 2019. Wearables by Bose - AR Audio Sunglasses. https://www.bose.co.uk/en{_}gb/products/frames.html

[6] Rosie Campbell, Richard Felton, and Charlotte Hoarse. 2014. Smart Wallpaper. file:///C:/Users/Mark/Downloads/IF{_}110.pdf

[7] Francesco Cricri, Sujeet Mate, Igor D. D. Curcio, and Moncef Gabbouj. 2009. Mobile and Interactive Social Television â€” A virtual TV room. In *2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops*. IEEE, 1–8. https://doi.org/10.1109/WOWMOM.2009.5282411

[8] Poppy Crum. 2019. Hearables: Here come the: Technology tucked inside your ears will augment your daily life. *IEEE Spectrum* 56, 5 (may 2019), 38–43. https://doi.org/10.1109/mspec.2019.8701198

[9] P Dragicevic. 2015. HCI Statistics without p-values. (2015). https://hal.inria.fr/hal-01162238/

[10] Javier Pastor Fernando Boronat, Senior, IEEE, Dani Marfil, Mario Montagud. 2017. HbbTV-compliant Platform for Hybrid Media Delivery and Synchronization on Single- and Multi-Device Scenarios. *IEEE Transactions on Broadcasting* (2017). https://riunet.upv.es/bitstream/handle/10251/102497/IE3_Trans_on_broadcasting_fboronat2017_IDBTS-17-112.pdf?sequence=3

[11] Matthieu Fradet, Caroline Baillard, Anthony Laurent, Tao Luo, Philippe Robert, Vincent Alleaume, Pierrick Jouet, and Fabien Servant. 2017. MR TV Mozaik: A New Mixed Reality Interactive TV Experience. In *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017*. https://doi.org/10.1109/ISMAR-Adjunct.2017.53

[12] Jon Francombe. 2018. Vostok-K Incident: Immersive Audio Drama on Personal Devices. https://www.bbc.co.uk/rd/blog/2018-10-multi-speaker-immersive-audio-metadata

[13] Jon Francombe and Kristian Hentschel. 2019. Evaluation of an immersive audio experience using questionnaire and interaction data. In *International Congress on Acoustics 2019*. https://www.bbc.co.uk/rd/publications/whitepaper352

[14] Jon Francombe, James Woodcock, Richard J. Hughes, Kristian Hentschel, Eloise Whitmore, and Tony Churnside. 2018. Producing Audio Drama Content for an Array of Orchestrated Personal Devices. In *Audio Engineering Society Convention 145*. http://www.aes.org/e-lib/browse.cfm?elib=19726

[15] Louise Fryer. 2016. *An Introduction to Audio Description*. Routledge, Milton Park, Abingdon, Oxon; New York, NY: Routledge, [2016] | Series: Translation Practices Explained. https://doi.org/10.4324/9781315707228

[16] David Geerts, Evert Van Beek, and Fernanda Chocron Miranda. 2019. Viewers' visions of the future: Co-creating hyper-personalized and immersive TV and video experiences. In *TVX 2019 - Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM Press, New York, New York, USA, 59–69. https://doi.org/10.1145/3317697.3323356

[17] Samuel Gibbs. 2020. Amazon launches Alexa smart ring, smart glasses and earbuds. https://www.theguardian.com/technology/2019/sep/26/amazon-launches-alexa-smart-ring-smart-glasses-and-earbuds

[18] David Gómez, Juan A. Núñez, Mario Montagud, and Sergi Fernández. 2018. ImmersiaTV: Enabling customizable and immersive multi-screen TV experiences. In *Proceedings of the 9th ACM Multimedia Systems Conference, MMSys 2018*. ACM Press, New York, New York, USA, 506–508. https://doi.org/10.1145/3204949.3209620

[19] Stephen Groening. 2016. ‘No One Likes to Be a Captive Audience’: Headphones and In-Flight Cinema. *Film History: An International Journal* (2016). https://muse.jhu.edu/article/640056/summary

[20] Gunnar Harboe, Crysta J. Metcalf, Frank Bentley, Joe Tullio, Noel Massey, and Guy Romano. 2008. Ambient social tv. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ’08*. ACM Press, New York, New York, USA, 1. https://doi.org/10.1145/1357054.1357056

[21] Kristian Hentschel and Jon Francombe. 2019. Framework for Web Delivery of Immersive Audio Experiences Using Device Orchestration. In *Adjunct Proceedings of ACM TVX 2019*. ACM.

