BIG PICTURES AND SMALL SCREENS; HOW TELEVISION SOUND RESEARCH CAN WORK WITH, AND FOR, HARD OF HEARING VIEWERS.

L Ward Acoustics Research Centre, University of Salford, Manchester, UK
B Shirley Acoustics Research Centre, University of Salford, Manchester, UK
WJ Davies Acoustics Research Centre, University of Salford, Manchester, UK

1 Introduction

Hearing loss affects one in six people in the United Kingdom and, given an ageing population, this figure is increasing.[1] Numerous studies highlight that improvements in the intelligibility of television sound are required to increase television's accessibility to individuals with hearing loss.[2–4] Recent developments in broadcast technology, in particular the advent of object-based broadcasting, show the potential to deliver real improvements in television accessibility.[4–7]

Effective strategies to improve the intelligibility of television sound are predicated on intelligibility evaluation methods which are ecologically valid, repeatable and representative of the varied characteristics of hearing loss. This paper presents a review of such evaluation methods in the context of their use in, or applicability to, television sound research. The paper first outlines the prevalence and characterisation of hearing loss, as well as definitions of intelligibility. Current problems with television accessibility for hard of hearing viewers and how object-based broadcasting may offer solutions are summarised. Intelligibility evaluation methods for hearing impaired individuals and objective metrics of intelligibility are then described. The use of such methods previously in television sound research, in addition to the use of related subjective methods, is reviewed. The repeatability, ecological validity and how these methods may be sensitive, or insensitive, to the characteristics of hearing loss are then examined. Finally, the importance of including end-users in the research and evaluation process is discussed.

2 Hearing Loss

2.1 Prevalence

The 'Hearing Matters' report, compiled by the hearing loss charity *Action on Hearing Loss* in 2015 indicates that 11 million people, or approximately one in six, in the UK are affected by hearing loss.[1] These statistics are mirrored in countries with similar demographics, with one in six Australians[8] (2006) and Americans[9] (2003-2004) having some hearing loss. 'Hearing Matters' estimates that in the UK 6.7 million people 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[10]) and many of those who have had a hearing aid fitted, do not use them regularly.[10,11] *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 ageing population[12] 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 like concerts.[16]

2.2 Characterisation of 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 common types of sensorineural hearing loss.[18] Clinically the severity

of hearing loss is characterised by pure-tone threshold audiometry and used to group individuals into four categories: mild, moderate, severe and profound.[1,19] Those with mild hearing loss (in the range $20-40$dB) generally struggle understanding speech in noisy situations but may still being able to understand speech in quiet unaided.[1,20,21] Those with moderate hearing loss ($41-70$dB) often have difficulty understanding speech under any condition without the assistance of a hearing aid.[1] The majority of people in the UK with some degree of hearing loss have have mild to moderate hearing loss (91.7%). Those with severe ($71-95$dB) to profound hearing loss (> 95dB) often rely on lip-reading, powerful hearing aids or cochlear implants and may utilise sign language as their primary language.
Audiometric thresholds provide simple tools for discussing hearing loss severity. However, it has long been acknowledged that audiometric thresholds alone do not fully account for the variability in individuals' ability to understand speech in noise.[22–24] It has been demonstrated that even listeners with normal audiometric thresholds can vary significantly in their ability to understand speech in noise.[25,26] This difference between audiometric thresholds and speech in noise performance was modelled by Plomp in 1978,[24] who defined two loss classes: Attenuation, loss due to the reduction in perceived volume, as a result of reduced audiometric thresholds and, Distortion, which is comparable with a decrease in the effective speech to noise ratio. Other works have termed this Distortion class supracochlear loss[23] or suprathreshold loss.[27,28] The remainder of this work will utilise the term suprathreshold.
The possible contributors to this suprathreshold include tinnitus,[29] loudness recruitment (reduced dynamic range),[30] loss of temporal fine structure[31] or reduction in frequency resolution,[32] as well as cognitive and age related factors.[26,33] Some of these factors may be caused by 'hidden hearing loss' hypothesised to be from cochlear neuropathy (the loss of the high-threshold auditory nerve fibres).[34,35] This variety of factors, and their interactions, highlight that each individual's hearing loss, and its causes, are quite unique.

3 Speech Intelligibility

Speech intelligibility is often defined as the proportion of words which are correctly heard, strictly differentiating it from comprehension.[36] It can also be defined as the proportion of words understood,[37] which incorporates elements of comprehension and quality. This work will utilise the latter, broader definition. The literature distinguishes between two forms of speech intelligibility: signal-dependent , where the ability to retrieve the message is based solely on the speech signal, and complementary, which utilises other non-speech cues from the speech signal, such as syntax and semantics, as well as multi-modal cues such as facial expressions.[38] These complementary cues have been shown to play a greater role in speech perception when hearing is challenged, either by impairment or masking from competing sources.[39,40] It is theorized that improvement in intelligibility given complementary cues is due to the manner in which the brain composes perceptual auditory objects: using expectations of what it believes the object will be in order to predict parts of the object for which no input signal is currently available.[41]

