



Intelligibility versus comprehension: understanding quality of accessible next-generation audio broadcast

Ben Shirley¹ · Lauren Ward¹

Published online: 15 July 2020
© The Author(s) 2020

Abstract

For traditional broadcasting formats, implementation of accessible audio strategies for hard of hearing people have used a binary, intelligibility-based approach. In this approach, sounds are categorized either as speech, contributing to comprehension of content, or non-speech, which can mask the speech and reduce intelligibility. Audio accessibility solutions have therefore focused on speech enhancement type methods, for which several useful standard objective measures of quality exist. Recent developments in next-generation broadcast audio formats, in particular the roll out of object-based audio, facilitate more in-depth personalisation of the audio experience based on user preferences and needs. Recent research has demonstrated that many non-speech sounds do not strictly behave as maskers but can be critical for comprehension of the narrative for some viewers. This complex relationship between speech, non-speech audio and the viewer necessitate a more holistic approach to understanding quality of experience of accessible media. This paper reviews previous work and outlines such an approach, discussing accessibility strategies using next-generation audio formats and their implications for developing effective assessments of quality.

Keywords Broadcast · Accessibility · Hearing impairment

1 Introduction

It is estimated that around 1 in 6 people have some degree of hearing loss [1] and an ageing population means that this proportion is expected to rise [2, 3]. Over 90% of people with hearing loss have mild to moderate loss [1] and can make some use of television audio. However, surveys indicate people with some hearing loss regularly experience difficulty in understanding speech in television broadcast [4, 5]. Difficulties in understanding speech on TV can be attributed to a number of factors including familiarity with the accent being spoken, clarity of dialogue, and reproduction quality at the consumer device [4, 6, 7]. One of the most commonly reported issues is the balance between different sounds, specifically the balance of speech with non-speech sounds that have potential to mask parts of the dialogue [8].

Improvements in media accessibility are needed to address the barriers currently preventing those with hearing loss from fully accessing broadcast content.

The remainder of this section gives a brief overview of the prevalence and characteristics of hearing loss which affect individuals' engagement with broadcast content. Currently implemented media access services for hard of hearing individuals are also outlined. Section 2 reviews other media accessibility strategies for hearing impaired individuals which have been explored, utilising current broadcasting technologies. Strategies developed using recent object-based audio formats are described in Sect. 3. An alternate approach is presented representing a paradigm shift in accessible audio from intelligibility to a narrative comprehension-based approach. An outline of a proposed accessibility solution based on this paradigm, exploiting end-user personalisation and object-based audio technology is given in Sect. 4. The potential benefits of the approach, as well as the challenges for standardization of an object-based audio personalisation approach to accessible broadcast audio is then made.

✉ Ben Shirley
b.g.shirley@salford.ac.uk

Lauren Ward
L.Ward7@edu.salford.ac.uk

¹ University of Salford, Manchester, UK

1.1 Prevalence of hearing loss

In 2015, 11 million people in the UK, or approximately one in six, were affected by hearing loss according to the ‘Hearing Matters’ report, compiled by the hearing loss charity *Action on Hearing Loss* [1]. These statistics are reflected in other countries with similar demographics, with one in six Australians [9] (2006) and Americans [10] (2003–2004) having some hearing loss. ‘Hearing Matters’ estimates that 6.7 million people in the UK could benefit from the use of a hearing aid [1]. However, generally only a small proportion of these people actually have one fitted (24% in Australia [11]) and many of those who have had a hearing aid fitted, do not use them regularly [11, 12]. *Action on Hearing Loss* project that by 2035, the number of individuals with hearing loss in the UK will rise to 15.6 million, or one in five people [1]. This projected increase is partly due to an aging population [2] as presbycusis, age-related hearing loss [13], is the single largest cause of hearing loss in the UK [1]. Another major cause is noise-induced hearing loss, which is often from occupational exposure [14, 15], though can also be from recreational activities such as concerts [16].

1.2 Characterisation of hearing loss

Hearing loss is often characterised by the location of the impairment within the auditory system: *conductive* hearing loss is due to problems within the ear canal, ear drum or middle ear, *sensorineural* hearing loss is due to problems with the inner ear and *mixed* loss is due to both [17]. Presbycusis and noise-induced hearing loss are the most common types of sensorineural hearing loss [18] and account for the greatest proportion of losses in the population. The severity of hearing loss is usually characterised clinically by pure-tone threshold audiometry and used to group individuals into four categories: mild, moderate, severe and profound [1, 19]. People with mild hearing loss (in the range 20–40 dB) can struggle to understand speech in noisy situations but may still be able to understand speech in quiet unaided [1, 20, 21]. People with moderate hearing loss (41–70 dB) often have difficulty understanding speech with or without background noise without a hearing aid [1]. The majority of people in the UK with some degree of hearing impairment have mild to moderate hearing loss (91.7%) [1]. People with severe (71–95 dB) to profound (> 95 dB) hearing loss often rely on lip-reading, hearing aids or cochlear implants and may identify sign language as their primary language. The large majority of people with a hearing impairment, those with mild to moderate hearing loss, can often enjoy AV broadcast without

using subtitles (also known as closed captions) or signing, and could benefit from improvements that can be made to broadcast audio.

