Evaluating hearables for augmenting TV audio in shared viewing environments

David Geary, Kristian Hentschel & Damian Murphy

***Abstract*—The TV remains the preferred device for shared audiovisual media consumption in the home; however, its limited ability to cater to individual preference in these scenarios can be a hindrance. The increasing ubiquity of devices in the home and the ability to connect them through wireless communication protocols presents a good opportunity to overcome this challenge. Hearables with 'hear-through' transparency features offer the potential to provide personal audio in shared TV viewing environments. A listening test (n=30) was conducted to observe how various TV mix content could be rendered over hearables, with concurrent TV audio perceived over a transparency mode, and how those options affect preference and dialogue clarity of content. The results show that mirroring the full mix or the dialogue on the hearables is preferred equally to the open-ear TV condition, however is less preferred than the hearables only condition. An improvement in dialogue clarity was observed in cases where dialogue in the original mix was challenging to perceive. Variance in both preference and dialogue clarity attributes highlights the requirement for personal mix control in shared viewing environments.**

***Index Terms*—multi-device, audio, hearables, TV, internet of sounds, audio augmented reality, listening test.**

I. INTRODUCTION

TV has served as a hub for social activity in the home for many decades. However, the allure of TV, has slowly been eroded through technological developments since the turn of the century, particularly in personal devices and personalisable media. As a result, 'multi-screening' has become more common in the living room, where the TV and personal devices are used simultaneously for separate tasks while attention switches between them. Contrary to this, ongoing research and creative practice are increasingly finding new ways of using multiple devices in a complementary manner to form new 'multi-device' applications and experiences. These have been made possible by the development of wireless communication protocols such as Bluetooth and WiFi, and mobile computing technologies. These affordances have culminated in the formation of research areas such as the *Internet of Things* (IoT), and more recently, the *Internet of Sounds* (IoS) for audio and music related applications [1]. New consumer devices on the market additionally present new possible applications and experiences. One example for audio applications, is the emergence of hearables, wireless hearing devices that not only deliver audio content but can also have additional features such as microphones for audio capture and processing, accelerometers, GPS, or even biological sensors [2]. For these reasons, hearables have been the subject of research for Extended Reality (XR) audio applications. As an example, hearables in the form of audio glasses have been used to augment TV audio by allocating (or 'mirroring') parts of the TV mix on the glasses [3]. This study builds on this to observe whether consumer wireless earbud-style hearables with 'hear-through' capabilities have the potential to provide useful personal audio augmentations for improving TV viewing in shared viewing environments. In particular, the paper focuses on formally evaluating a simulation of the listening experience under ideal conditions and comparing it to listening to the TV on its own.

D. Geary is with the AudioLab in the School of Physics, Engineering and Technology, University of York

K. Hentschel is with BBC Research & Development

D. Murphy is with the AudioLab in the School of Physics, Engineering and Technology, University of York

This work investigates two research questions.

1) How does mirroring of audio content on wireless earbud-style hearables affect preference and dialogue clarity versus standard TV viewing?
2) Does the type of programme affect these attributes for this form of reproduction?

The paper is organised into the following sections. An overview of relevant literature is presented in Section II. The methods used for preparing the audio material and the listening test are described in Section III. The results of the listening test are outlined in Section IV and discussed in Section V. Limitations and future work are considered in Section VI, and the conclusion is provided in Section VII.

II. RELATED WORK

A. Hearables and transparency modes

The definition of hearables varies depending on the application of focus. In audio research, hearables are devices that not only contain loudspeakers for audio reproduction but additional hardware and software for audio processing, control and augmentation [4]. Another field with significant interest in hearables is medical sciences, whose focus and subsequent definitions are concerned with application of these devices for bio-signal monitoring [5]. In both cases, hearables are wireless devices that are worn in, on, or around the ear that contain a range of hardware and software features for data collection and signal processing, typically driven by a mobile device, such as a smartphone, using Bluetooth. Given these criteria, hearables come in many different form factors including headphones, earbuds, audio glasses, and hearing aids. In addition to audio reproduction, most modern consumer hearables have sound control features such as Active Noise Cancellation for attenuating unwanted external noise, and 'transparency' for presenting the external sound environment back to the listener. Other features of these devices can include accelerometers, compass, and GPS, which enable novel interactions with audio material, including motion and location tracking, which are necessary for XR applications.