3.1 Complementary Intelligibility

A significant body of research exists demonstrating the release from masking that the inclusion of complementary intelligibility cues can produce. These cues include, but are not limited to, spatial separation,[42] contextual cues[43,44] priming the listener with the speaker's voice or speech content,[45,46] and multi-modal cues.[47,48] Complementary intelligibility cues are particularly pertinent to television sound research as television content uses a rich array of these cues: visual context, context through sound effects, dialogue based context as well as accessibility services like subtitling and audio description.
Spatial release from masking, or the benefit on speech intelligibility gained when the masker and the target speech are spatially separated, has long been documented for normal hearing listeners.[42] Hearing impaired listeners also benefit from spatial release from masking, though to a reduced degree.[49] Research by Bilger in 1984 showed the benefit that contextual cues have on speech intelligibility.[44] It showed that speech recognition in noise by older listeners with sensorineural hearing loss more than

doubled when predictability of the speech was increased (from recognising 37% of keywords up to 76%). Recent work by Ward[43] has shown the inclusion of non-speech contextual cues, in the form of sound effects related to the keywords, improves word recognition rates from 35.8% to 60.7% for normal hearing listeners. Follow-on work has demonstrated that the effect is present for some hard of hearing listeners: those with mild loss exhibit comparable levels of benefit as normal hearing listeners.[50] It was also shown, for low predictability speech, that the degree of benefit reduces as a listeners' better ear hearing decreases. Studies have investigated the multi-modal interaction of different complementary intelligibility cues, such as the interaction between speech and visual cues for normal hearing listeners.[47,48] Aguert showed that by the age of five, children can find utility in the situational context of speech.[48] Research by Spehar has investigated the effects of different types of contextual cues, showing that participants benefited from both visual and speech-based context.[47]

4 Broadcast Accessibility for the Hard of Hearing

4.1 Problems with Intelligibility

Complaints about the intelligibility of dialogue on television have abounded of late.[51,52] However such complaints are not new, with the first BBC research paper on the issue published over 25 years ago.[53] A more recent BBC study reported that nearly 60% of viewers had trouble hearing what was said in some TV programmes.[3] A survey by the *Royal National Institute for the Deaf* in 2008 reported that 87% of hard of hearing listeners struggled to understand speech on television.[2]

The BBC study identified four key factors that made speech hard to understand: clarity of speech, unfamiliar or strong accents, background noise, and background music.[3] This indicates that at least some intelligibility problems on television are fundamentally speech in noise problems. Whilst individually any of these factors can negatively impact on understanding, when the factors are combined their effect is compounded.[54] In a recent BBC White Paper, Armstrong defines a problem space, elaborating on these factors and highlighting the numerous points in the broadcast chain where intelligibility can be degraded.[4] These range from the original performance and content capture, through production and broadcast to reproduction in the home. A similar problem space defined by Mapp also includes three listener-based factors which can affect intelligibility: hearing acuity, attention and alertness, and familiarity/fluency of language.[55] Armstrong considers these human factors in the terms of individuals' 'media access needs',[5] of which there are two types: sensory needs and cognitive needs.

For truly intelligible television audio all these possible points of failure need to be managed. This review will focus on sensory needs, what Mapp terms 'hearing acuity'.[55] This will include both cochlear and suprathreshold loss, as both these types of loss play a role in determining how problematic speech in noise scenarios are for individual listeners. The advent of object-based broadcasting, and personalisable content, gives a basis for significant improvements in meeting individuals' sensory media access needs. In particular, it allows viewers to mitigate any challenges they may have understanding speech against the background sounds through the ability to alter the volume of different sound elements to their preference.

4.2 Object-Based Broadcasting for Accessibility

In object-based broadcasting, different broadcast components including speech and non-speech audio elements can be treated as independent 'objects'. These objects, transmitted along with metadata, remain separate until they are rendered at point of service based on their associated metadata.[56,57] This forms the basis for personalisation methods, which allow for separate control of non-speech audio objects by the end-user. With such advances, the challenge for delivering accessible television sound now lies less in technical delivery but in understanding what listener requirements for accessibility are.[4] In particular, object-based broadcasting allows an end-user to adjust the balance between not only speech and non-speech sounds, but between individual groups of non-speech sounds.[7] This significantly increases the possible customisable parameters. As such, there is a need for a greater understanding of the appropriate balance of these non-speech objects which gives an optimally intelligible mix for different hard of hearing listeners. As some of these objects, such as sound effects, can often play narratively important roles within the content, evaluation of their effect on intelligibility needs to consider the potential for these objects to act as both complementary intelligibility cues and as maskers. Evaluating any implemented object-based broadcasting accessibility strategy presents three challenges: ensuring that the intelligibility evaluation methods are ecologically valid, repeatable and representative of the varied characteristics of hearing loss that may affect intelligibility.

5 Evaluating Human Intelligibility

As previously discussed, pure-tone audiometric thresholds do not accommodate the variability in individuals' speech in noise performance.[22,26] Speech perception in noise tests aim to evaluate this. There are a large variety of these tests including, but not limited to, the Hearing in Noise Test,[58] Listening in Spatialised Noise-Sentences Test,[59] Oldenburg Sentence Test,[60] Bamford-Kowal-Bench Speech in Noise Test[61] and the Revised Speech Perception in Noise Test (R-SPIN).[44,62] These tests take two main forms: everyday sentence tests, using meaningful real-life sentences[44,61] (e.g. the sentence 'The clown has a funny face'[63]) and matrix sentence tests, which follow a strict sentence form for each stimuli and require training[60] (e.g. 'Thomas wants nine cheap beds'[64]). Regardless of type, these sentences are usually phonetically balanced across the stimuli and validated to ensure repeatable results. These tests often follow an adaptive paradigm, which varies the speech to noise ratio until the 'Speech Reception Threshold' (SRT), the point at which 50% of the speech is intelligible, is determined.[58,60,65] The alternate approach is to utilise a static signal to noise ratio and quantify performance by the percentage of words correctly identified.[62,66] The most commonly used maskers are multi-talker babble[62] or speech shaped noise.[66]