Although a convenient and quick classification method, audiometry does not provide a complete or accurate descriptor for individual hearing impairment. It has been shown that audiometry does not fully explain an individual’s ability to understand speech in noise [22, 23] and that even people with normal audiograms have variable performance in understanding speech in noise [24] which is in some ways analogous to the perception of speech on television broadcast.

1.3 Media access services for people with hearing impairments

Current access services for people with hearing impairments vary considerably across Europe and although some standardization processes have taken place, the availability of accessible broadcast is still highly dependent on territory. In the UK access services such as signing and subtitles are mandatory across a proportion of programming with the amount of programming that must be accessible varying according to audience share. For example the BBC, as a public service broadcaster with a large audience share is mandated to, and provides, subtitling across almost 100% of its broadcast output and signing for 5% [25]. France mandates 100% subtitling and sign language for at least 3 news programmes per day [26], Spain mandates subtitling for 45% of private and 55% of public channels and sign language for 1 (private channels) or 3 (public channels) hours per week. A useful summary of current practice can be found in [27].

Even for standardized accessibility services like subtitling there is considerable dissent on implementation detail, for example the balance of importance between speed of subtitle delivery and synchronization to content. Some interesting research is being carried out in the area, with investigations into subtitling presentations for 360 video formats being pursued by the BBC [28] and the use of so-called dynamic subtitles [29]. However, for many viewers with mild to moderate hearing subtitles and sign language are not an optimal solution to intelligibility of speech and the balance of sound elements within the mix is critical to comprehension of TV content. While loudness levels of programming are now strictly defined by standards [30], the level of the dialogue as compared to other elements is not. Some broadcasters do make reference to dialogue levels in delivery specifications [31], [Netflix, Netflix Sound Mix Specifications and BestPractices vOC-1-1. Tech. rep. (2018)] however, they are often

poorly defined and vary considerably between broadcasters and between programme genres.

2 Accessible audio in channel-based broadcast

This section presents a chronological review of the approaches investigated for use with channel-based broadcast; a term used here to describe broadcast of content which is transmitted as a premixed, linear stream either via terrestrial analogue or digital transmission or digital transmission over IP.

Many commentators on accessibility have suggested transmission of a supplementary audio channel for hard of hearing listeners, often termed a ‘clean audio’ channel, as being an ideal accessibility solution for those with hearing impairment [32–36]. DVB specifications describe ‘clean audio’ as *audio providing improved intelligibility* [37]. Two approaches to achieve this have been proposed: *broadcast mix* where an additional mix with lower levels of non-speech sounds is transmitted by the broadcaster [6, 38] and *receiver mix* which generates a ‘speech enhanced’ mix at the set top box using signal processing [39, 40].

Early work in this area was conducted by Mathers in 1991 with the BBC and *Royal National Institute for the Deaf* among other partners [38]. This used audiovisual clips with either + 6 dB, – 6 dB or unchanged background sound levels. Subjective ratings of quality were elicited from participants and found that a -6dB reduction in the level of background noise produced only a small improvement. It emphasizes the need for controlled, objective studies in the area.

A 1998 position paper by Emmett suggests a separate dialogue only mix would be the optimal solution however given the impracticality of implementing this solution in the production process, a number of post-processing and spectral solutions are also proposed [41]. Shortly after this, in 1999, the DICTION project utilized the R-SPIN test to evaluate processing which reduced background sounds in analogue television [39]. Both objective responses (target keywords) and subjective ratings of clarity were elicited as part of the research. Carmichael’s work indicated that while, at the time of the research, signal processing could make speech sound clearer, it could not improve word recognition performance.

During the transition from analogue to digital broadcasting many researchers sought to leverage the capabilities of the new formats, including the transmission of 5.1 surround sound, to achieve improved speech understanding [40]. The Clean Audio Project, funded by the ITC and then Ofcom, began in 2003 [42] and improved ratings of clarity of dialogue, sound quality and enjoyment (assessed

using blind, forced choice AB comparison) by changing the mix of centre, compared to non-centre, channels for hard of hearing people. Non-centre channel attenuation for improved speech intelligibility using 5.1 broadcast has been standardized in ETSI [43] and referenced in other broadcast standards [44–46]. Although implemented successfully in some territories this method has had limited success owing to variations in the use of the centre channel by broadcasters.