Acoustic transparency has become more prevalent with increased use of wearable audio technology outside the home. While outside, we gain lots of information about our surroundings through sound, whether that is to ensure safety or communication with others. Therefore, having some form of acoustic transparency enables listeners to attend to this information. Transparency can be achieved passively (e.g. audio glasses) or through 'hear-through' signal processing, which is used in ear canal occluding devices such as headphones and earbuds. Hear-through processing attempts to compensate for the sound modification imparted by the device, and to approximate the open ear response through equalisation [6]. A technical evaluation of commercial and prototype hearable transparency modes is presented by Denk et al. [7]. The authors identified three characteristics for achieving good transparency: synchronisation between left and right earpieces, suitable equalisation to open-ear response and reduction of delay and comb filtering effects. A large variance in achieving the open ear response was observed across the tested devices. A further perceptual evaluation of these devices revealed that most were not able to achieve sound quality comparable to the open ear, and that sound quality is largely determined by the extent to which the open ear transfer function is reproduced [8].

B. Improving TV audio

Generally there are two approaches to improving audio in TV experiences. First is to improve the audio itself through addressing aspects of production, distribution and delivery of content, all of which are the aims of object-based audio (OBA). Features of OBA include accessibility, interactivity options, and spatial audio; however, the full potential of OBA is yet to be realised as considerable changes to existing broadcast infrastructure and production workflows are still necessary [9]. Audio can also be improved through additional signal processing. For example, speech enhancement techniques using neural networks have been found to be effective for reducing listening effort [10]. A problem that remains is that the TV, as a facilitator of shared media experiences, cannot itself accommodate the range of individual preferences in a shared viewing environment. This problem is especially apparent for viewers with specific accessibility requirements who currently may feel guilt or reluctance for imposing their TV sound requirements onto others. In many cases, a compromise between listeners may be achievable. However, where there is no satisfactory resolution, the ability to have personalised audio without diminishing the shared experience would be desirable. This can be achieved through using additional connected devices to either enhance or complement TV audio. One approach is to use soundzones, multiple distinct acoustic spaces (or zones) within the same physical space with minimal interference, created with loudspeaker arrays. These zones are generated using signal processing techniques such as acoustic contrast control and pressure control, together with beam-forming technology [11]. Soundzone solutions can be packaged into soundbars for domestic usage; however, there are still technical challenges to be overcome before mass market adoption. Small portable loudspeakers placed close to individuals, amplifying the full mix, not just the components of interest, offer some of the same affordances, albeit still affecting the experience of other listeners within a shared setting.

Hearables provide an alternative method where, for instance, accessibility improvements for hearing-impaired listeners can be provided to only those who require them. Bluetooth streaming accessories, for example, that provide a method for sending TV audio directly can yield increased speech quality for hearing aid users [12]. Bone conduction headphones when used to mirror the TV audio have been demonstrated to provide benefits of increased clarity and spatialisation, especially for elderly listeners [13]. McGill et. al. [3] explored the use of acoustically transparent audio glasses to enhance TV viewing through mirroring various channel combinations of the TV mix on the glasses, and providing additional speech augmentations such as audio description and director's commentary. Benefits of mirroring were observed across all measured attributes including spatial realism, clarity, attention to dialogue and enjoyment, with the largest benefit reported where the full TV mix was mirrored on the audio glasses. Synchronised smartphones and tablets have been used to increase the loudspeaker count of TV programme experiences to provide a more immersive spatial audio experience [14]. Most recently, Geary et. al. [15] designed a prototype audiovisual experience where personal audio in the hearables was augmented with loudspeakers playing diffuse sounds and perceived by the listener through a transparency mode. Feedback from demonstrations revealed the potential for using transparency modes to provide personal speech enhancement for loudspeaker delivered audio. Transparency modes enable listeners to perceive their external sound environment while enjoying personal sound content, and are especially useful in outdoor environments. However, their potential to be used as a method for improving audio experiences in domestic settings have yet to be explored.

III. METHOD

To investigate the proposed research questions, a within-subjects multiple-stimulus listening test was designed and conducted. Additional data on attitudes towards hearables and social listening was collected through an exit survey at the end of the test. This section describes the experimental methods and the test procedure.

