Proceedings of the 3rd Workshop on Intelligent Music Production, Salford, UK, 15 September 2017

ADAPTING AUDIO MIXES FOR HEARING IMPAIRMENTS

Trevor R. Agus and Chris Corrigan

SARC, School of Arts, English and Languages
Queen's University Belfast
{t.agus,c.corrigan}@qub.ac.uk

ABSTRACT

A hearing impairment can make it harder to pick apart typical soundtracks, whose dialog, music, and sound effects have likely been mixed with normal-hearing listeners in mind. This paper reviews the potential for enhancing audio mixes, with a focus on preserving the original intentions of the sound engineer, even if this involves changing the mix. Hearing aid strategies are contrasted with more extended enhancements that would be possible off-line. The solutions proposed range from remastering from an established mix, through remixing from available stems, to more extensive processing of individual stems. The arrival of object-based audio makes these solutions more feasible and offers the opportunity to test and develop psychoacoustical theories in the complex but controlled world of the sound engineer.

1. INTRODUCTION

Roughly 15% of us have a clinically significant hearing loss in both ears [1, 2]. Television channels currently provide no alternative services for hearing-impaired (HI) viewers, short of subtitles or sign language. This perhaps stems in part from a lack of strategies for adapting audio mixes to be heard through HI ears. This review looks at existing and potential strategies, building on audiological research and hearing aid technologies. The goal of providing an artistically rich experience despite peripheral limitations parallels new directions in audio description [3]. Although this review focuses on television soundtracks, many of the principles apply more broadly to film soundtracks, music and computer games.

Previous research into audio-enhancement strategies for HI listeners has focused on amplifying speech relative to the rest of the soundtrack. Mathers [4] showed that selectively boosting speech by 6 dB made speech seem more intelligible for HI listeners, although only to the same degree as normal-hearing (NH) listeners listening to speech attenuated by 6 dB. When HI listeners set their own levels [5], they boosted speech by around 10 dB relative to the standard mix, although this varied from clip to clip.

Proposed technological solutions have also focused on increasing speech levels in the mix. Applying noise-reduction methods post-production have been relatively ineffective [6–8], but object-based audio may provide easier access to clean speech [5, 9]. Strategies based on a model of loudness or speech intelligibility have been proposed to boost the relative level of speech [10, 11].

Despite the focus on speech, the provision of music and sound effects also contribute to the viewing experience [4]. Thus, this review focuses more broadly on audibility of all components, where possible.

2. FACETS OF HEARING IMPAIRMENTS

The symptoms of hearing impairments vary from person to person, but can be characterized by a raft of psychoacoustical measurements (see [12] for a review).

The diagnostic symptom of a hearing impairment is the inaudibility of quiet sounds. This is typically measured in an audiology clinic as the level of the quietest audible pure tones, which usually varies across the frequency range.

If inaudibility were the only problem, suitable amplification would counteract it. However, more intense sounds require less amplification to attain equal loudness; the loudest comfortable sounds for HI listeners are typically not much more intense than for NH listeners. Therefore, amplifying quiet sounds enough would by default amplify louder sounds too much, causing discomfort or even long-term damage. Therefore, amplification must be combined with suitable dynamic range compression.

Even for comfortably audible sounds, HI listeners can also have a reduced ability to separate out components of sounds that are close in frequency [13]. HI listeners typically require more favorable signal-to-noise ratios [14].

HI ears may also convey less precisely the timing cues [15] that aid pitch perception [16] and sound localization [17], which in turn help distinguish simultaneous talkers [18, 19]. Temporal cues are also implicated in the reduced sensitivity to differences between consonance and dissonance for elderly listeners [20].

Auditory perception of our complex sound-world relies on a diverse range of cues, many of which are affected by the interrelated facets of hearing impairments.