[22] JL Hintze and RD Nelson. 1998. Violin plots: a box plot-density trace synergism. *The American Statistician* (1998). http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1998.10480559

[23] Andrew F. Jarosz and Jennifer Wiley. 2014. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving* (2014). https://doi.org/10.7771/1932-6246.1167

[24] Robert E. Kass and Adrian E. Raftery. 1995. Bayes Factors. *J. Amer. Statist. Assoc.* 90, 430 (jun 1995), 773–795. https://doi.org/10.1080/01621459.1995.10476572

[25] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. 2018. eSense: Open Earable Platform for Human Sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems - SenSys ’18*. ACM Press, New York, New York, USA, 371–372. https://doi.org/10.1145/3274783.3275188

[26] Philipp Mayring. 2014. Qualitative content analysis: Theoretical foundation, basic procedures and software solution. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (2014).

[27] Mark McGill, John Williamson, and Stephen A. Brewster. 2015. It Takes Two (To Co-View). In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX ’15*. ACM Press, 23–32. https://doi.org/10.1145/2745197.2745199

[28] Mark McGill, John Williamson, and Stephen A. Brewster. 2015. Who’s the Fairest of Them All. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX ’15*. ACM Press, New York, New York, USA, 83–92. https://doi.org/10.1145/2745197.2745200

[29] 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

[30] Mark McGill, John H. Williamson, and Stephen A. Brewster. 2015. A review of collocated multi-user TV. *Personal and Ubiquitous Computing* 19, 5-6 (jun 2015), 743–759. https://doi.org/10.1007/s00779-015-0860-1

[31] Mark McGill, Graham Wilson, David McGookin, and Stephen Brewster. 2020. Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. https://doi.org/10.1145/3313831.3376702

[32] Takashi Mikami, Masataka Nakahara, and Kazutaka Someya. 2016. Compatibility Study of Dolby Atmos Objects’ Spatial Sound Localization Using a Visualization Method. http://www.aes.org/e-lib/online/browse.cfm?elib=18157

[33] Mike Armstrong. 2016. From Clean Audio to Object Based Broadcasting. *BBC Research & Development White Paper* WHP 324 (2016). http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP324.pdf

[34] Miscellaneous. 2020. Bose AR SDK Overview. https://developer.bose.com/guides/bose-ar/sdk-overview

[35] Miscellaneous. 2020. FFmpeg - A complete, cross-platform solution to record, convert and stream audio and video. https://www.ffmpeg.org/

[36] Miscellaneous. 2020. NAudio Audio and MIDI library for.NET. https://github.com/naudio/NAudio

[37] Mario Montagud, Isaac Fraile, Juan A. Nuñez, and Sergi Fernández. 2018. ImAc: Enabling immersive, accessible and personalized media experiences. In *TVX 2018 - Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM Press, New York, New York, USA, 245–250. https://doi.org/10.1145/3210825.3213570

[38] Richard D. Morey. 2019. Computation of Bayes Factors for Common Designs [R package BayesFactor version 0.9.12-4.2]. *cran* (2019). https://cran.r-project.org/web/packages/BayesFactor/index.html

[39] Danielle Navarro. 2018. Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1). In *University of New South Wales*. Chapter Chapter 17. https://learningstatisticswithr.com/book/bayes.html

[40] Timothy Neate, Michael Evans, and Matt Jones. 2017. Enhancing Interaction with Dual-Screen Television Through Display Commonalities. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX ’17*. ACM Press, New York, New York, USA, 91–103. https://doi.org/10.1145/3077548.3077549