A. Listening conditions

Five different listening conditions were investigated; three involved combining TV and hearables output where the hearables mirror part or all of the full TV mix, which consisted of dialogue (D), and music and effects (ME). The conditions were 'full mix on both devices' (*FM_FM*), 'dialogue on the hearables and full mix on the TV' (*D_FM*) and 'music and effects on the hearables and full mix on the TV' (*ME_FM*). The other two conditions were single device output only. One being the full mix on the TV without the hearables inserted (*TV_ONLY*), and the other where the full mix was played only on the hearables (*H_ONLY*). An overview of the conditions can be seen in Table I. The full mix was preserved on the TV to focus on shared viewing scenarios where not everyone might have hearables. To observe any effects across different types of TV content, three BBC programmes were selected from *Documentary*, *Drama* and *Sports* genres. For each of the three programme types, three 20 second sections were captured to form the test items. Separate dialogue (D) and background stems (ME) were provided by the BBC for each programme. For the documentary and drama programme, the background consisted of music and varying sound effects. The sports programme background contained crowd noise and pitch-side sound effects. The full mixes for each programme were created by summing together the dialogue and background stems. Further explanation of how the test items were prepared is in the next section.

Condition	Hearables audio	TV audio
FM_FM	D, ME	D*, ME*
D_FM	D	D*, ME*
ME_FM	ME	D*, ME*
TV_ONLY	-	D, ME
H_ONLY	D, ME	-

TABLE I
A SUMMARY OF EACH LISTENING CONDITION, HIGHLIGHTING THE AUDIO ASSETS THAT ARE ALLOCATED TO EACH DEVICE. *D* CORRESPONDS TO DIALOGUE AND *ME* CORRESPONDS TO MUSIC AND EFFECTS. THE (*) INDICATES THAT THE SIGNALS ARE PERCEIVED THROUGH THE TRANSPARENCY MODE OF THE HEARABLES.

B. Stimuli preparation

Wireless, battery-powered devices such as hearables are a challenge to incorporate into controlled listening test environments. Wireless connections, such as Bluetooth, are less reliable than wired connections, and charging requirements can limit how long they might be used for and so disrupt test processes. With earbud-style devices, there is the additional challenge of ensuring the same fit for all participants, where failing to do so can result in drastically different listening experiences. For these reasons, the listening experience of augmenting TV audio using hearables in transparency mode (Figure 1) was simulated. This was achieved by splitting each combined device condition into two components: the perceived TV signal and the direct hearables signal. A GRAS KEMAR head and torso was used with the hearables inserted and their transparency mode

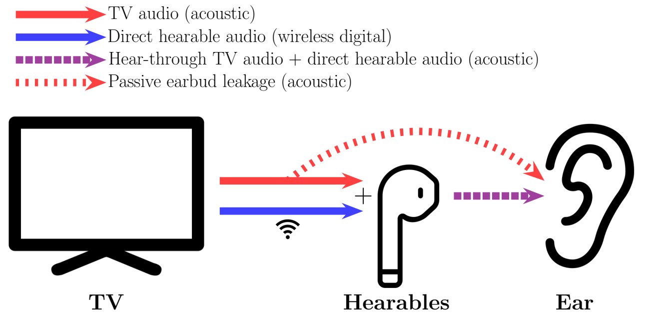

Fig. 1. Concept diagram outlining the listening experience for augmenting TV audio with hearables, showing the relevant audio signal paths.

enabled to create the TV component. This TV component includes the TV audio as captured through the hearables' transparency mode and any passive earbud leakage. The hearables component contains the relevant stereo audio mix (*D, ME, FM*) sent directly to the hearables fitted to KEMAR via Bluetooth and recorded directly with the transparency mode disabled. These two components were then manually synchronised and superimposed to approximate listening to the TV via the transparency mode with additional direct audio from the hearables. Other perceptual evaluation studies of hearables have used similar preparation methods to reproduce hearables signals over headphones [8].