5.1 Complementary intelligibility

Many speech perception in noise tests also facilitate the evaluation of complementary intelligibility factors. The Listening in Spatialised Noise-Sentences Test accommodates spatial release from masking[59] by using binaural reproduction of co-located and spatially separated speech and maskers to calculate the spatial advantage gained from separated sources. The effect of semantic context on intelligibility is quantified by the R-SPIN.[44,62] This is achieved through the use of high and low predictability everyday sentence stimuli, where the speech preceding the keyword in these sentences either gives the listener clues to the keyword, e.g. *'Stir your coffee with a* ***spoon****'*, or no clues. e.g. *'Bob could have known about the* ***spoon****'* (where the keyword is noted in bold). The R-SPIN has also been adapted to investigate the effects of non-speech contextual cues on intelligibility, through the addition of relevant sound effects to some of the stimuli.[6,43] The GRID corpus is an audio-visual corpus which has controlled audio-visual stimuli and matrix test sentences.[67] Whilst not developed specifically as a evaluation tool for hearing impaired populations, it allows for the effect of multi-modal cues on intelligibility to be evaluated. The effect of static visual situational context, in the form of an illustration prior to the target sentence stimuli, is evaluated by the Illustrated Sentence Test.[47,68] The stimuli following the illustration can be presented in an auditory-only modality to evaluate intelligibility or in a visual-only modality to evaluate lip-reading.

6 Objective intelligibility metrics

Objective intelligibility metrics are tools that produce proxy measures for human intelligibility judgements. These metrics are based on the effects that noise, signal processing or a space may have on signal features. The way these metrics account for these effects fall into two main types.[69] The first type are those which quantify the *masked audibility* of speech in noise, utilising estimates of both the speech and noise to determine masking level. Examples of these include the Articulation Index,[70] Speech Intelligibility Index (SII)[71] and the Glimpse Proportion.[72] In the case of the Glimpse Proportion, this masked audibility is evaluated by determining the number of masked spectro-temporal segments which are above a pre-set audibility criterion (often 3dB).[72] The other type of metric quantifies the *distortion* caused by the masker by measuring the similarity between the clean and noisy speech.[69] This type of metric includes the Normalised-Covariance Measure,[73] the Coherence Speech Intelligibility Index (an extension of the SII),[74,75] the Modulation Spectrum Area[76] and the Short Term Objective Intelligibility metric (STOI).[77] In the case of the Coherence Speech Intelligibility Index the distortion is measured by determining the degree to which the noisy or modulated output of a system is linearly related to the clean input.[74,75] The Speech Transmission Index (STI) is also a metric of the distortion type and calculates the modulation transfer function of the transmission path.[78] It is most commonly used to evaluate the effect of room acoustics rather than noise on speech intelligibility, but can be utilised for speech in noise evaluations.[37] The SII[71] and the STI[37] have been standardised. These metrics can either be intrusive, which require both the noisy speech and a clean reference signal,[71–73,77] or non-intrusive, which do not require a clean reference signal.[76,79]

As automatic speech recognition techniques have significantly improved over the past decade, matching human recognition levels for some tasks,[80] the concept of automatic speech recognition based intelligibility metrics has seen renewed attention.[81–83] One approach is to utilise the speech recognition algorithm to re-synthesise the clean reference signal required by intrusive metrics such as the STOI.[83] An alternate approach is to design speech recognition systems which use the same auditory models as humans.[84] A recent study has shown that the relevant features utilised by automatic speech recognition coincide with those used by humans, in particular the spectral peaks of the speech signal and dips in the noise masker.[80] Through shared auditory models automatic speech recognition systems can be taught to make predictions of human intelligibility.[84]

6.1 For hearing loss

Many intelligibility metrics have either been adapted for, or specifically developed to, model hearing loss. Humes demonstrated that the STI could predict hearing impaired performance by modelling elevated audiometric thresholds as internal noise.[85] The Modulation Spectrum Area metric, which developed from the STI's modulation transfer function, has been used directly as an intelligibility correlate for cochlear implant users in reverberant conditions.[76] The SII, though developed for normal hearing listeners, has also been used to predict SRTs in noise and different room acoustic conditions.[86] A number of adaptations to the SII have been explored to determine its efficacy for predicting hearing impaired listeners' SRTs in fluctuating noise, incorporating hearing loss as an input parameter to the model, yielding small improvements to the original SII's predictions.[87]

Metrics which have been specifically developed for use with hearing impaired listeners include the Hearing Aid Speech Quality and Perception Indicies (HASQI and HASPI respectively)[88,89] and the Normalised Covariance Measure.[73,90] Hearing loss is often modelled in objective metrics, through widening of filter bandwidths, to represent loss of frequency selectivity and attenuating of filter banks, to represent loss of hearing sensitivity.[91] In the HASQI auditory model, which has been extended to form the basis for predicting both intelligibility (HASPI) and quality (HASQI), loudness recruitment is also accounted for by a reduction in dynamic range compression.[92] The Normalised Covariance Measure is based upon a perceptual model which can accommodate different processing deficits, including an individual's temporal forward masking ability.[73]

Approaches based on speech recognition have also been designed to accommodate for hearing impairment.[81,82] Fontan aims to develop a system which can predict intelligibility and comprehension test performance for hard of hearing listeners.[82] However their system has only been validated with normal hearing listeners and stimuli which has been altered to mimic varying degrees of presbycusis. The Framework for Auditory Discrimination Experiments (FADE), developed at the University of Oldenburg, has been shown to successfully predict the outcome of matrix sentence tests and psychoacoustical experiments.[81] The system utilises a front-end which models hearing impairment, utilising Plomp's Attenuation and Distortion model.[24] The Attenuation is modelled similarly to other metrics which reduce the level in each frequency band based on the audiogram.[81] The Distortion, or suprathreshold loss, is modelled as uncertainty by adding Gaussian white noise with a variable standard deviation. This variable value can be set based on a number of 'typical' audiogram shapes, or can be determined with an independent data set.