[41] Timothy Neate, Matt Jones, and Michael Evans. 2017. Cross-device media: a review of second screening and multi-device television. *Personal and Ubiquitous Computing* 21, 2 (apr 2017), 391–405. https://doi.org/10.1007/s00779-017-1016-2

[42] Nreal. 2020. Building Mixed Reality for Everyone. https://www.nreal.ai/

[43] D. Oetting and H. Fuchs. 2013. Advanced clean audio solution: dialogue enhancement. In *International Broadcasting Convention (IBC) 2013 Conference*. Institution of Engineering and Technology, 1.2–1.2. https://doi.org/10.1049/ibc.2013.0002

[44] Rita Oliveira, Jorge Ferraz de Abreu, and Ana Margarida Almeida. 2018. Audio Description of Television Programs: A Voluntary Production Approach. Springer, Cham, 150–160. https://doi.org/10.1007/978-3-319-90170-1_11

[45] Joseph Plazak and Marta Kersten-Oertel. 2018. A survey on the affordances of ‘hearables’. 3, 3 (jul 2018). https://doi.org/10.3390/inventions3030048

[46] Jussi Rämö and Vesa Välimäki. 2012. Digital Augmented Reality Audio Headset. *Journal of Electrical and Computer Engineering* 2012 (oct 2012), 1–13. https://doi.org/10.1155/2012/457374

[47] Red Letter Media. 2019. Director’s Commentaries. https://redlettermedia.bandcamp.com/audio

[48] Jacob M. Rigby, Sandy J.J. Gould, Duncan P. Brumby, and Anna L. Cox. 2019. Development of a questionnaire to measure immersion in video media: The Film IEQ. In *TVX 2019 - Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. https://doi.org/10.1145/3317697.3323361

[49] RNIB. 2019. Audio description (AD). https://www.rnib.org.uk/information-everyday-living-home-and-leisure-television-radio-and-film/audio-description

[50] Agnieszka Roginska and Paul Geluso. 2017. *Immersive sound: The art and science of binaural and multi-channel audio*. 1–364 pages. https://doi.org/10.4324/9781315707525

[51] Pejman Saeghe, Sarah Clinch, Bruce Weir, Maxine Glancy, Vinoba Vinayagamoorthy, Ollie Pattinson, Stephen Robert Pettifer, and Robert Stevens. 2019. Augmenting Television With Augmented Reality. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. https://dl.acm.org/citation.cfm?id=3325129

[52] Neelima Sailaja, Adrian Gradinar, James Colley, Paul Coulton, Andy Crabtree, Ian Forrester, Lianne Kerlin, and Phil Stenton. 2019. The living room of the future. In *TVX 2019 - Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. 95–107. https://doi.org/10.1145/3317697.3323360

[53] J Saldaña. 2015. *The coding manual for qualitative researchers*. Sage.

[54] Nitin Sawhney and Chris Schmandt. 2000. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction* 7, 3 (2000), 353–383. https://doi.org/10.1145/355324.355327

[55] Ben Shirley and Paul Kendrick. 2006. The Clean Audio project: Digital TV as assistive technology. *Technology and Disability* 18, 1 (2006), 31–41. https://doi.org/10.3233/tad-2006-18105

[56] Ben Guy Shirley, Melissa Meadows, Fadi Malak, James Stephen Woodcock, and Ash Tidball. 2017. Personalized object-based audio for hearing impaired TV viewers. *Journal of the Audio Engineering Society* (2017). http://www.aes.org/e-lib/browse.cfm?elib=18562

[57] Christian Simon, Matteo Torcoli, and Jouni Paulus. 2019. MPEG-H Audio for Improving Accessibility in Broadcasting and Streaming. (sep 2019). arXiv:1909.11549 http://arxiv.org/abs/1909.11549

[58] Y Tang and TJ Cox. 2018. Improving intelligibility prediction under informational masking using an auditory saliency model, In International Conference on Digital Audio Effects. *Proceedings of the International Conference on Digital Audio Effects 2018*, 113–119. http://usir.salford.ac.uk/id/eprint/47113/