The recordings were made in a quiet recording studio (RT60 = 568 ms) with similar acoustics to an average living room. For the TV, a Mitsubishi 42" LDT422V monitor with SP-422V speakers was used and the hearables were a pair of Sony Linkbuds S. Measured passive attenuation of these revealed the attenuation shelf began in the 400-800 Hz region and attenuation gradually increased up to attenuation of 30 dB at 20 kHz. They were selected as a typical example of a popular commercially available device available at the time of the study, and offered a relatively flat hear-through response. The dummy head was positioned facing the TV at a distance of 2.3m at an angle of 0°. The TV signal was measured at 65 dB SPL at the dummy head. All recordings were normalised to -23 LUFS. For creation of the combined device conditions, the TV component and the hearable component were superimposed and the resulting stimuli normalised to -23 LUFS. This method resulted in the combined device conditions containing both device signals in an equal ratio, and the overall loudness matched that of the single device conditions. In this experiment, the user was not given control over individual stimuli levels and so an equal ratio was selected to ensure that both device outputs were able to be perceived by the listener. This also ensured an equal overall loudness of conditions for these listening tests. Keeping all conditions at equal loudness was deemed more important than faithfully reproducing realistic mixing ratios between the two devices, as loudness is known to influence preference ratings.

C. Listening test

A total of 30 participants took part in a 45 minute-long multiple stimulus listening test, without defined reference and anchor stimuli. The participants, 20 male and 10 female, were between the ages of 22 and 56 and were recruited through university communication channels. All participants self-reported 'normal hearing', gave informed consent and were remunerated for their participation. Ethical approval for the experiment (*Geary20231006*) was granted by the University of York ethics committee. Participants were asked to rate each of the five listening conditions on a scale of 1-100 based on their overall preference where 1 represents lowest preference and 100 represents highest preference, for 10 programme items (3×3 programme types plus one random repeat item). Afterwards they listened to the same programme items but this time rating the conditions on *dialogue clarity*. The order of the conditions in each item and the items themselves were randomised for each participant, and participants could freely switch between conditions while making the ratings. The items were presented at the same playback level over a pair of Beyerdynamic DT990 Pro headphones and were equalised for a flat frequency response using Sonarworks SoundID Reference software. The test interface was implemented using the webMUSHRA framework and was displayed on a laptop, connected with an external mouse and keyboard. After the listening test, each participant completed a short exit survey to gather information on hearable ownership and usage, and attitudes around aspects of shared TV viewing experiences. The questions were either presented as statements where participants answered using a 5-point Likert scale or as questions with free text responses.

IV. Results

Statistical analysis of these results was conducted using separate two-way repeated measures ANOVAs for *overall preference* and *dialogue clarity* attributes. Two factors were assessed: *listening condition* (*TV_ONLY, H_ONLY, FM_FM, D_FM and ME_FM*) and *programme type* (*documentary, drama, and sports*). Data normality was assessed using a Shapiro-Wilks test on the studentised residuals. Further post-hoc analysis was carried out using paired-samples t-tests with Bonferroni correction. Correlation between the attributes was determined through Pearson's correlation testing. For both attributes, the ANOVA result is reported and subsequent p-value reporting relate to pairwise comparisons from the t-tests.

A. Overall preference

Figure 2 presents the results for the *overall preference* attribute. *Overall preference* was normally distributed (p>.05) except for *documentary* × *H_ONLY* (p=.017) and *documentary* × *ME_FM* (p=.024). A statistically significant two-way interaction was identified between *listening condition* and *programme type* (F=20.7, df=8, p<.001, partial η^2=.417). Both *listening condition* (F=96.3, df=4, p=<.001, η^2=.769) and *programme type* (F=18.4, df=2, p=<.001, η^2=.388) factors were significant. The *H_ONLY* condition across all programme types was significantly preferred versus all other conditions

