Personalization of Object-based Audio for Accessibility using Narrative Importance

Ben Shirley
b.g.shirley@salford.ac.uk
University of Salford, UK

Lauren Ward*
l.ward7@edu.salford.ac.uk
University of Salford, UK
BBC Research & Development, UK

Emmanouil Theofanis Chourdakis*
e.t.chourdakis@qmul.ac.uk
Queen Mary University London, UK
BBC Research & Development, UK

ABSTRACT

An increasing incidence of hearing impairment and of reported problems with broadcast audio is leading to an increased demand for personalized audio services. Previous research has treated these issues as a 'speech in noise' problem; sounds are viewed as speech (good) or as competing masker (bad). This binary approach to accessible audio disregards the important role of some non-speech sounds in facilitating understanding of broadcast programme narrative. This work, as part of the S3A project, has taken a more holistic approach to audio personalization using categories of narrative importance to provide complex manipulations of broadcast audio based on *narrative comprehension*, instead of simply *intelligibility*. A simple, intuitive user-interface allows the user to adjust the complexity of audio scenes based on their personal hearing needs, metadata is generated at production using plugins to generate appropriate metadata and audio previews of user-narrative importance settings. This paper outlines the concept of narrative importance, the production tools and the end-user interface designed to deliver it. Response to these tools from target users and production staff are discussed as well as ongoing work.

KEYWORDS

accessibility, broadcast audio, hearing impaired, object-based audio

ACM Reference Format:
Ben Shirley, Lauren Ward, and Emmanouil Theofanis Chourdakis. 2019. Personalization of Object-based Audio for Accessibility using Narrative Importance. In *Manchester, 2019: ACM International Conference on Interactive Experiences for Television and Online Video, Manchester, UK*. ACM, New York, NY, USA, 5 pages.

1 INTRODUCTION

Hearing impairment is estimated to affect around 1 in 6 people in the UK [Action on Hearing Loss 2015] with similar numbers reflected throughout Europe and North America [Agrawal et al. 2008; Shield 2006]. The majority of hearing impaired people have mild or moderate hearing loss (usually defined as between 20dB and 70dB loss in their better hearing ear [British Society of Audiology 2011]). Such listeners still make use of the soundtrack while watching television broadcasts and the majority consider that hearing well when watching TV/video was 'very important' or 'extremely important' (84% in a 2018 study) [Strelcyk and Singh 2018]. An ageing demographic suggests that the proportion of people with hearing loss is likely to rise significantly [Office for National Statistics 2015; Roth et al. 2011].

Media coverage of inaudible television speech in broadcast has become commonplace over recent years [Fullerton 2017; Plunkett 2016], even being debated in the UK Parliament's upper house [Hansard 2017]. A survey by the BBC found that 60% of respondents had difficulty hearing what was said in broadcasts at some point during a single evening [Cohen 2011]. An earlier study by the Royal National Institute for the Deaf reported that 87% of hard of hearing viewers struggled to understand speech on television [Royal National Institute for Deaf People 2008]. This coupled with rising rates of hearing loss makes addressing the challenge of broadcast accessibility increasingly important.

2 PREVIOUS ACCESSIBLE BROADCAST AUDIO RESEARCH

Early accessible audio research has largely approached the challenge as a speech in noise problem; speech is seen as desirable, non-speech is seen as 'background' that can mask speech. In conventional (channel-based) broadcast original 'clean' speech, separate from the rest of the sound, is usually unavailable so creating accessible broadcast audio has often been tackled using speech enhancement techniques which can add unwanted artifacts to the speech [Armstrong 2011].

More recent object-based audio (OBA) formats that are beginning to be used in broadcast have the potential to mitigate this problem considerably. Using OBA it is possible to send individual elements of the broadcast sound scene as independent audio objects with accompanying descriptive metadata

TVX 2019, June 03–05, 2019, Manchester, UK

which describes how the audio should be replayed. User personalization of OBA has been proposed for personalization of sports broadcast [Mann et al. 2013; Mehta and Ziegler 2014; Meltzer et al. 2014] and alternate language provision [Bleidt et al. 2015; Bleisteiner et al. 2018; Brun 2018] and now a key driver for personalization is accessibility [Shirley and Ward 2019; Ward and Shirley 2019; Ward et al. 2018]. Some research has started utilizing OBA to present accessible audio which includes the ability to adjust the balance between 'foreground' and 'background' sound [Bleisteiner et al. 2018; Walton et al. 2016, 2018] and an automated intelligibility control of OBA using measurable objective models of speech intelligibility has been proposed [Tang et al. 2018].