3. THE LIMITATIONS OF HEARING AIDS

Hearing aids provide a rudimentary automatic remastering of audio in real-time (see [21] for a summary). Although many hearing aid specifications are proprietary, they all feature amplification of low-intensity sounds. Compression also avoids over-amplification of highintensity sounds, reacting quickly to onsets to prevent discomfort or damage. The frequency-dependent parameters of amplification and compression are set by trained audiologists based on an individual's hearing loss. Optimal settings also vary depending on the properties of the audio source; speech and music modes can be selected manually or automatically.

Hearing aids process all sounds picked up by their microphones, including speech, music, effects, background noise in the room, and self-generated noise. Compressing these sounds simultaneously means that fluctuations in the level of one component affects all others: a steady sound is "pumped" by level changes in another sound.

Hearing aids need relatively low latencies. Even delays on the order of tens of milliseconds can provide a disturbing echo of the unprocessed sound, particularly when it is the wearer's own voice [22]. This limits the complexity of the algorithms that could manipulate the sound. Additional limitations on computing power come from the batteries, which would be drained faster by more complex algorithms hearing aid algorithms

Pairs of hearing aids are generally not connected. They process the sound independently and cannot work together to preserve binaural spatialization cues.

Despite practical constraints, hearing aids offer a great improvement in some settings and can considerably improve wearers' quality of life. However, for off-line materials, off-line processing offers more potential.

4. REMASTERING FOR HI LISTENERS

Where stems of a soundtrack are no longer available, off-line processing could improve on hearing aids, even using the same principles of frequency-dependent gain and dynamic range compression, but without limitations on bit depth or algorithm complexity.

A further advantage of off-line remastering is that the compression algorithm could map the known dynamic range of the soundtrack to the reduced comfortably audible range of the HI listener. It could prepare for faster onsets rather than reacting to them in milliseconds, suppressing high-amplitude transients without unwanted compression artefacts such as pumping. Rudimentary source separation could even treat different parts of the soundscape (speech or music) separately and appropriately.

5. REMIXING FOR HI LISTENERS

In many situations, it may be possible to enhance the audio starting from its constituent stems, which are often available at late stages of the production process. This allows each stem to be compressed independently [23], avoiding the pumping effects of co-compression.

Adjusting levels of signals independently is one of the main tasks of audio mixing. Perceptually, adjusting levels affects loudness (the perceptual correlate of level) and masking (the reduced audibility of one sound due to the presence of another). Both loudness and masking can be modelled based on the contribution of a sound to each frequency region (e.g. [24]).

Ward et al. [25] generated a rough mix automatically by equalizing the loudness of each stem, as a starting point for a sound engineer; masking can also be manipulated (e.g. [26]). When adapting audio mixes for HI listeners, we can assume that a sound engineer has already balanced the stems for NH ears. Thus, the goal of enhancement could be to maintain the loudness of each stem for HI listeners. Preserving loudness patterns for HI listeners would require a model to predict the loudness for a HI listener (e.g. [27]).

It seems likely that, for many mixes, it would be possible for optimization algorithms to match the normal listener's patterns of partial loudness for a HI listener. However, such an algorithm could reach its limit for particularly dense mixes: the reduced frequency selectivity of HI listeners may mean that neighboring frequency bands would mutually mask to an unavoidable extent. Here, more drastic changes to the mix may be necessary, such as omitting a less important stem (cf. [28, 29]).

Preserving the patterns of partial loudness may be a clear first step in adapting audio mixes for HI listeners, although may omit other psychoacoustically important aspects. For example, "auditory saliency", the ability of a sound to grab the listeners' attention, may not depend solely on loudness. There are some models of auditory saliency (e.g. [30]), but they are difficult to test experimentally. There are also secondary masking effects, sometimes referred to as "informational masking" (e.g. [31]), in which sounds that would be deemed audible by simple models are inaudible due to higher-level factors, such as perceived similarity between simultaneous sounds.

6. AUDIO EFFECTS FOR HI LISTENERS

Up to this point, we have implicitly assumed that the signal should be preserved as far as possible: level changes and filtering should not be abrupt. Greater enhancement could be achieved by more disruptive processing of individual stems. Separate processing of percussive and harmonic sounds improves music for cochlear-implant users [32].