7 Evaluating Intelligibility in Television Sound Research

7.1 Previous Approaches

The type of stimuli described in Section 5 has seen some use in television sound research. The DICTION project utilised the R-SPIN to evaluate their developed processor, which removed background sounds in television for improved intelligibility.[93] They elicited both objective responses (target keywords) and subjective ratings of clarity however it showed that the processor did not meet its aims. In the Clean Audio project the R-SPIN stimuli were adapted to evaluate the effect speech presented as a phantom centre compared with a central loudspeaker had on intelligibility.[94] This work demonstrated that a central loudspeaker had a measurable intelligibility gain over a phantom centre. Other experiments within the Clean Audio project elicited subjective ratings of clarity of dialogue, sound quality and enjoyment using blind AB comparison.[95] Subjective ratings of speech quality and general sound quality were also used by Fraunhofer in the Enhanced Digital Cinema project.[96] As in the DICTION project, the algorithm was shown not to improve quality of speech or overall sound quality. These studies all involved the target population, either hard of hearing or elderly, within their research process.

A handful of researchers have made use of objective metrics in their studies. Müsch proposed that the Speech Intelligibility Index could be used within a speech enhancement algorithm to evaluate intelligibility of 5.1 broadcast content but did not validate it.[97] Work by Mapp evaluated the effect which the home listening environment may have on the perceived intelligibility utilising the STI.[55] Whilst not a listener centred effect, Mapp leveraged the STI to show the majority (of tested) living rooms had good intelligibility and did not vary much. This suggests that the room effects are likely not a major cause of poor intelligibility of television and film content.

7.2 Approaches for Object-Based Broadcasting

Object-based broadcasting's personalisation potential has been used to evaluate intelligibility and quality subjectively by allowing participants to choose their preferred balance of various object categories.[7,98–101] These levels are elicited by asking the participants to: optimise their understanding of the content,[7] make the speech easy to follow[98] or set preferred balance given the broadcast content[99,100] or in light of background noise.[101] All studies used audio-visual material, however only one of these studies had hard of hearing participants.[7] The common result from all these studies was significant interpersonal variation, with normal hearing listeners often preferring higher levels of background noise to enhance the feeling of 'being there'.[100] Torcoli's evaluation methodology differs from the others in that it utilises a validation phase after the preference selection.[98] Termed the Adjustment/Satisfaction Test, the participants are asked to judge how satisfied they are with their level balance, with reference to the default mix.

Some research has exploited the speech in noise tests described in Section 5. In a follow-up experiment to earlier Dialogue Enhancement work by Fraunhofer,[99] Fuchs utilised the Oldenburg Sentence Test to evaluate a dialogue enhancement strategy for hearing impaired listeners.[102] This implementation used both speech-shaped noise and 'applause' type noise and both hard of hearing and normal hearing listeners. This study showed that the dialogue enhancement strategy improved intelligibility for hard of hearing listeners, coming close to doubling word recognition. Research by Ward has also used speech in noise tests. This research evaluated the effect of non-speech audio objects on intelligibility utilising the R-SPIN and showed they improved word recognition for some hard of hearing listeners.[43,50] Recent work by Tang has shown that the Binaural Distortion Weighted Glimpse Proportion,[103] a development from the original Glimpse Proportion,[72] can be effective for evaluating broadcast content and setting appropriate speech to background ratios.[104] The latter two stages of the study investigating this employed a test methodology similar to the Adjustment/Satisfaction Test,[98] though the stages were implemented as two separate experiments. Participants in the first stage were asked to select optimal balances between foreground speech and background sound before these levels were validated in a second experiment. Studies by Ward have also explored how the Glimpse Proportion may be utilised to quantify masking effects of non-speech sound elements.[43,50] The STOI metric has been investigated, among other performance and quality metrics, for detecting distortions incurred by dialogue enhancement systems.[105]

8 Discussion

It is evident that advances in broadcast technology, in particular object-based broadcasting, has made intelligibility evaluation more complex. Furthermore, many of the standardised approaches to evaluating audio systems, such as MUSHRA, are not suitable for innovative technology such as personalisable content.[106] However, as this review highlights, there are a large number of evaluation methods available which, to varying degrees, meet the criteria of being ecologically valid, repeatable and representative of hearing loss characteristics.

8.1 Ecologically Valid

Ecological validity is defined here as how accurately the experimental conditions mimic real-life scenarios and subsequently, how well the results will likely generalise. Personalisation based evaluations[7,98–101] are the most ecologically valid methodologies; their experimental conditions most closely match the anticipated user experience. Furthermore through utilising real audio-visual media, these tests best accommodate for the inherent complementary intelligibility factors within television content.
The ecological validity of speech in noise tests depends on the type of sentences and noise used. Everyday sentence tests like the R-SPIN,[44,62] as compared with matrix tests, provide sentence stimuli more similar to television dialogue. Multi-talker babble noise also adequately represents many common problematic background sounds within television content, such as the ambience in crowded places like restaurants or shows recorded in front of live audiences. This use of everyday sentences and representative noise likely contributes to R-SPIN's regular use in television research.[43,93,94,102] Fuchs' use of the Oldenburg Sentence Test[60] in applause type noise[102] shows another adaptation which balances ecological validity with controlled stimuli. Usage of noises potentially encountered in the listening environment, like the café noise used by Walton,[101] or in the content, like mixtures of music and ambience, may also improve the ecological validity of speech in noise tests. However, even with relevant noises such tests remain limited by their audio-only modality. The Illustrated Sentence Test, with static visual context, goes some way to amending this.[68] The GRID corpus[67] improves on this further with matched vision of the speaker, allowing for the effect of visual complementary intelligibility cues, such as lip-reading, to be accounted for. However, as the speaker is face on to the camera throughout, it would only generalise well to content like news broadcasts and not to most dramatic content.
Objective metrics cannot account for complementary intelligibility cues within television content, however they can be effective for evaluating the effect of background sounds on signal-dependent intelligibility.[43,104] Furthermore their ecological validity is heavily dependent on their efficacy across different broadcast noise types. Studies have shown that whilst state-of-the-art metrics make good estimates of intelligibility in single masker types, across-noise predictions are generally poor.[69] As such, without evaluation in different broadcast noise types, the ecological validity of most objective measures remains unknown and would require systematic reviews of existing methods in broadcast type noise. The Binaural Distortion Weighted Glimpse Proportion is the only metric to be evaluated for broadcast use and was evaluated in speech shaped noise, music with and without vocals and babble noise.[104]