Comprehension or Intelligibility?

Research on accessible audio for broadcast has largely focused on intelligibility of speech, usually defined in terms of number of words recognized [Fontan et al. 2015]. The assumption has been that all non-speech sounds are potentially maskers of speech and are not useful. However understanding of broadcast programmes is dependent on more than simple understanding of speech. Other sounds, such as sound effects and music, can have important roles in signalling, continuity and scene setting. Consider the case of a live broadcast of a football game: speech (commentary) is undoubtedly important however the sound of the referee whistle is also important to understand the game's narrative. Imagine the movie 'Jaws' without its iconic musical score - all of the tension is lost and a scene becomes simply a person swimming in the sea.

Non-speech sound effects have also been shown to assist intelligibility of speech. Recent research [Ward et al. 2017] found that the inclusion of sound effects related to keywords could improve word recognition rates in noise from 35.8% without sound effects to 60.7% with sound effects for normal hearing listeners. This was also shown to be the case for some hard of hearing listeners although the effect was highly dependent on hearing acuity. People with mild hearing loss (20dB to 40dB loss in best ear) had results similar to people with normal hearing.

It can be argued that accessible audio should focus more on *comprehension* of broadcast instead of measures of *intelligibility*. This was the focus of research carried out by Shirley et al in the FascinatE project which utilized a prototype OBA format to allow user-control of levels of separate objects for commentary, crowd and on-pitch sounds in a football game [Shirley and Oldfield 2015]. Further work using the MDA (Multi-Dimensional Audio) OBA format presented drama and sports media clips to hard of hearing participants with separate controls for categories of *speech*, *music*, *background* and *foreground* audio objects [Shirley et al. 2017]. Whereas speech was always set to a higher level than other objects a significant number of participants stated that the *foreground* objects helped to understand the narrative of the media content.

3 NARRATIVE IMPORTANCE AND ITS IMPLEMENTATION

Earlier OBA work showed promise in moving beyond a simple binary (speech vs. non-speech) approach to accessible audio. However the user interface could be considered overly complex; few people would wish to adjust 5 different level controls for each programme as was the case in the research discussed earlier using the MDA audio format (1 gain control for each of 4 categories and an overall volume). The research presented in this paper utilizes hierarchical narrative importance categorization and presents the user with a single control which has different effects on the different categories of sounds.

Audio objects in object-based audio have accompanying metadata. For personalization in accessibility, an additional field of metadata describing narrative importance has been added for each object. Producers tag objects during post-production with appropriate levels of narrative importance based on the importance of the sound in conveying the narrative of the programme. The metadata is retained in the OBA audio stream to the user. The user interface, a single control, is then used to manipulate levels of objects within each category based on user preference set by the control. The remainder of this paper describes the work to implement and evaluate tools which enable this approach to accessible audio.

User Interface

The control acts as an audio *complexity control* for the media and can be seen in Figure 1. At maximum setting the audio mix is as the producer intended, fully immersive and with all objects at default, producer-set levels. At minimal setting the audio mix contains only elements that are essential or highly important to understanding the narrative. Consequently the mix is less complex. Each object category stem is continuously variable throughout the full range of the control with differing attenuation factors applied based on the narrative importance of that object-category.