Later in the project the intelligibility of speech presented as a phantom centre compared with a central loudspeaker was evaluated [40, 47] using a modified R-SPIN test [48]. This showed measurable improvement in intelligibility when using a central loudspeaker.

A 2007 BBC experiment into a music-free documentary soundtrack was conducted using the red button service for ‘The Nature of Britain: Secret Britain’ [6]. While positive feedback was received, this exercise highlighted the significantly increased production overheads needed to produce a bespoke music-free mix for hard of hearing people.

In 2008 a special session on Hearing Enhancement at the 125th AES Convention refocused attention onto speech enhancement methods [49]. In one paper from this session, Müsch argues that audio processing can reduce the cognitive effort required for comprehension [50]. His work discussed algorithms which utilized several techniques to detect the presence of speech in centre channel and to attenuate other competing sounds in the same, and other, channels. The aim of the techniques used was twofold; to decrease listener effort and, as a consequence, to improve intelligibility. Müsch argues that the cognitive load used to filter out background sound and ‘clean up’ the speech means that there is reduced attention for the higher-level cognitive processing used to contextualize sentences and therefore fill any ‘gaps’ caused by words not heard. In effect, additional cognitive load is reducing the capacity to take advantage of complementary intelligibility. Also as part of this session, some results from Fraunhofer’s Enhanced Digital Cinema project were reported [51]. In this work, they used pattern recognition, voice activity detection and machine learning methods to enhance the speech. Subjective ratings of speech quality and general sound quality were obtained from two cohorts: one comprised of normal hearing expert listeners and one comprised of hard of hearing children. These showed that the sound quality and speech quality of their proposed method was rated comparable to unprocessed audio by the hard of hearing cohort, while the experienced normal hearing listeners rated the sound quality of the unprocessed audio higher, though with comparable speech quality.

The DTV4All project [52] and the subsequent HBB4-ALL project [53], also took advantage of speech commonly being panned centrally. The project’s open-source ‘center cut’ software [54] to process stereo audio and

extract centrally panned (presumed speech) content. This was assessed in 2 studies with mixed results [53]; the centre cut approach showed little or no effect for some broadcast media though showed improvement in others. However, these experiments were not conducted blind, with participants aware of which content was deemed the ‘clean audio’ mix, this potentially affecting reliability of these results. The premise of the HBB4ALL project, exploiting the capabilities of the new HBB 2.0 specifications to improve accessibility, reinforces a theme that revision in broadcast standards and technology can be an impetus for accessibility improvements and research.

In 2010 the BBC Vision Audibility project repeated a similar experiment to Mathers, providing three mixes with varying background sound levels to participants: + 4 dB, – 4 dB and unchanged [6]. This showed the greater level of background sound inhibited speech understanding but less background sound didn’t always provide intelligibility improvements. This further highlights that a personalized accessibility solution is required.

Around the same time as the DTV4ALL project began and ‘center cut’ software was released, Vickers investigated a frequency domain two to three channel up-mix approach for speech enhancement [55]. Results indicated that existing upmixing algorithms either provided inadequate centre channel separation or produced ‘watery sound’ or ‘musical noise’ artefacts although little perceptual evaluation was undertaken. As recently as 2015, similar centre channel speech enhancement method has been investigated [56]. Objective evaluation using PESQ showed that the algorithm caused no degradation and perceptual testing showed preference for their proposed enhancement method. However, validation was conducted with a small cohort of young listeners which may not have sufficient ecological validity when designing sound systems for (mainly older) people with hearing impairments.

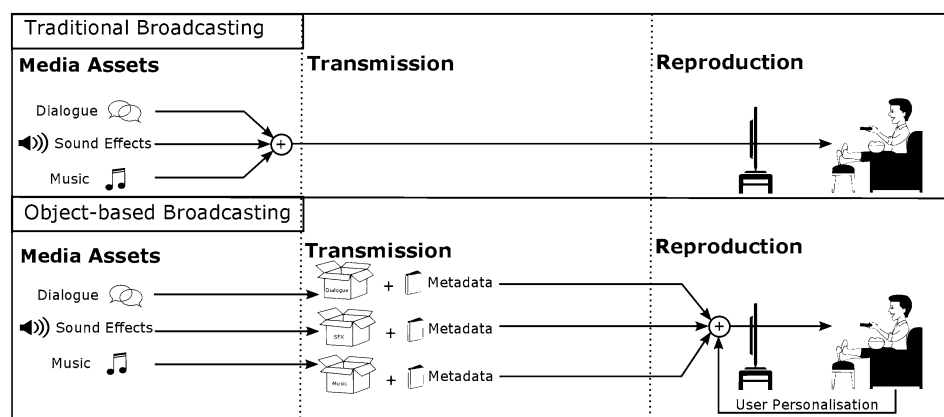
Whenever new standards or significant technological shift occurs, accessibility research has focused on exploiting these new capabilities and this has proven to be a useful approach

[40, 53]. At other times focus has tended to return to speech enhancement methods though many of these have been shown to have little effect on intelligibility. As Armstrong’s useful 2011 review of speech enhancement methods concludes ‘Audio processing cannot be used to create a viable ‘clean audio’ version for a television audience’ [57]. Shirley [58] has rejected this conclusion for the specific case of television broadcast on the basis that broadcast audio is produced subject to known guidelines and conventions which can inform the speech enhancement process. This however is highly dependent on a standardized approach to broadcast audio production (with speech in the centre channel of a 5.1 mix) and broadcasters have not always adhered to a speech-in-centre approach for creative reasons, or to improve downmix compatibility.