Mixes are already adapted by sound engineers using harmonic exciters, envelope modifiers, transient designers and mutual compression. Where cues are less reliably perceived by HI listeners (e.g. for clarity of speech, pitch, or consonance/dissonance perception), preservation could include exaggeration of the available cues.

Currently signal preservation seems more preferable to applying these more drastic signal-processing techniques. However, this may simply reflect that we do not yet have the tools required to manipulate or enhance the important auditory cues without undesirable side-effects. This in turn may relate to our incomplete understanding of the complex cues that are involved in the perception of even relatively simple sounds (e.g. [33]).

7. SPATIALISATION

Most modern soundtracks are available in stereo, if not more channels. This provides particular challenges and opportunities for enhancement. HI listeners' localization is typically less precise [34]. To preserve spatial perceptions, it may be necessary to exaggerate directional and distance cues, including interaural level differences, interaural timing differences, and reverberation (cf. [35, 36])

Alternatively, these less effective spatial cues could be sacrificed or appropriated to reduce masking. In a complex soundscape, energetic masking can be predicted from whichever ear has the more favorable signal-to-noise ratio, at least to a first approximation [37]. Spatialization can also reduce informational masking [19].

8. PRACTICAL CONSIDERATIONS

Although mixing studios provide some visualizations and statistics of the audio, sound engineers rely primarily on their own ears and judgment. Assuming that they have near-normal hearing, they will not be able to make these judgments on behalf of HI listeners. An HI sound engineer could mix for their own hearing, but perhaps not for others', nor could they perceive the original mix as it was intended.

Various tools could help sound engineers to mix for others' hearing. Reduced audibility could be emulated by mixing in a background noise that is spectrally shaped to mask any sounds that would be inaudible for a given hearing impairment, forcing the sound engineer to work in the narrower dynamic range that remains. Alternating between original and suitably filtered monitors could also restrict the dynamic range. Monitoring with an expanded dynamic range could mimic this aspect of the hearing impairment, but would also introduce unhelpful artefacts. The effects of reduced temporal resolution, increased masking, and reduced clarity could be emulated rather inexactly by blurring the spectrum, which can be achieved by convolving the audio output with low-frequency noise. Even a single enhanced mix for HI listeners poses considerable technical challenges.

Scaling this up for diverse audiences is even more challenging: a broad range of hearing impairment would imply a range of different optimal mixes. This is an argument for automating the enhancement of mixing.

When it is only practical to generate one or two enhanced mixes, it may be necessary to aim at a certain range of hearing impairments, analogous to off-the-shelf reading glasses. Here the twin goals would be to preserve the target perceptions as much as possible for the least sensitive hearing in the range while causing minimal discomfort to the most sensitive hearing. This strategy would also be necessary when several viewers with different hearing are watching together over loudspeakers.

9. ARTISTIC CONSIDERATIONS

Soundtracks are designed, not just to make words and music audible, but to convey a certain atmosphere. In high-end productions, the final audio mix reflects the artistic decisions of a sound designer and likely also the director. They may be reluctant to let others adapt their audio mix.

However, an audio mix that is designed with NH listeners in mind is inevitably changed when heard by HI listeners, perceptually so even in its original form. The enhancements proposed here describe attempts to preserve the artistic intent of the audio mix, minimizing perceptual changes despite differences in peripheral hearing. This is the intent of all the proposed adaptations, no matter how severe: remastering, remixing, and audio effects.

10. BROADER CONTEXT

Enhancing audio for the many HI viewers is likely to have benefits for the broader public. Cinematic productions are watched at home at lower levels, so narrowing its dynamic range should improve audibility while preserving good neighborly relations. In-car listening masks the quietest sounds, limiting dynamic range. Object-based audio allows flexible spatialization, but this could affect the mutual masking of the constituent sounds. It may be necessary to adapt the mix accordingly. Audio-enhancement technologies could also benefit non-native viewers or even simply accommodate viewers' personal preferences.