Production Tools

During post-production audio objects must be tagged by the producer, dubbing mixer or sound designer with metadata appropriate to the function of the audio object in understanding the narrative. The producer-interface has been designed for familiarity and simplicity, in order to avoid introducing additional time, and therefore cost, to media production, and so that new skills do not need to be learned. The process of assigning an object, or an audio track, to a category is

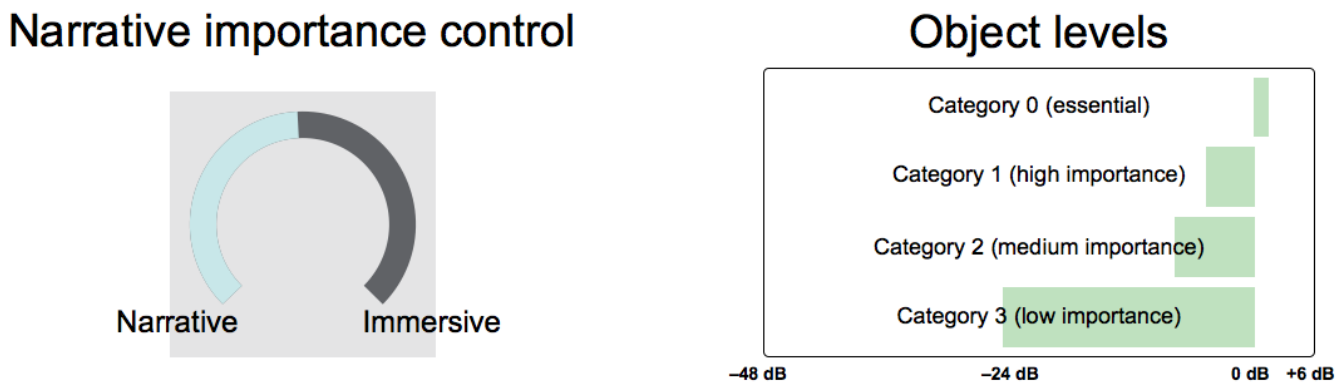

Figure 1: Narrative importance user-control showing relative levels of each object-category

therefore made analogous to assigning the audio track to a mix bus or auxiliary track. As one dubbing mixer described it,"*instead of selecting* L/R *on the channel I just press* 0, 1, 2 *or* 3 *- it's not going to add any more time to my work.*"

The effect has been implemented as a VST plugin for Linux and Windows as well as an AAX Pro Tools plugin. It was developed using the FAUST [Orlarey et al. 2009] signal processing programming language and FAUST2JUCE [Letz et al. 2015]. Figure 2 shows the DSP block diagram of the effect and figure 3 a screen-grab of its operation. It consists of a single dial control (as in the user interface) which ranges from 0 to 100 (101 points) with 0 producing the mix with narrative elements only, and 100 the full immersive mix. The dial controls the gain levels for each of the 4 stereo input tracks $x_{i,c}$ and mixes them to a single stereo output track y. Additionally, values of the dial are smoothed using an exponential envelope in order to avoid 'scratches' due to processing rate being much higher than the control rate.

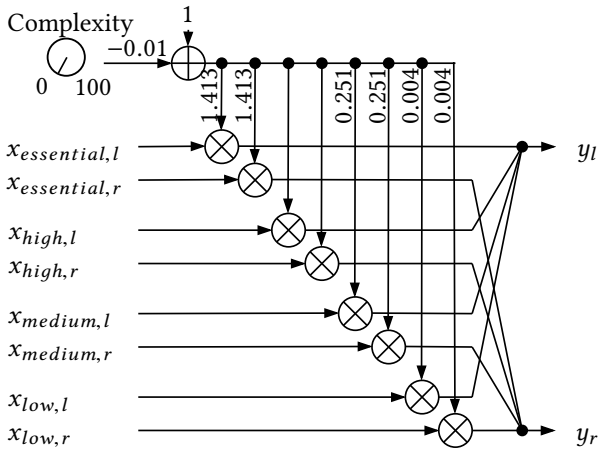

Figure 2: DSP block diagram of the implemented vst effect. Input channels $x_{i,l}$ and $x_{i,r}$ are the left and right channels of track with narrative importance i. The numbers at the top are the narrative importance gain levels converted to linear scale.