3 Accessible audio in object-based broadcast

The development of next-generation object-based audio (OBA) formats presents the opportunity to again exploit technological improvements for accessibility applications. In OBA the challenge of separating speech from competing sounds in the audio mix is potentially solved. Speech signals can be broadcast as audio objects, independent of the remainder of the mix, making speech processing and remixing at the set top box a much more straightforward task. A graphical representation of the two different approaches to broadcasting can be seen in Fig. 1. Instead of broadcasting a mix that is specific to a specific loudspeaker arrangement, OBA facilitates broadcast of individual elements of a sound scene, together with descriptive metadata that indicates how those elements should be presented. Presentation of sound elements that are important to comprehension of broadcast narrative as individually controllable audio objects is technically achievable and can be a useful tool to facilitate a personalised audio experience based on specific access requirements.

Fig. 1 Use of audio objects for personalised audio presentation compared to channel-based audio



Formats such as Dolby Atmos, MPEG-H and DTS:X, although initially presented as facilitating immersive audio mixes and periphonic (with height) sound, are also capable of providing personalised audio presentation to viewers based on individual preference (Fig. 1). Demonstrations of these formats have exhibited features including choice of sports commentator (e.g. home or away) and alternate language options and the personalisation potential of these formats is beginning to be utilised for broadcast accessibility purposes. One such demonstrator came out of the FascinatE project [59], which developed an interactive end-to-end interactive broadcast system that incorporated object-based audio and high-resolution panoramic video. As part of production, separate audio objects were created allowing the viewer to choose their own point of view and for the audio scene to adapt to the visual viewpoint. In one test case at a football game individual audio objects were created for crowd, commentary and on-pitch sounds which could be re-balanced by the viewer based on their individual requirements and preferences. This use of objects was proposed as a means of implementing Clean Audio [58] recommendations by attenuating non-speech objects in the mix based on viewer preference [60].

3.1 Evaluation

Standard measures exist to measure intelligibility of speech which can be useful for assessment of this approach and some have been utilised for accessible or personalised audio. Recent work by Tang has shown that the Binaural Distortion Weighted Glimpse Proportion [61], a development from the original Glimpse Proportion [62], can be effective for evaluating broadcast content and setting appropriate speech to background ratios [63]. Studies by Ward have also explored how the Glimpse Proportion method may be utilised to quantify masking effects of non-speech sound elements [64, 65].

The use of standard metrics for evaluating dialog enhancement has been explored more widely by Torcoli [66]. In this study, nine objective measures from audio and speech coding, speech enhancement and blind source separation applications are compared for their efficacy in detecting the type of likely distortions a dialog enhancement system may produce. The study presents useful reference data to aid selecting appropriate metrics depending on expected distortions but notes that no measure provides a monotonic response to all tested distortions. This means that in utilising objective measures, multiple measures would likely need to be used in complement. Even so, these measures are limited in their objectivity limiting their evaluation to quality and intelligibility, rather than overall comprehensibility or enjoyment of the content.

3.2 The role of non-speech sound

Until recently research aimed at improving broadcast audio for hard of hearing people has adopted a binary paradigm: sound items are either speech, and therefore useful, or they are non-speech that can act as a masker for speech. More recent research has leveraged OBA to go beyond this and investigate the importance of non-speech sound elements for comprehension of broadcast media. Research demonstrated by the University of Salford and DTS at IBC (2015/16/17) and NAB (2016) used metadata to define audio objects based on their type so that less important, potential masker, sounds could be attenuated for people with hearing impairments and the level of sounds that were important to understanding narrative could be increased. The research defined audio objects into 4 categories: speech, music, foreground effects (relating to action on the screen) and background effects, which were developed based on audio object categorisation by Woodcock [67]. An adapted object-based media player was used to present separate level controls for each object category to hearing impaired test participants.