8.2 Representative of Hearing Loss Characteristics

Speech in noise tests inherently account for the effect of suprathreshold hearing loss and tests of this type have been used to good effect in television research.[43,93,94,102] Ward's study of the effects of non-speech sounds on intelligibility for hard of hearing listeners showed that this type of controlled stimuli can be leveraged to better explain the relationships between hearing impairment and complementary intelligibility cue usage[50] than subjective experiments with similar aims.[7] Whilst some research has used a single speech to background ratio for all listeners for each stimuli type,[102] the approaches used by Ward[50] and in the DICTION project[93] adjusted the signal to background ratio for each hard of hearing participant. This approach limits the generalisability of the results. Speech in noise tests designed to yield an SRT, like the Hearing in Noise Test,[58] or modified versions of single speech to noise ratio tests, such as the multiple signal to noise ratio version of the R-SPIN, would likely accommodate for various hearing abilities more effectively.[107]

Whilst objective intelligibility metrics can be quite easily modified to accommodate reduction in audiometric thresholds, the degree to which they model suprathreshold loss varies.[73,88–91] The use of widened filters to represent loss of frequency selectivity begins to accommodate for basic suprathreshold factors and has the advantage of simplicity. The greater the number of internal parameters the models contain however the greater the difficulty in calibrating the model, and the greater the potential for the inaccuracies in the model to be larger than the difference between approaches under evaluation.[23] The Gaussian noise model used in FADE[81] to accommodate for suprathreshold loss appears to effectively balance modelling of suprathreshold characteristics with ease of calibration and minimal internal parameters.

Validation of methods with hard of hearing listeners is also key. Approaches like Fontan's, where validation is undertaken with normal hearing listeners and stimuli modified to mimic hearing loss,[82] do not necessarily represent hard of hearing listener performance. They are limited in two ways: the modelling of hearing loss used to adjust the stimuli is unlikely to model all suprathreshold factors and, more importantly, any mechanisms listeners have developed to compensate for their hearing loss, such as increased used of complementary intelligibility cues, will not be present in normal hearing populations.

8.3 Repeatable

Repeatability here refers to evaluation methods which have either been validated for test-retest reliability or have taken steps within their methodology to minimise, or characterise, intra-personal variation in responses. Objective intelligibility metrics are, by their nature, repeatable as they utilise static models of intelligibility which is one of their distinct advantages. Most speech in noise tests have been validated with the target population, also giving them good repeatability. However, this repeatability may be limited, if significant alterations are made to the stimuli or test methodology.

Personalisation style tests do have some repeatability limitations, partly as the methodology is new and has not been extensively validated. They also limit their generalisability by requesting a user personalise their levels, without validating their choices against a reference or default, to confirm that the selection is perceived as an improvement. The type of experimental approach used by Tang to validate an objective intelligibility metric for broadcast and by the Adjustment/Satisfaction Test[98] improve the repeatability of these tests through their addition of a validation stage.

9 Final Remarks

This paper has reviewed how ecologically valid, repeatable and representative of the characteristics of hearing loss various methods for evaluating intelligibility are. It has highlighted the usefulness of controlled objective tests for revealing and modelling patterns between hearing loss and intelligibility of television sound. At a number of points this work has alluded to, though not directly addressed, personal preference. Preference must be considered in accessibility research as what is objectively better may not always be what is preferential to the target users.[50] Taking an example from the established field of subtitle research: it has been recommended that subtitles should not exceed objective measured maximum reading speeds however target users prefer that subtitles are time-aligned, even if they exceed this speed.[108,109] Such preferences are only gained from evaluation methods involving the target users. So whilst objective metrics provide useful, repeatable and often representative evaluation methods, only evaluation with end-users can ensure ecological validity.

10 Acknowledgements

Lauren Ward is funded by the General Sir John Monash Foundation. The contributions of Yan Tang, Cammi Motley and Phillipa Demonte to this work are gratefully acknowledged.