Figure 3: Screen-grab of the implemented VST effect

User and Producer Assessment

Demonstrations, focus groups and semi-structured interviews have been carried out with hard of hearing people to identify their response to the prototype user-interface and to identify improvements that can be made [Ward et al. 2018]. Responses indicated a very positive impression of the interface and use of narrative importance and enthusiasm for the additional control it gave them over the audio. Most participants considered that retaining some of the background sounds, even in the least complex personalized mix, helped to retain what they described as the content's 'depth' and 'colour'. Several participants expressed a wish for greater control - for a greater range so that audio-objects in the less important audio categories could be reduced still further where desired. Mix-sessions were undertaken with 2 professional sound mixers to more clearly understand the process by which narrative importance metadata could be generated and how choices were made. Both found the concept easy to understand with one commenting, *'that's how I decide how loud and element should be in the mix anyway - how important*

it is to the story.' The sessions produced useful pointers as to how production workflows incorporating narrative importance metadata could be made more efficient. Neither expressed any wariness about *their* mix being altered by the viewer at home although it is very likely that this sentiment may not extend to premium movie production.

4 DISCUSSION

It is clear that for some hard of hearing people accessibility is more than simply about speech enhancement and that other narratively important elements of the soundtrack can also be important to comprehension. Demonstrations of prototypes of the user-control with production professionals have indicated that utilizing narrative importance metadata in this way need not significantly increase production workload. The next challenge is to identify how narrative importance metadata could be implemented in broadcast and to broadly assess its impact on user experience. To this end some of the current work is focusing on implementation in the Standard Media Player (SMP) in collaboration with BBC R&D. The SMP is a responsive accessible player which delivers BBC media content across News, Sport, Weather, iPlayer, Radio and live events. Other ongoing work is looking at how different producers interpret narrative importance of sounds in TV broadcast and how consistent this interpretation is for individual sound mixers.

5 CONCLUSIONS

This paper has introduced the idea of utilizing narrative importance metadata into object-based accessible audio for people with hearing impairments. It has documented development of both user interface and production tools consisting of VST plugins that can be used in many common digital audio workstations and outlined next steps for further development and implementation.

6 ACKNOWLEDGMENTS

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. Lauren Ward is funded by the General Sir John Monash Foundation.

REFERENCES

Action on Hearing Loss. 2015. Hearing Matters Report. https://www.actiononhearingloss.org.uk/how-we-help/information-and-resources/publications/research-reports/hearing-matters-report/

Y. Agrawal, E. A. Platz, and J. K. Niparko. 2008. Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. *Archives of internal medicine* 168, 14 (2008), 1522–1530.

Mike Armstrong. 2011. Audio processing and speech intelligibility: a literature review. In *BBC Research & Development Whitepaper.*

R. Bleidt, A. Borsum, H. Fuchs, and S. M. Weiss. 2015. Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement. *SMPTE Motion Imaging J.* 124, 5 (July 2015), 1–13. https://doi.org/10.5594/j18579

W. Bleisteiner, A. Silzle, R. Schmidt, T. Liebl, O. Warusfel, M. Ragot, and N. Epain. 2018. *D5.6: Report on Audio subjective tests and User tests.* Technical Report. The Orpheus Project. https://orpheus-audio.eu/wp-content/uploads/2018/07/orpheus-d5.6_report-on-audio-subjective-and-user-tests_v1.3.pdf.

British Society of Audiology. 2011. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking.

Rupert Brun. 2018. Successful Demonstration of Interactive Audio Streaming Using MPEG-H Audio at Norwegian Broadcaster NRK. /www.audioblog.iis.fraunhofer.com/mpegh-nrk/

D. Cohen. 2011. Sound Matters, BBC College of Production. http://www.bbc.co.uk/academy/production/article/art20130702112136134

Lionel Fontan, Julien Tardieu, Pascal Gaillard, Virginie Woisard, and Robert Ruiz. 2015. Relationship between speech intelligibility and speech comprehension in babble noise. *J. Speech, Lang. Hear. Res.* 58, 3 (2015), 977–986.

H. Fullerton. 2017. BBC drama SS-GB criticised for "mumbling" and bad sound quality in first episode. http://www.radiotimes.com/news/2017-02-26/bbc-drama-ss-gb-criticised-for-mumbling-and-bad-sound-quality-in-first-episode

Hansard. 2017. Television Broadcasts: Audibility. https://hansard.parliament.uk/lords/2017-04-04/debates/F84C55A0-3D8B-41F7-A19C-CC216F8C7B0B/TelevisionBroadcastsAudibility

Stephane Letz, Sarah Denoux, Yann Orlarey, and Dominique Fober. 2015. Faust audio DSP language in the Web. In *Proceedings of the Linux Audio Conference (LAC-15), Mainz, Germany.*

M. Mann, A. W. P. Churnside, A. Bonney, and F. Melchior. 2013. Object-based audio applied to football broadcasts. In *ACM international workshop on Immersive media experiences.* ACM, 13–16.