This system was evaluated under laboratory conditions with a cohort of hard of hearing listeners [68]. Results indicated that some hearing impaired viewers found that the *foreground effects* category improved their ability to follow narrative. This approach is based on *comprehension* of narrative, rather than simply *intelligibility* of the speech, in media content. Quantitative studies by Ward have built on this approach, investigating improvements in intelligibility (quantified by word recognition in noise tasks) by introducing non-speech sounds which are salient to the speech. Studies with normal hearing listeners showed that the presence of a salient non-speech sound improved word recognition by 69.5% for sentences with minimal semantic cues and by 18.7% for sentences with high levels of semantic cues (significant at [$p < 0.001$]) [64]. This study was repeated with a hard of hearing cohort which showed that for sentences with minimal semantic cues, the benefit gained by the introduction of salient non-speech sounds correlates to level of hearing impairment [65]. While those with mild-moderate amounts of hearing impairment gained some quantifiable intelligibility improvement from the inclusion of relevant non-speech sounds, those with higher degrees of hearing loss did not. These results support and begin to explain the mechanisms behind the preferences demonstrated in [68].

Recent research has continued this more holistic approach to the problem, focusing on narrative comprehension rather than intelligibility. The following sections describe an object-based approach to enhancing comprehension of broadcast audio using hierarchical narrative importance (NI) metadata and identifies resulting challenges to standardisation of accessible broadcast.

4 An accessibility approach based on narrative importance

While the categories of sound utilised in previous work were useful, these categorisations were limited. For example, 'music' gives an accurate descriptor of the physical characteristics of a sound object but gives no indication of the importance it may have in presenting the narrative of the program. In the case of the category 'music', the music object may be included in a mix for a number of reasons including ambience, but also may an integral part of storytelling. Music in AV media could be diegetic music which the characters interact with, or used to build emotion or tension which would otherwise not be apparent and is integral to the narrative. Sound effects may also be critical to understanding narrative and can even be helpful in improving the intelligibility of speech content [69]. To address this variation of narrative importance a new scheme has been proposed by the authors based on a hierarchical approach to categorising the narrative importance of objects in an object-based audio scene. This *narrative importance* approach also lays the groundwork for an usable and accessible control interface based on gain adjustment of objects relative to their importance in *comprehension* of the media narrative.

A single dial controlling the complexity of audio presentation based on narrative importance (NI) metadata was developed to combine powerful user-personalisation with ease of access [70]. NI metadata was used during media production to hierarchically categorise audio objects within audio scenes based on the role of each sound in conveying the narrative; each object being assigned an NI metadata value between 0 (essential) to 3 (least important). Metadata was generated by the producer, in order to ensure that the producer's intent for the content is maintained, using a VST plugin in a digital audio workstation. Gain adjustments were applied on reproduction to each category of sounds based on the user-selected level on the control. The range of personalisation is then able to transition smoothly between a fully immersive mix at one end of the scale, and a mix containing only the narratively important elements at the other. This allowed the user to adjust the complexity of the reproduced audio mix based on their needs, while ensuring that comprehension of the narrative was always maintained. An example of hierarchical mapping for an example scene is shown in Fig. 2, which uses the audio drama 'The Turning Forest' [71].

Full details of this implementation can be found in [70]. Early results have shown qualitative improvements in intelligibility and comprehension for hard of hearing listeners while maintaining the creative integrity of the producer's work. Although potentially a complex (and

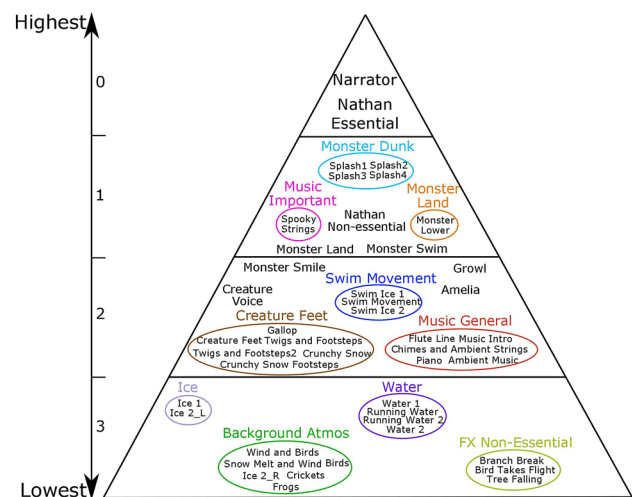


Fig. 2 Hierarchical categorisation of audio objects in a scene from The turning forest

therefore expensive) process the impact on production workflows has been minimized by utilizing an interface very familiar to all sound mixers, tracks containing audio-objects are simply routed to one of 4 busses, depending on the narrative importance of each track. The contents of these busses can then be automatically tagged with appropriate metadata.