[59] The Data Visualisation Catalogue. 2019. Density Plot. https://datavizcatalogue.com/methods/density_plot.html

[60] Matteo Torcoli, Alex Freke-Morin, Jouni Paulus, Christian Simon, and Ben Shirley. 2019. Background ducking to produce esthetically pleasing audio for TV with clear speech. In *AES 146th International Convention*. http://www.aes.org/e-lib/online/browse.cfm?elib=20308

[61] Unity Technologies. 2019. Unity Game Engine. https://unity3d.com/

[62] M. Oskar van Deventer, Michael Probst, and Christoph Ziegler. 2018. Media Synchronisation for Television Services Through HbbTV. In *MediaSync*. Springer International Publishing, Cham, 505–544. https://doi.org/10.1007/978-3-319-65840-7_18

[63] Radu-Daniel Vatavu. 2013. There's a world outside your TV. In *Proceedings of the 11th european conference on Interactive TV and video - EuroITV '13*. ACM Press, New York, New York, USA, 143. https://doi.org/10.1145/2465958.2465972

[64] Vinoba Vinayagamoorthy. 2018. How Users Perceive Delays in Synchronous Companion Screen Experiences - An Exploratory Study. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18*. ACM Press, New York, New York, USA, 57–68. https://doi.org/10.1145/3210825.3210836

[65] Vinoba Vinayagamoorthy, Maxine Glancy, Christoph Ziegler, and Richard Schäffer. 2019. Personalising the TV Experience using Augmented Reality An Exploratory Study on Delivering Synchronised Sign Language Interpretation. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300762

[66] Vinoba Vinayagamoorthy, Rajiv Ramdhany, and Matt Hammond. 2016. Enabling Frame-Accurate Synchronised Companion Screen Experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '16*. ACM Press, New York, New York, USA, 83–92. https://doi.org/10.1145/2932206.2932214

[67] James Vincent. 2016. Are bone conduction headphones good enough yet? https://www.theverge.com/2016/10/24/13383616/bone-conduction-headphones-best-pair-aftershokz

[68] Vue. 2020. Meet Vue - Smart Glasses. https://www.enjoyvue.com/

[69] Jack Wallen. 2015. Earables: The next big thing. https://www.techrepublic.com/article/earables-the-next-big-thing/

[70] Tim Walton, Michael Evans, David Kirk, and Frank Melchior. 2016. A subjective comparison of discrete surround sound and soundbar technology by using mixed methods. In *140th Audio Engineering Society International Convention 2016, AES 2016*. Audio Engineering Society. http://www.aes.org/e-lib/browse.cfm?elib=18290

[71] Lauren Ward, Matthew Paradis, Ben Shirley, Laura Russon, Robin Moore, and Rhys Davies. 2019. Casualty accessible and enhanced (A&E) audio: Trialling object-based accessible TV audio. In *147th Audio Engineering Society International Convention 2019*. http://www.aes.org/e-lib/browse.cfm?conv2=147{&}ebrief=563

[72] Lauren A. Ward and Ben G. Shirley. 2019. Personalization in object-based audio for accessibility: A review of advancements for hearing impaired listeners. *AES: Journal of the Audio Engineering Society* 67, 7-8 (aug 2019), 584–597. https://doi.org/10.17743/jaes.2019.0021

[73] Christoph Ziegler, Christian Keimel, Rajiv Ramdhany, and Vinoba Vinayagamoorthy. 2017. On time or not on time: A user study on delays in a synchronised companion-screen experience. In *TVX 2017 - Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM Press, New York, New York, USA, 105–114. https://doi.org/10.1145/3077548.3077557

[74] Christoph Ziegler, Richard Schäffer, Vinoba Vinayagamoorthy, Maxine Glancy, Paul Debenham, and Alastair Bruce. 2018. Personalising the TV experience with augmented reality technology: Synchronised sign language interpretation. In *TVX 2018 - Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*. https://doi.org/10.1145/3210825.3213562