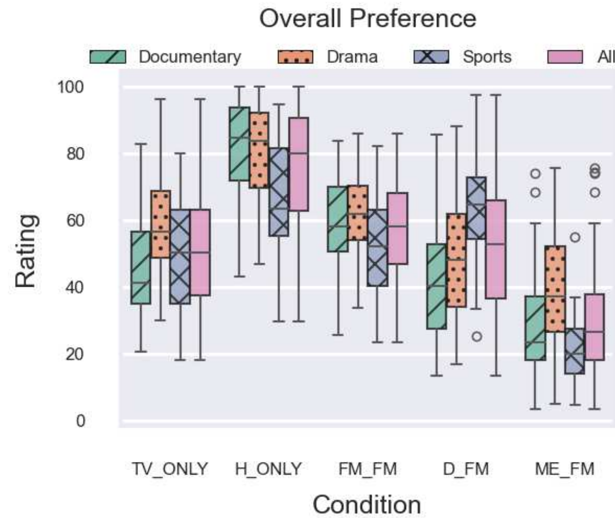

Fig. 2. A boxplot of overall preference ratings for the five listening conditions. Each condition is further grouped by programme type and their combination, represented as the different colour plots. Each boxplot displays the median, the interquartile range (IQR), and outliers. The whiskers include data points distanced 1.5 times the IQR from the nearest quartile.

(p<.001). The combined device conditions *FM_FM* and *D_FM* were comparable to the *TV_ONLY* listening experience. *ME_FM* was rated significantly worse than all other conditions (p<.001). Drama content was significantly preferred over documentary and sports (p<.001) when aggregating all conditions. There is a significant preference for *D_FM* (p<.001) when listening to the sports programme over the other programmes. Interestingly, there was no significant difference in preference between *D_FM* and *ME_FM* conditions for the drama programme (p=.101).

B. Dialogue clarity

Dialogue clarity was normally distributed (p>.05) except for *drama* × *FM_FM* (p=.017), *documentary* × *ME_FM* (p<.001), *drama* × *ME_FM* (p=.027) and *sports* × *ME_FM* (p=.008). A statistically significant two-way interaction was identified between *listening condition* and *programme type* (F=20.9, df=8, p<.001, partial η^2=.419). Both *listening condition* (F=121.5, df=4, p=<.001, η^2=.807) and *programme type* (F=82.7, df=2, p=<.001, η^2=740) were significant factors. Displayed in Figure 3, *dialogue clarity* is broadly similar across all conditions except for the *ME_FM* condition where it is significantly worse than all others (p<.001). There is greater dialogue clarity for the *H_ONLY* condition over *TV_ONLY*, *FM_FM* and *D_FM* conditions for documentary and drama programmes, however, this difference is small (mean rating δ<15). No significant difference was observed in *dialogue clarity* for the *D_FM* condition over the others, aside from within the sports programme where there was a significant improvement (p<.001). *Dialogue clarity* varies significantly across programme types (p<.001) with the drama programme dialogue having the most clarity and the sports programme dialogue having the least. *D_FM* significantly raises the

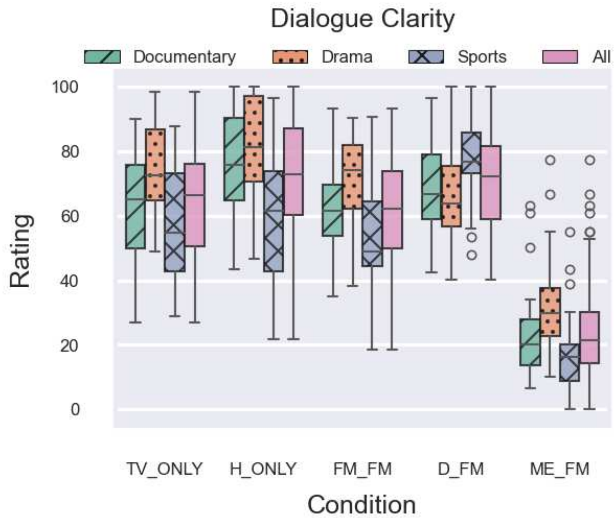

Fig. 3. A boxplot of dialogue clarity ratings for each of the five listening conditions. Each condition is further grouped by programme type and their combination, represented as the different colour plots.

dialogue clarity of the sports programme to above that of the other programmes (p<.001). In contrast, for the drama programme, the *D_FM* condition reduces the clarity of the dialogue, when compared against the single device conditions such as *H_ONLY* (p<.001). Overall preference and dialogue clarity were positively correlated in all combinations of the independent variables (r>0.5, p<.05) aside from the drama × *D_FM* condition (r=.21, p=.26).