Finally, efforts to automate audio enhancements are likely to feed back into hearing research, both basic and audiological. The audio-mix paradigm provides a forum of directly applicable research that tests basic psycho-acoustical theories in scenarios that are as complicated as most of our everyday auditory environments. By side-stepping the practical limitations of the hearing aid, a broader range of the parameter space can be explored. Lessons learnt by manipulating the artificial soundscapes of soundtracks could inform the design of future hearing aids.

11. REFERENCES

[1] A. C. Davis, "The prevalence of hearing impairment and reported hearing disability among adults in Great Britain," *Int. J. Epidemiol.*, vol. 18, no. 4, pp. 911–7, 1989.

[2] F. R. Lin, J. K. Niparko, and L. Ferrucci, "Hearing Loss Prevalence in the United States," *Arch. Intern. Med.*, vol. 171, no. 20, pp. 1851–2, 2011.

[3] M. Lopez, G. Kearney, and K. Hofstädter, "Enhancing Audio Description: Sound Design, Spatialisation and Accessibility in Film and Television," in *Proceedings of the Institute of Acoustics*, 2016, vol. 38, no. 2, pp. 174–82.

[4] C. D. Mathers, "A study of sound balances for the hard of hearing," *BBC Res. White Pap.*, 1991.

[5] B. Shirley, M. Meadows, F. Malak, J. Woodcock, and A. Tidball, "Personalized object-based audio for hearing impaired TV viewers," *J. Audio Eng. Soc.*,

vol. 65, no. 4, pp. 293–303, 2017.

[6] A. R. Carmichael, “Evaluating digital ‘on-line’ background noise suppression: Clarifying television dialogue for older, hard-of-hearing viewers,” *Neuropsychol. Rehabil.*, vol. 14, no. March, pp. 241–9, 2004.

[7] M. Armstrong, “Audio Processing and Speech Intelligibility: a literature review,” *BBC Res. White Pap.*, 2011.

[8] C. Uhle, O. Hellmuth, and J. Weigel, “Speech enhancement of movie sound,” in *125th Audio Engineering Society Convention*, 2008.

[9] J. Paulus *et al.*, “MPEG-D Spatial Audio Object Coding for Dialogue Enhancement (SAOC-DE),” in *138th Audio Engineering Soc. Convention*, 2010.

[10] T. Komori, T. Takagi, K. Kurozumi, and K. Murakawa, “An investigation of audio balance for elderly listeners using loudness as the main parameter,” in *125th Audio Engineering Soc. Convention*, 2008.

[11] H. Müsch, “Aging and sound perception: Desirable characteristics of entertainment audio for the elderly,” in *125th Audio Engineering Soc. Convention*, 2008.

[12] B. C. J. Moore, *Cochlear hearing loss: Physiological, psychological and technical issues*, 2nd ed. Chichester, UK: Wiley, 2007.

[13] B. R. Glasberg and B. C. J. Moore, “Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,” *J. Acoust. Soc. Am.*, vol. 79, no. 4, pp. 1020–33, 1986.

[14] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, “The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,” *J. Acoust. Soc. Am.*, vol. 72, no. 6, pp. 1788–803, 1982.

[15] K. Hopkins and B. C. J. Moore, “The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise,” *J. Acoust. Soc. Am.*, vol. 130, no. 1, pp. 334–49, 2011.

[16] G. A. Moore and B. C. J. Moore, “Perception of the low pitch of frequency-shifted complexes.,” *J. Acoust. Soc. Am.*, vol. 113, no. 2, pp. 977–85, 2003.

[17] B. M. Sayers, “Acoustic-Image Lateralization Judgments with Binaural Transients,” *J. Acoust. Soc. Am.*, vol. 36, no. 6, pp. 923–6, 1964.

[18] D. S. Brungart, “Informational and energetic masking effects in the perception of two simultaneous talkers,” *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1101–9, 2000.