11 REFERENCES

1. Action on Hearing Loss. Hearing Matters Report, 2015. URL: https://www.actiononhearingloss.org.uk/how-we-help/information-and-resources/publications/research-reports/hearing-matters-report/, [Accessed: 2/10/2017].
2. Royal National Institute for Deaf People. Annual survey report 2008, 2008.
3. D. Cohen. Sound matters, BBC College of Production. March 2011. URL: http://www.bbc.co.uk/academy/production/article/art20130702112136134, [Accessed: 16/6/2016].
4. M. Armstrong. BBC White Paper WHP 324: From Clean Audio to Object Based Broadcasting, Oct, 2016. URL: http://www.bbc.co.uk/rd/publications/whitepaper324, [Accessed: 28/09/2017].
5. M. Armstrong and M. Crabb. Exploring ways of meeting a wider range of access needs through object-based media - workshop. In Conf. on Accessibility in Film, Television and Interactive Media, York, UK, Oct. 2017.
6. L. Ward. Accessible broadcast audio personalisation for hard of hearing listeners. In TVX, pages 105–108, Hilversum, Netherlands, June 2017. ACM.
7. B.G. Shirley, M. Meadows, F. Malak, J.S. Woodcock, and A. Tidball. Personalized object-based audio for hearing impaired TV viewers. J.Audio.Eng.Soc, 65(4):293–303, 2017.
8. Access Economics. Listen hear! the economic impact and cost of hearing loss in Australia. Report for The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and Victorian Deaf Society, 2006. URL: https://audiology.asn.au/public/1/files/Publications/ListenHearFinal.pdf, [Acessed: 2/10/2017].
9. Y. Agrawal, E.A. Platz, and J.K. Niparko. Prevalence of hearing loss and differences by demographic characteristics among us adults: data from the national health and nutrition examination survey, 1999-2004. Arch.Intern.Med, 168(14):1522–1530, 2008.
10. Senate Community Affairs References Committee et al. Hear us: inquiry into hearing health in australia. Available at [Verified 17 November 2015], 2010.
11. H. Aazh, D. Prasher, K. Nanchahal, and B.C.J. Moore. 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):152–161, 2015.
12. Office for National Statistics. National population projections: 2014-based statistical bulletin, Oct, 2015. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2015-10-29#older-people, [Accessed: 2/10/2017].
13. G.A. Gates and J.H. Mills. Presbycusis. Lancet, 366(9491):1111–1120, 2005.
14. K.T. Palmer, M.J. Griffin, H.E. Syddall, A. Davis, B. Pannett, and D. Coggon. Occupational exposure to noise and the attributable burden of hearing difficulties in Great Britain. Occup.Environ.Med, 59(9):634–639, 2002.
15. S. Sadhra, C.A. Jackson, T. Ryder, and M.J. Brown. Noise exposure and hearing loss among student employees working in university entertainment venues. Ann.Occup.Hyg., 46(5):455–463, 2002.
16. M. Maassen, W. Babisch, K.D. Bachmann, H. Ising, G. Lehnert, P. Plath, P. Plinkert, E. Rebentisch, G. Schuschke, M. Spreng, et al. Ear damage caused by leisure noise. Noise.Health, 4(13):1, 2001.
17. R.J.H. Smith, A.E. Shearer, M.S. Hildebrand, and G. Van Camp. Deafness and hereditary hearing loss overview. Gene.Rev., URL: https://www.ncbi.nlm.nih.gov/books/NBK1434/, [Last Updated: 27/7/2017], 2014.
18. P.M. Rabinowitz. Noise-induced hearing loss. Am.Fam.Physician, 61(9):2759–2760, 2000.
19. British Society of Audiology. Recommended procedure: pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels. Sept. 2011. URL: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf, [Accessed: 2/10/2017].
20. A.W. Bronkhorst and R. Plomp. Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. J.Acoust.Soc.Am., 92(6):3132–3139, 1992.

21. M.K. Pichora-Fuller, B.A. Schneider, and M. Daneman. How young and old adults listen to and remember speech in noise. J.Acoust.Soc.Am., 97(1):593–608, 1995.
22. A.J. Vermiglio, S.D. Soli, D.J. Freed, and L.M. Fisher. 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):779–788, 2012.
23. M. Huckvale and G. Hilkhuysen. On the predictability of the intelligibility of speech to hearing impaired listeners. In 1st Int. CHAT Workshop, URL: `http://spandh.dcs.shef.ac.uk/chat2017/papers/CHAT_2017_huckvale.pdf`, Stockholm, Sweden, Aug. 2017.
24. R. Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J.Acoust.Soc.Am, 63(2):533–549, 1978.
25. D. Ruggles and B. Shinn-Cunningham. Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. J.Assoc.Res.Otolaryngol., 12(3):395–405, 2011.
26. C. Füllgrabe, B.C.J. Moore, and M.A. Stone. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Front.Aging.Neurosci., 6, 2014.
27. H.M. Bharadwaj, S. Masud, G. Mehraei, S. Verhulst, and B.G. Shinn-Cunningham. Individual differences reveal correlates of hidden hearing deficits. J.Neurosci., 35(5):2161–2172, 2015.
28. B.C.J. Moore and B.R. Glasberg. Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. J.Acoust.Soc.Am., 94(4):2050–2062, 1993.
29. C.W. Newman, J.A. Wharton, B.G. Shivapuja, and G.P. Jacobson. Relationships among psychoacoustic judgments, speech understanding ability and self-perceived handicap in tinnitus subjects. Audiology, 33(1):47–60, 1994.
30. E. Villchur. Simulation of the effect of recruitment on loudness relationships in speech. J.Acoust.Soc.Am., 56(5):1601–1611, 1974.
31. K. Hopkins and B.C.J. Moore. The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. J.Acoust.Soc.Am., 125(1):442–446, 2009.
32. R. Badri, J.H. Siegel, and B.A. Wright. Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms a. J.Acoust.Soc.Am., 129(2):852–863, 2011.
33. A. Wingfield, P.A. Tun, and S.L. McCoy. Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. Curr.Dir.Psychol.Sci., 14(3):144–148, 2005.
34. C.J. Plack, D. Barker, and G. Prendergast. Perceptual consequences of “hidden” hearing loss. Trends.Hear., 18:2331216514550621, 2014.
35. R. Schaette and D. McAlpine. Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. J.Neurosci., 31(38):13452–13457, 2011.
36. Lionel Fontan, Julien Tardieu, Pascal Gaillard, Virginie Woisard, and Robert Ruiz. Relationship between speech intelligibility and speech comprehension in babble noise. Journal of Speech, Language, and Hearing Research, 58(3):977–986, 2015.
37. IEC60268-16: Objective rating of speech intelligibility by speech tranmission index. Standard, International Electrotechnical Commission, 2011.
38. N. Miller. Measuring up to speech intelligibility. Int.J.Lang.Commun.Disord., 48(6):601–612, 2013.
39. A.A. Zekveld, M. Rudner, I.S. Johnsrude, D.J. Heslenfeld, and J. Rönnberg. Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. Brain.Lang., 122(2):103–113, 2012.
40. B. Lindblom. On the communication process: Speaker-listener interaction and the development of speech. Augment.Altern.Commun., 6(4):220–230, 1990.
41. J.K. Bizley and Y.E. Cohen. The what, where and how of auditory-object perception. Nature.Rev.Neurosci., 14(10):693–707, 2013.
42. I.J. Hirsh. The relation between localization and intelligibility. J.Acoust.Soc.Am., 22(2):196–200, 1950.
43. L. Ward, B.G. Shirley, Y. Tang, and W.J. Davies. The effect of situation-specific acoustic cues on speech intelligibility in noise. In Proc. 18th Interspeech, pages 2958–2962, Stockholm, Sweden, Aug. 2017. ISCA.
44. R.C. Bilger. Speech recognition test development, In: E. Elkins ed. Speech recognition by the hearing impaired, volume 14, pages 2–15. 1984.