C. Survey results

Over half of participants said they owned hearables (16, 53%). For participants whose hearables had a 'transparency mode' (12), when asked how often they use this mode, the most common response was sometimes (4, 33%), with the other responses spread evenly across the scale. The common reasons for using the mode were for the need to pay attention to surroundings whilst outside the home or when communicating with others. Of those that do not own hearables, most either would like to own hearables in the future or would consider it (9, 64%). Most participants positively valued (agree/strongly agree) socialising with others in shared viewing environments (19, 63%), and would like more control over their personal listening experience (21, 70%). There was a slightly more mixed response on whether participants would feel comfortable using hearables to enhance their listening experience during co-viewing (somewhat agree, 43%; neither agree or disagree, 23%; and somewhat disagree 27%). Participants were similarly split between those who would and would not use hearables to augment the TV.

D. Audio analysis

To better understand the influence of the programme type on attribute ratings, Figure 4 shows Mel-spectrograms for a drama and sports programme item under the *H_ONLY* condition. The dialogue in each plot is represented as the visible vertically spaced 'ripples' of the formant patterns in the speech signal. Overall, the dialogue is more prominent relative to the background in the drama item than in the sports item, where the average speech-to-background level difference was 11.3 dB for the drama example and 3.3 dB for the sports example. Not only was the background quieter in the drama programme, it also has less frequency content in speech critical regions (500 - 2 kHz), thus resulting in less masking. This observation may contribute to the observed differences in preference and dialogue clarity ratings between programmes in the single device conditions, therefore stressing the need for programme type, or even individual mixes, to be considered when providing such audio augmentations.

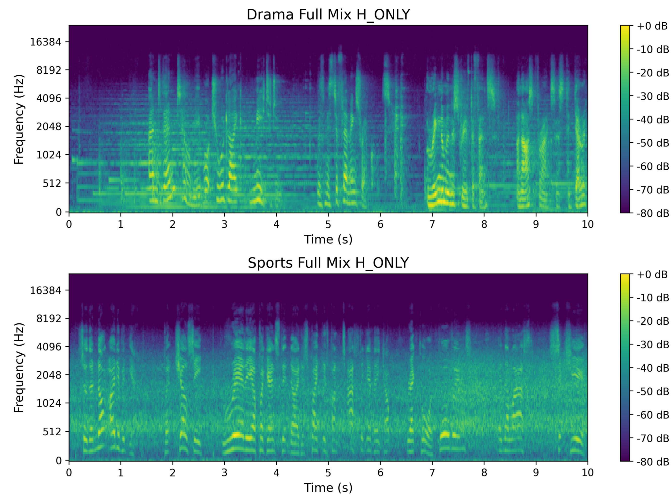

Fig. 4. Mel-spectrograms of the first 10 seconds of the stimuli used for drama and sports programme items under the *H_ONLY* listening condition.

V. Discussion

Previous work identified that mirroring music and other effects on a personal device may provide a more immersive audio experience through louder and improved spatialisation of those assets when compared to just listening on the TV [3]. However, in this experiment, any benefit of mirroring music and effects is negated by quieter dialogue, at the 1:1 loudness ratio of the TV and hearables used for the combined device stimuli. Across the listening conditions, the drama programme is largely preferred over the other programme types. This could be due to the full mix balance being the most satisfactory compared with the other programme types but also could be a simple preference for the drama genre within our participant pool. Concerning the significant preference for *D_FM* when listening to the sports programme over the other programmes, this could be attributed to the perceived importance of the music and effects content to the overall programme. The crowd noise in the sports programme, for example, may be less important or less pleasant to listen to in sports than music and effects in the drama programme. Related to this, the little difference between the *D_FM* and *ME_FM* conditions for the drama programme suggest that both conditions shift the mix balance away from the ideal seen in *H_ONLY*, and that dialogue and the music and effects might be of nearer equivalence in importance than for the other programme types. Furthermore, the full mix balance of the drama programme is already dialogue focused (Figure 4), therefore, additional music and effects does not impact too much on the dialogue.