4.1 Implementation in next-generation object-based audio

Challenges to implementation fall into 2 main areas that could be described as; technological challenges and cultural challenges. The technology for delivering personalized audio presentation using OBA is in place although broadcasts are at an early stage. Broadcasts using OBA formats have so far been mainly concerned with providing immersive audio, or in delivering existing content using new formats however there is considerable industry activity in developing accessibility personalization features. These efforts to implement technological solutions are sometimes hindered by cultural factors. Production teams, particularly in high-cost productions, are understandably unwilling to cede control of *their mix* to an AV receiver. However, many producers, sound designers and dubbing mixers working in television production have quite a different view. During ongoing research on utilizing NI metadata for personalization sound mixers saw real value to facilitating personalized reproduction. One summed up this view: "Giving the audience the ability to adjust the balance to meet their listening needs could be liberating for the content creator, who could make a mix more like the one they love without having to worry quite so much about those with a hearing difficulty or in a noisy

environment.” It seems likely that television production, driven by ratings from an ageing population, is likely to become the early adopters of such technology.

5 Challenges for standardisation

Technological developments in broadcast audio, and new object-based audio formats, are providing a powerful platform to facilitate accessible audio for people who are hard of hearing. They can reduce or remove the need to ‘unmix’ broadcast audio in the home to enhance speech intelligibility and so can also avoid the artefacts inherent in many speech separation algorithms. Furthermore, object-based audio can facilitate personalised broadcast media based on individual needs and requirements, no longer relying on a one-size-fits-all approach. Results from the studies described earlier suggest that, while it can never replace subtitles for people with more severe hearing impairments, personalised audio presentation can provide a more enjoyable and effective solution than subtitles for many people and the use of production metadata to inform remixing carried out at the set top box allows producer intent to be maintained throughout the chain [70].

While presenting great promise, personalised audio also presents substantial challenges for standardisation activity. Signing can be mandated, mandatory subtitles can have standardised text appearance, font and size. Even intelligibility can be ‘objectively’ measured using the standard metrics described in Sect. 3.1. Comprehension, which the authors argue is a more representative and useful quantity to aim to improve, is a more difficult metric to define and subsequently measure. Signal-based intelligibility measures cannot take into account the saliency or usefulness of non-speech sounds in predicting intelligibility and the usefulness of non-speech sounds in understanding speech and narrative varies considerably between individuals [65].

Standardisation of accessible audio for hard of hearing people following the ‘Clean Audio Project’ [72] has been specified as mandatory where clean audio is provided. An example reference system is indicated allowing broadcasters and broadcast technology companies to implement in a manner most appropriate for their requirements. This has been sufficient for implementation in some territories but not in the majority, although it can be argued that this has been more the result of variation in use of the centre channel in production than on any flexibility in the specifications. It seems likely that, given the likely variability of personalisation implementations, that an approach based on a reference system may be effective. Flexibility will certainly be needed as NI-based personalisation is based on an approach to object-based audio broadcast that is some years off and

there will continue to be much legacy channel-based content for the foreseeable future.

6 Conclusions

Those with hearing loss often have difficulty fully accessing and engaging with broadcast content and this problem affects a significant portion of viewing audiences. Considerable research has been carried out aimed at providing so-called ‘clean audio’ media accessibility solutions for people who are hard of hearing. A review of new approaches triggered by the introduction of new broadcast audio formats has been presented although uptake of these solutions has been limited owing to the problem of ‘unmixing’ speech from broadcast audio. The potential for personalised accessible audio using recent object-based audio formats have been shown to provide an excellent vehicle to improve broadcast audio for people who are hard of hearing and a novel solution has been described using object-audio metadata to rebalance audio based on narrative importance.

Acknowledgements Lauren Ward is funded by the General Sir John Monash Foundation.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Action on Hearing Loss. Hearing Matters Report: <https://www.actiononhearingloss.org.uk/how-we-help/information-and-resources/publications/research-reports/hearing-matters-report/> (2015). Accessed 2 Oct 2017
2. Office for National Statistics. National population projections: 2014-based statistical bulletin (2015) <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2015-10-29#older-people>
3. Roth, T.N., Hanebuth, D., Probst, R.: Prevalence of age-related hearing loss in Europe: a review. *Eur. Arch. Oto-Rhino-Laryngology* **268**(8), 1101 (2011)