45. Z. Yang, J. Chen, Q. Huang, X. Wu, Y. Wu, B.A. Schneider, and L. Li. The effect of voice cuing on releasing chinese speech from informational masking. Speech.Commun., 49(12):892–904, 2007.
46. R.L. Freyman, U. Balakrishnan, and K.S. Helfer. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J.Acoust.Soc.Am, 115(5):2246–2256, 2004.
47. B. Spehar, S. Goebel, and N. Tye-Murray. Effects of context type on lipreading and listening performance and implications for sentence processing. J.Speech.Lang.Hear.Res., 58(3):1093–1102, 2015.
48. M. Aguert, V. Laval, L. Le Bigot, and J. Bernicot. Understanding expressive speech acts: the role of prosody and situational context in french-speaking 5-to 9-year-olds. J.Speech.Lang.Hear.Res., 53(6):1629–1641, 2010.
49. T.L. Arbogast, C.R. Mason, and G. Kidd Jr. The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. J.Acoust.Soc.Am, 117(4):2169–2180, 2005.
50. L. Ward and B.G. Shirley. Television dialogue; balancing audibility, attention and accessibility. In Conf. on Accessibility in Film, Television and Interactive Media, York, UK, Oct. 2017.
51. H. Fullerton. BBC drama SS-GB criticised for "mumbling" and bad sound quality in first episode, 26 Feb 2017. URL: http://www.radiotimes.com/news/2017-02-26/bbc-drama-ss-gb-criticised-for-mumbling-and-bad-sound-quality-in-first-episode.
52. J. Plunkett. Heard this before? BBC chief speaks out over happy valley mumbling, 8 Apr 2016. URL: https://www.theguardian.com/media/2016/apr/08/bbc-happy-valley-mumbling-jamaica-inn-sarah-lancashire.
53. CD Mathers. A study of sound balances for the hard of hearing. NASA STI/Recon Technical Report N, 91, 1991.
54. M. Armstrong. BBC White Paper WHP 190: Audio Processing and Speech Intelligibility: a literature review, Jan, 2011. URL: http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP190.pdf.
55. P. Mapp. Intelligibility of Cinema & TV Sound Dialogue. In Proc. 141st Audio.Eng.Soc.Con., Sep 2016.
56. J. Herre, H. Purnhagen, J. Koppens, O. Hellmuth, J. Engdegård, J. Hilper, L. Villemoes, L. Terentiv, C. Falch, A. Hölzer, et al. MPEG spatial audio object coding - the ISO/MPEG standard for efficient coding of interactive audio scenes. J.Audio.Eng.Soc., 60(9):655–673, 2012.
57. J. Popp, M. Neuendorf, H. Fuchs, C. Forster, and A. Heuberger. Recent advances in broadcast audio coding. In Proc. 9th Int. Symp. Broadband Multimedia Systems and Broadcasting, pages 1–5, London, 2013. IEEE.
58. M. Nilsson, S.D. Soli, and J.A. Sullivan. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J.Acoust.Soc.Am., 95(2):1085–1099, 1994.
59. S. Cameron and H. Dillon. The listening in spatialized noise-sentences test (lisn-s): Comparison to the prototype lisn and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. J.Am.Acad.Audiol., 19(5):377–391, 2008.
60. K. Wagener, V. Kühnel, and B. Kollmeier. Development and evaluation of a german sentence test i: Design of the oldenburg sentence test. Zeitschrift Fur Audiologie, 38:4–15, 1999.
61. Etymotic. BKB-SIN^{TM} Speech-in-Noise test Version 1.03.
62. D.N. Kalikow, K.N. Stevens, and L.L. Elliott. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J.Acoust.Soc.Am., 61(5):1337–1351, 1977.
63. J. Bench, Å. Kowal, and J. Bamford. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br.J.Audiol., 13(3):108–112, 1979.
64. HearCom. Matrix sentence test, 2010. RL: http://hearcom.eu/prof/DiagnosingHearingLoss/AuditoryProfile/SpatialHearing.html, [Accessed: 02/10/2017].
65. R.A. McArdle, R.H. Wilson, and C.A. Burks. Speech recognition in multitalker babble using digits, words, and sentences. J.Am.Acad.Audiol., 16(9):726–739, 2005.
66. K.R. Duncan and N.L. Aarts. A comparison of the HINT and Quick Sin Tests. J.Speech Lang.Pathol.Audiol., 30(2):86, 2006.
67. M. Cooke, J. Barker, S. Cunningham, and X. Shao. An audio-visual corpus for speech perception and automatic speech recognition. J.Acoust.Soc.Am., 120(5):2421–2424, 2006.
68. N. Tye-Murray, S. Hale, B. Spehar, J. Myerson, and M.S. Sommers. Lipreading in school-age children: the roles of age, hearing status, and cognitive ability. Journal of Speech, Language, and Hearing Research, 57(2):556–565, 2014.
69. Y. Tang, M. Cooke, and C. Valentini-Botinhao. Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech. Comp.Speech.Lang., 35:73–92, 2016.