[19] R. L. Freyman, K. S. Helfer, D. D. McCall, and R. K. Clifton, “The role of perceived spatial separation in the unmasking of speech,” *J. Acoust. Soc. Am.*, vol. 106, no. 6, pp. 3578–88, 1999.

[20] O. Bones and C. J. Plack, “Losing the Music: Aging Affects the Perception and Subcortical Neural Representation of Musical Harmony,” *J. Neurosci.*, vol. 35, no. 9, pp. 4071–80, 2015.

[21] H. Dillon, *Hearing Aids*, 2nd ed., vol. 30, no. 8. Stuttgart: Boomerang Press, 2012.

[22] M. A. Stone and B. C. J. Moore, “Tolerable hearing-aid delays: IV. effects on subjective disturbance during speech production by hearing-impaired subjects.,” *Ear Hear.*, vol. 26, no. 2, pp. 225–35, 2005.

[23] Z. Ma, B. D. E. Man, P. D. L. Pestana, D. Black, and J. D. Reiss, “Intelligent Multitrack Dynamic Range Compression,” *J. Audio Eng. Soc.*, vol. 63, no. 6, pp. 412–26, 2015.

[24] B. C. J. Moore, B. R. Glasberg, A. Varathanathan, and J. Schlittenlacher, “A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition,” *Trends Hear.*, vol. 20, 2016.

[25] D. Ward, J. D. Reiss, and C. Athwal, “Multi-track mixing using a model of loudness and partial loudness,” in *133rd Audio Engineering Soc. Convention*, 2012.

[26] S. Hafezi and J. D. Reiss, “Autonomous multitrack equalization based on masking reduction,” *J. Audio Eng. Soc.*, vol. 63, no. 5, pp. 312–23, 2015.

[27] B. C. J. Moore and B. R. Glasberg, “A revised model of loudness perception applied to cochlear hearing loss,” *Hear. Res.*, vol. 188, no. 1–2, pp. 70–88, 2004.

[28] A. Tsilfidis, C. Papadakos, and J. Mourjopoulos, “Hierarchical Perceptual Mixing,” *126th Audio Eng. Soc. Conv.*, 2009.

[29] G. D. Kohlberg, D. M. Mancuso, D. A. Chari, and A. K. Lalwani, “Music Engineering as a Novel Strategy for Enhancing Music Enjoyment in the Cochlear Implant Recipient,” *Behav. Neurol.*, 2015.

[30] C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis, “Mechanisms for allocating auditory attention: An auditory saliency map,” *Curr. Biol.*, vol. 15, no. 21, pp. 1943–7, 2005.

[31] T. R. Agus, M. A. Akeroyd, S. Gatehouse, and D. Warden, “Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise,” *J. Acoust. Soc. Am.*, vol. 126, no. 4, pp. 1926–40, 2009.

[32] W. Buyens, B. van Dijk, M. Moonen, and J. Wouters, “Music mixing preferences of cochlear implant recipients: a pilot study.,” *Int. J. Audiol.*, vol. 53, no. 5, pp. 294–301, 2014.

[33] T. R. Agus, C. Suied, S. J. Thorpe, and D. Pressnitzer, “Fast recognition of musical sounds based on timbre,” *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. 4124–33, 2012.

[34] W. M. Whitmer, B. U. Seeber, and M. A. Akeroyd,

"The perception of apparent auditory source width in hearing-impaired adults," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3548–59, 2014.

[35] P. D. Pestana, J. D. Reiss, and Á. Barbosa, "User preference on artificial reverberation and delay time parameters," *J. Audio Eng. Soc.*, vol. 65, no. 1–2, pp. 100–7, 2017.

[36] E. Perez-Gonzalez and J. D. Reiss, "A real-time semiautonomous audio panning system for music mixing," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, 2010.

[37] B. A. Edmonds and J. F. Culling, "The spatial unmasking of speech: Evidence for better-ear listening," *J. Acoust. Soc. Am.*, vol. 120, no. 3, p. 1539, 2006.