4. D. Cohen. Sound matters, BBC College of Production. <http://www.bbc.co.uk/academy/production/article/art20130702112136134> (2011). Accessed 16 June 2016
5. Royal National Institute for Deaf People: Annual survey report **2008**, (2008)
6. Armstrong, M.: <http://www.bbc.co.uk/rd/publications/whitepaper324> (2016)
7. Mapp, P.: In: Proceedings of 141th Audio Engineering Society Convention (2016). <http://www.aes.org/e-lib/browse.cfm?elib=18436>
8. Shirley, B., Ward, L., Chourdakakis, E.T.: In: Personalization of object-based audio for accessibility using narrative importance: ACM TVX. Manchester (2019)
9. Economics, A.: Report for The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and Victorian Deaf Society (2006)
10. Agrawal, Y., Platz, E.A., Niparko, J.K.: Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999–2004. *Arch. Internal Med.* **168**(14), 1522 (2008)
11. Committee, S.C.A.R. et al.: Available at https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Completed_inquiries/2008-10/hearing_health/report/index [Verified 17 November 2015] (2010)
12. Aazh, H., Prasher, D., Nanchahal, K., Moore, B.C.J.: Hearing-aid use and its determinants in the UK National Health Service: a cross-sectional study at the Royal Surrey County Hospital. *Int. J. Audiol.* **54**(3), 152 (2015)
13. Gates, G.A., Mills, J.H.: Presbycusis. *The Lancet* **366**(9491), 1111 (2005)
14. Palmer, K.T., Griffin, M.J., Syddall, H.E., Davis, A., Pannett, B., Coggon, D.: Occupational exposure to noise and the attributable burden of hearing difficulties in Great Britain. *Occup. Environ. Med.* **59**(9), 634 (2002)
15. Sadhra, S., Jackson, C.A., Ryder, T., Brown, M.J.: Noise exposure and hearing loss among student employees working in university entertainment venues. *Ann. Occup. Hyg.* **46**(5), 455 (2002)
16. Maassen, M., Babisch, W., Bachmann, K.D., Ising, H., Lehnert, G., Plath, P., Plinkert, P., Rebentisch, E., Schuschke, G., Spreng, M., et al.: Ear damage caused by leisure noise. *Noise Health* **4**(13), 1 (2001)
17. Smith, R.J.H., Shearer, A.E., Hildebrand, M.S., Van Camp, G.: Deafness and hereditary hearing loss overview. Accessed 20 Aug (2014)
18. Rabinowitz, P.M.: Noise-induced hearing loss. *Am. Fam. Physician* **61**(9), 2759 (2000)
19. British Society of Audiology: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf (2011)
20. Bronkhorst, A., Plomp, R.: Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *J. Acoust. Soc. Am.* **92**(6), 3132 (1992)
21. Pichora-Fuller, M.K., Schneider, B.A., Daneman, M.: How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* **97**(1), 593 (1995)
22. Vermiglio, A.J., Soli, S.D., Freed, D.J., Fisher, L.M.: The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *J. Am. Acad. Audiol.* **23**(10), 779 (2012)
23. Huckvale, M., Hilkhuisen, G.: In: Proceedings of 1st International Workshop on Challenges in Hearing Assistive Technology, Stockholm, Sweden (2017)
24. Füllgrabe, C., Moore, B.C.J., Stone, M.A.: Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Ageing Neurosci.* **6**, 347 (2014)
25. Ofcom's code on television access services. Online (2017)
26. Rapport annuel relatif à l'accessibilité des programmes de télévision aux personnes handicapées et à la représentation du handicap à l'antenne, exercice 2015 (2016)
27. ERGA, Erga special task report on the provision of greater accessibility to audiovisual media services for persons. Tech. Rep. ERGA (2016)12, European Regulators Group for Audiovisual Media Services (2016)
28. Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., Glancy, M.: In: Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, pp. 3–8. ACM (2017)
29. Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., Jay, C.: In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, pp. 103–112. ACM (2015)
30. Loudness Normalisation and Permitted Maximum Level of Audio Signals. Standard, EBU, Geneva, CH (2014)
31. BBC, Best practice guide sound mixing for bbc programmes. Tech. rep. (2018). <http://dpp-assets.s3.amazonaws.com/wp-content/uploads/specs/bbc/TechnicalDeliveryStandardsBBCAudioMixGuidelines.pdf>
32. Ellis, et al., K.: Media International Australia, Incorporating Culture & Policy (153), 53 (2014)
33. Hoeg, W., Lauterbach, T.: Digital Audio Broadcasting: Principles and Applications of DAB, DAB+ and DMB. Wiley, Hoboken (2009)
34. Utray, F., de Castro, M., Moreno, L., Ruiz-Mezcua, B.: Monitoring accessibility services in digital television. *Int. J. Digit. Multimed. Broadcast.* **2012**, 219–228 (2012)
35. Looms, P.: Policy and Marketing Strategies for Digital Media, pp. 43–59. Routledge, New York (2014)
36. Australian Communications Consumer Action Network (ACCAN), Slater, J., Lindstrom, J.I., Astbrink, G.: Broadband Solutions for Consumers with Disabilities. Quest Publishing, New York, 2010
37. ETSI. Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream (2009)
38. Mathers, C.D.: NASA STI/Recon Technical Report N **91** (1991)
39. Carmichael, A.: Evaluating digital on-line background noise suppression: clarifying television dialogue for older, hard-of-hearing viewers. *J. Neuropsychol. Rehabil.* **14**(1–2), 241 (2004)
40. Shirley, B., Kendrick, P.: The Clean Audio project: Digital TV as assistive technology. *Technol. Disabil.* **18**(1), 31 (2006)
41. Emmett, J.: In: 104th Audio Eng. Convention. Audio Engineering Society, Soc (1998)
42. Shirley, B.G., Kendrick, P.: In: Audio Engineering Society Convention 116. Audio Engineering Society (2004)
43. Etsi, T.: ETSI TS **101**(154), 9 (2009)
44. EBU. EBU - TECH 3333: EBU HDTV Receiver Requirements. Geneva (2009)
45. NorDig. Nordig unified requirements for integrated receiver decoders for use in cable, satellite, terrestrial and ip-based networks v2.4. 6.2.4 Clean Audio (2009)
46. Open IPTV Forum. OIPF Release 2 Specification Volume 2 - Media Formats (2011)
47. Shirley, B., Kendrick, P., Churchill, C.: The effect of stereo cross-talk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source. *J. Audio Eng. Soc.* **55**(10), 852 (2007)
48. Bilger, R., Nuetzel, J., Rabinowitz, W., Rzeckowski, C.: Standardization of a test of speech perception in noise. *J. Speech. Lang. Hear. Res.* **27**(1), 32 (1984)