Dialogue clarity varies significantly across programme types. This is most likely the result of the difference in full mix composition between the three programme types, where not only is the content of the *ME* track different across programmes, but also the relative loudness of the *D* and *ME* within the full mix, as illustrated in Figure 4. Surprisingly, increasing the relative level of the dialogue (*D_FM*) in the documentary and drama programmes provided no additional benefit for dialogue clarity. Given that *overall preference* and *dialogue clarity* are positively correlated, preference for the different conditions is likely influencing the dialogue clarity ratings to some extent. Separating the influence of the transparency mode specifically from that of changes to the dialogue/background mix in the combined device conditions is a challenge when interpreting the results of this experiment. However, information can be obtained by comparing the single device conditions (*TV_ONLY, H_ONLY*) to the *FM_FM* condition, where the mix balance is unchanged across all three conditions. Given that *FM_FM* was largely comparable to the TV only experience for both attributes, the spectral modification of the transparency mode could be tolerable for many listeners when there is some direct audio signal present. Nonetheless, this will depend on the quality of the transparency mode used and the level balance between hear-through and direct audio. An experiment focusing on these details would be useful for confirmation.

VI. Limitations and future work

There are a number of important caveats and limitations to this work that require reflection. Firstly, the full mix of each programme was created by summing the dialogue (D) and background (ME) stems together. This process, while forming a plausible sounding mix, would have undoubtedly omitted additional mix processing such as side-chain compression of the background to the dialogue which would have presumably improved dialogue clarity and overall preference ratings. In addition, the sports programme consisted of assets used in an OBA production. Therefore, no channel-based mix balance appropriate for TV broadcast was available. It is likely that the commentary would have been set at a higher relative level to the background in a broadcast suitable mix, potentially resulting in higher attribute ratings than observed in this experiment. Secondly, as transparency mode characteristics across devices are variable [8], the results of this work may not be generalisable to other devices. Thirdly, the ear canal resonance produced by the KEMAR head and torso was not corrected in our stimuli and may have influenced attribute ratings. The simulation of the combined device listening experience in this experiment assumed near-perfect synchronisation between the hearables and the TV, accounting for a synchronisation solution that would be necessary in a real-world implementation. This circumvents the problem of latency introduced by the transparency mode and Bluetooth transmission. Further work would need to be completed to identify the impact of latency on a listening experience such as this and to consider possible solutions. Finally, as mentioned in Section III, only a single level balance between the hearable and the TV was tested in this experiment. This is unlikely to be optimal and the preferred balance is likely influenced by factors such as what mix content is reproduced on the hearable, the programme type, and personal preference. Ultimately in the real-life experience, the user would have control over the level of the hearable. Further experimentation is required to identify trends in listener preference for level balance between TV and hearables.

VII. Conclusion

The emergence and rise of personal devices and media has diminished the value of TV for many viewers. The TV is designed to facilitate a single shared experience; however, the lack of flexibility to accommodate individual preference in social settings has become a problem for TV experiences. Consumer earbud-style hearables with transparency modes offer a potential solution for enabling personal audio within shared TV viewing environments, without more complex loudspeaker arrays. A listening test was conducted to identify how best to enhance TV audio using a hearable transparency mode by transmitting various parts of the TV mix to the hearable, known as content 'mirroring'. This type of reproduction was compared with other common forms of TV listening, and measured using *overall preference* and *dialogue clarity* attributes. The results show that where the full mix or dialogue is mirrored on the hearable, it is preferred equally to listening to just the TV output, but less preferred than listening just over the hearables. Music and effects mirrored on the hearables was significantly less preferred than the other conditions, and is likely due to the reduction in dialogue quality and level in that condition. Thus, we recommend that mirroring music and effects on the hearables should be avoided or done cautiously as to limit impact on dialogue comprehension. Benefits of mirroring the dialogue on the hearables were observed for dialogue clarity ratings in programme material where there was a low dialogue to background ratio in the full mix. However, no benefit was observed in programmes with good dialogue separation. The type of TV programme was found to significantly affect both measured attributes and is likely to be a consequence of differences in audio mix elements and their respective balance within each programme. *Overall preference* and *dialogue clarity* were found to be significantly positively correlated in all but one combination of the independent variables. Variance across the results, presumably due to personal preference reinforces the importance of personalisation and choice when it comes to TV audio, and emphasises the requirement for personal sound control within shared environments. This sentiment is echoed in the participant responses to the exit survey. Further work should focus on the impact of latency on the listening experience, developing synchronisation solutions, and exploring the social implications of shared viewing scenarios with connected hearables.