Sripal Mehta and Thomas Ziegler. 2014. Personalized and immersive broadcast audio. In *International Broadcast Convention.* IET.

Stefan Meltzer, Max Neuendorf, Deep Sen, and Peter Jax. 2014. MPEG-H 3D Audio-The Next Generation Audio System. In *International Broadcast Convention.* IET.

Office for National Statistics. Oct, 2015. National Population Projections: 2014-based Statistical Bulletin. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2015-10-29#older-people

Yann Orlarey, Dominique Fober, and Stéphane Letz. 2009. FAUST: an efficient functional approach to DSP programming. *New Computational Paradigms for Computer Music* 290 (2009), 14.

J. Plunkett. 2016. Heard this before? BBC chief speaks out over Happy Valley mumbling. https://www.theguardian.com/media/2016/apr/08/bbc-happy-valley-mumbling-jamaica-inn-sarah-lancashire

Thomas Niklaus Roth, Dirk Hanebuth, and Rudolf Probst. 2011. Prevalence of age-related hearing loss in Europe: a review. *Eur. Arch. Otorhinolaryngol.* 268, 8 (2011), 1101–1107.

Royal National Institute for Deaf People. 2008. Annual Survey Report 2008.

Bridget Shield. 2006. Evaluation of the social and economic costs of hearing impairment. https://www.hear-it.org/sites/default/files/hear-it%20documents/Hear%20It%20Report%20October%202006.pdf

B. Shirley and R. Oldfield. 2015. Clean audio for TV broadcast: An object-based approach for hearing-impaired viewers. *J. Audio Eng Soc.* 63, 4 (2015), 245–256.

Ben Shirley and Lauren Ward. 2019. Intelligibility vs. Comprehension: Understanding Quality of Accessible Next-generation Audio Broadcast. *Universal Access in the Information Society* Special Issue on"Quality of Media Accessibility Products and Services" (2019). [Accepted; In press].

Ben Guy Shirley, Melissa Meadows, Fadi Malak, James Stephen Woodcock, and Ash Tidball. 2017. Personalized object-based audio for hearing impaired TV viewers. *J. Audio Eng Soc.* 65, 4 (2017), 293–303.

Olaf Strelcyk and Gurjit Singh. 2018. TV listening and hearing aids. *PloS one* 13, 6 (2018).

Y. Tang, B. M. Fazenda, and T. J. Cox. 2018. Automatic speech-to-background ratio selection to maintain speech intelligibility in broadcasts using an objective intelligibility metric. *Appl. Sci.* 8, 1 (2018), 59.

T. Walton, M. Evans, D. Kirk, and F. Melchior. 2016. Does Environmental Noise Influence Preference of Background-Foreground Audio Balance?. In *141st Audio Eng. Soc. Convention*. AES, Los Angeles, U.S.A.

Tim Walton, Michael Evans, David Kirk, and Frank Melchior. 2018. Exploring object-based content adaptation for mobile audio. *Personal Ubiquitous Comput.* (2018), 1–14.

Lauren Ward and Ben Shirley. 2019. Personalization in Object-based Audio for Accessibility: A Review of Advancements for Hearing Impaired Listeners. *Audio Engineering Society Journal* Special Issue on Object-Based Audio (2019). [Under Review].

Lauren Ward, Ben Shirley, and Jon Francombe. 2018. Accessible object-based audio using hierarchical narrative importance metadata. In *Audio Engineering Society Convention 145*. Audio Engineering Society.

Lauren Ward, Ben Shirley, Yan Tang, and William Davies. 2017. The effect of situation-specific acoustic cues on speech intelligibility in noise. In *Interspeech 2017*. ISCA, Stockholm, Sweden, 2958–2962.