49. Rumsey, F.: Hearing enhancement. *J. Audio Eng. Soc.* **57**(5), 353 (2009)
50. Misch, H.: In: *Proceedings of 125th Audio Engineering Society Convention*. Audio Engineering Society (2008)
51. Uhle, C., Hellmuth, O., Weigel, J.: In: *Audio Engineering Society Convention 125*. Audio Engineering Society (2008)
52. Brückner, W.: Joint recommendations for future access services, joint recommendations for future access services., Tech. Rep. D3.7 / D4.7, DTV4ALL (2011)
53. D4.4 – pilot-b evaluations and recommendations. Tech. rep., HBB4ALL (2016)
54. (2013). <http://www.virtualdub.org/>
55. Vickers, E.: In: *Audio Engineering Society Convention 127*, Audio Engineering Society (2009)
56. Geiger, J.T., Grosche, P., Parodi, Y.L.: In: *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, pp. 869–873 (2015)
57. Armstrong, M.: BBC white paper WHP 190: Audio processing and speech intelligibility: a literature review (2011). <http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP190.pdf>
58. Shirley, B.: Improving television sound for people with hearing impairments. Ph.D. thesis, University of Salford (2013)
59. FascinatE. Fascinate 2010, European commission fp7 grant no. 248138 (2010)
60. Shirley, B., Oldfield, R.: Clean audio for TV broadcast: an object-based approach for hearing-impaired viewers. *J. Audio Eng. Soc.* **63**(4), 245 (2015)
61. Tang, Y., Cooke, M., Fazenda, B.M., Cox, T.J.: A metric for predicting binaural speech intelligibility in stationary noise and competing speech maskers a. *J. Acoust. Soc. Am.* **140**(3), 1858 (2016)
62. Cooke, M.: A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* **119**(3), 1562 (2006)
63. Tang, Y., Fazenda, B.M., Cox, T.J.: Automatic speech-to-background ratio selection to maintain speech intelligibility in broadcasts using an objective intelligibility metric. *Appl. Sci.* **8**(1), 59 (2018)
64. Ward, L., Shirley, B., Tang, Y., Davies, W.: In: *Proceedings of Interspeech 2017: 18th Annual Conference of International Speech Communication Association*. ISCA, Stockholm, Sweden, pp. 2958–2962 (2017)
65. Ward, L., Shirley, B.: In: *Conference on Accessibility in Film, Television and Interactive Media*. York, UK (2017)
66. Torcoli, M., Uhle, C.: In: *Audio Engineering Society Convention 141*. Audio Engineering Society (2016)
67. Woodcock, J.S., Davies, W.J., Cox, T.J., Melchior, F.: Categorization of broadcast audio objects in complex auditory scenes. *J. Audio Eng. Soc.* **64**(6), 380–394 (2016)
68. Shirley, B.G., Meadows, M., Malak, F., Woodcock, J.S., Tidball, A.: Personalized object-based audio for hearing impaired TV viewers. *J. Audio Eng. Soc.* **65**(4), 293 (2017)
69. Ward, L., Shirley, B., Tang, Y., Davies, W. et al.: In: *INTER-SPEECH 2017, 18th Annual Conference of the International Speech Communication Association* (2017)
70. Ward, L., Shirley, B., Francombe, J.: In: *145th Audio Eng. Convention*. Audio Engineering Society, Soc (2018)
71. Woodcock, J., Pike, C., Melchior, F., Coleman, P., Franck, A., Hilton, A.: In: *Audio Engineering Society Convention 140*. Audio Engineering Society (2016)
72. ETSI. ETSI TS101154 V2.4.1 Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.