Finally, current transparency modes are not designed for this type of application; however, it is hoped that this work raises awareness around the benefits of personal audio control in shared viewing environments, and can contribute to the wider discussion around how concurrent use of personal and shared audio devices, and transparency features could improve or evolve in future iterations.

Acknowledgments

This work is supported in part by the UK Arts and Humanities Research Council (AHRC) XR Stories Creative Industries Cluster project, grant no. AH/S002839/1, and in part by a University of York funded PhD studentship, with additional support from BBC Research & Development.

References

[1] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The internet of sounds: Convergent trends, insights, and future directions," *IEEE Internet of Things Journal*, vol. 10, pp. 11264–11292, July 2023.

[2] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of "hearables"," *Inventions*, 2018.

[3] M. McGill, F. Mathis, M. Khamis, and J. Williamson, "Augmenting TV viewing using acoustically transparent auditory headsets," in *ACM International Conference on Interactive Media Experiences*, IMX '20, (New York, NY, USA), pp. 34–44, Association for Computing Machinery, June 2020.

[4] R. Gupta, J. He, R. Ranjan, W.-S. Gan, F. Klein, C. Schneiderwind, A. Neidhardt, K. Brandenburg, and V. Välimäki, "Augmented/Mixed reality audio for hearables: Sensing, control, and rendering," *IEEE Signal Process. Mag.*, vol. 39, pp. 63–89, May 2022.

[5] P. Crum, "Hearables: Here come the: Technology tucked inside your ears will augment your daily life," *IEEE Spectrum*, vol. 56, pp. 38–43, May 2019.

[6] R. Gupta, R. Ranjan, J. He, W.-S. Gan, and S. Peksi, "Acoustic transparency in hearables for augmented reality audio: Hear-through techniques review and challenges," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, Aug. 2020.

[7] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, "Acoustic transparency in hearables—technical evaluation," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 508–521, 2020.

[8] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, "Acoustic transparency in hearables—perceptual sound quality evaluations," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 495–507, 2020.

[9] P. Coleman, A. Franck, J. Francombe, Q. Liu, T. de Campos, R. J. Hughes, D. Menzies, M. F. Simón Gálvez, Y. Tang, J. Woodcock, P. J. B. Jackson, F. Melchior, C. Pike, F. M. Fazi, T. J. Cox, and A. Hilton, "An Audio-Visual system for Object-Based audio: From recording to listening," *IEEE Trans. Multimedia*, vol. 20, pp. 1919–1931, Aug. 2018.

[10] N. L. Westhausen, R. Huber, H. Baumgartner, R. Sinha, J. Rennies, and B. T. Meyer, "Reduction of subjective listening effort for TV broadcast signals with recurrent neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3541–3550, 2021.

[11] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, pp. 81–91, Mar. 2015.

[12] P. Smith and A. Davis, "The benefits of using bluetooth accessories with hearing aids," *Int. J. Audiol.*, vol. 53, pp. 770–773, Oct. 2014.

[13] H. Isono, "Television viewing for elderly people using bone-conducting headphones," *Proceedings of the Annual Meeting of Japan Ergonomics Society*, vol. 47spl, pp. 354–357, 2011.

[14] J. Francombe, "Get to the centre of the scrum - our immersive six nations rugby trial - bbc r&d." https://www.bbc.co.uk/rd/blog/2021-02-synchronised-audio-devices-sound-immersive, April 2021. (Accessed on 01/03/2024).

[15] D. Geary, J. Francombe, K. Hentschel, and D. Murphy, "Using design dimensions to develop a multi-device audio experience through workshops and prototyping," in *Proceedings of the 18th International Audio Mostly Conference*, AM '23, (New York, NY, USA), pp. 71–78, Association for Computing Machinery, Oct. 2023.

VIII. Biography Section

David Geary David Geary is a PhD student at the AudioLab in the School of Physics, Engineering and Technology at the University of York.

Kristian Hentschel Kristian Hentschel is a project engineer in the audio team at BBC Research and Development. He holds an MSci in Computing Science from the University of Glasgow.

Damian Murphy Damian Murphy received his PhD from the University of York, UK, and is now Professor of Sound and Music Computing at the AudioLab in the School of Physics, Engineering and Technology at the University of York.