70. H. Fletcher. An empirical theory of telephone quality. AT&T Internal Memorandum, 101(6), 1921.
71. ANSI S3.5. ANSI S3.5-1997 Methods for the calculation of the Speech Intelligibility Index, 1997.
72. M. Cooke. A glimpsing model of speech perception in noise. Journal of the Acoustical Society of America, 119(3):1562–1573, 2006.
73. I. Holube and B. Kollmeier. Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. J.Acoust.Soc.Am, 100(3):1703–1716, 1996.
74. G. Carter, C. Knapp, and A. Nuttall. Estimation of the magnitude-squared coherence function via overlapped fast fourier transform processing. IEEE.Trans.Audio.Electroacoust., 21(4):337–344, 1973.
75. J.M. Kates and K.H. Arehart. Coherence and the speech intelligibility index. J.Soc.Acoust.Am., 117(4):2224–2237, 2005.
76. F. Chen, O. Hazrati, and P.C. Loizou. Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure. Biomed.Signal.Process.Control., 8(3):311–314, 2013.
77. C.H. Taal, R.C. Hendriks, R. Heusdens, and J. Jensen. A short-time objective intelligibility measure for time-frequency weighted noisy speech. In ICASSP, pages 4214–4217. IEEE, 2010.
78. H.J.M. Steeneken and T. Houtgast. A physical method for measuring speech-transmission quality. J.Acoust.Soc.Am., 67(1):318–326, 1980.
79. T.H. Falk, C. Zheng, and W. Chan. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. IEEE.Trans.Audio.Speech.Lang.Process, 18(7):1766–1774, 2010.
80. C. Spille and B.T. Meyer. Listening in the dips: Comparing relevant features for speech recognition in humans and machines. Proc. 18th Interspeech, pages 2968–2972, 2017.
81. B. Kollmeier, M.R. Schädler, A. Warzybok, B.T. Meyer, and T. Brand. Sentence recognition prediction for hearing-impaired listeners in stationary and fluctuation noise with fade: Empowering the attenuation and distortion concept by plomp with a quantitative processing model. Trends.Hear., 20:2331216516655795, 2016.
82. L. Fontan, I. Ferrané, J. Farinas, J. Pinquier, J. Tardieu, C. Magnen, P. Gaillard, X. Aumont, and C. Füllgrabe. Automatic speech recognition predicts speech intelligibility and comprehension for listeners with simulated age-related hearing loss. J.Speech.Lang.Hear.Res., pages 1–12, 2017.
83. M. Karbasi, A.H. Abdelaziz, H. Meutzner, and D. Kolossa. Blind non-intrusive speech intelligibility prediction using twin-hmms. In Proc. 17th Interspeech, pages 625–629, 2016.
84. M.R. Schädler, A. Warzybok, S.D. Ewert, and B. Kollmeier. A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. J.Acoust.Soc.Am., 139(5):2708–2722, 2016.
85. L.E. Humes, D.D. Dirks, T.S. Bell, C. Ahlstbom, and G.E. Kincaid. Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners. J.Speech.Lang.Hear.Res., 29(4):447–462, 1986.
86. K.S. Rhebergen, N.J. Versfeld, and W.A. Dreschler. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. J.Soc.Acoust.Am., 120(6):3988–3997, 2006.
87. R.M. Meyer and T. Brand. Comparison of different short-term speech intelligibility index procedures in fluctuating noise for listeners with normal and impaired hearing. Acta.Acust.united.Ac., 99(3):442–456, 2013.
88. J.M. Kates and K.H. Arehart. The hearing-aid speech perception index (HASPI). Speech.Commun., 65:75–93, 2014.
89. J.M. Kates and K.H. Arehart. The hearing-aid speech quality index (HASQI) version 2. J.Audio.Eng.Soc, 62(3):99–117, 2014.
90. J. Ma, Y. Hu, and P.C. Loizou. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J.Acoust.Soc.Am., 125(5):3387–3405, 2009.
91. T.H. Falk, V. Parsa, J.F. Santos, K. Arehart, O. Hazrati, R. Huber, J.M. Kates, and S. Scollie. Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools. IEEE.Signal.Process.Mag., 32(2):114–124, 2015.
92. J.M. Kates. An auditory model for intelligibility and quality predictions. In Proc. Meetings on Acoustics ICA2013, volume 19, page 050184. ASA, 2013.
93. A.R. Carmichael. Evaluating digital “on-line" background noise suppression: Clarifying television dialogue for older, hard-of-hearing viewers. Journal of Neuropsychology Rehabilitation, 14(1-2):241–249, 2004.
94. B.G. Shirley, P. Kendrick, and C. Churchill. The effect of stereo crosstalk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source. J.Audio.Eng.Soc, 55(10):852–863, 2007.

95. B.G. Shirley and P. Kendrick. ITC Clean Audio Project. In Proc. 116th Audio.Eng.Soc.Con. Audio Engineering Society, 2004.
96. C. Uhle, O. Hellmuth, and J. Weigel. Speech enhancement of movie sound. In Proc. 125th Audio.Eng.Soc.Con. AES, 2008.
97. H. Müsch. Aging and sound perception: Desirable characteristics of entertainment audio for the elderly. In 125^{th} Con. Audio.Eng.Soc., 2008.
98. M. Torcoli, J. Herre, J. Paulus, C. Uhle, H. Fuchs, and O. Hellmuth. The Adjustment/Satisfaction Test (A/ST) for the Subjective Evaluation of Dialogue Enhancement. In Proc. 143rd Audio.Eng.Soc.Con., New York, USA, Oct 2017.
99. H. Fuchs, S. Tuff, and C. Bustad. Dialogue enhancement—technology and experiments. EBU Technical review, 2:1, 2012.
100. M. Mann, A.W.P. Churnside, A. Bonney, and F. Melchior. Object-based audio applied to football broadcasts. In Proc. Int. Workshop on Immersive Media Experiences, pages 13–16. ACM, 2013.
101. T. Walton, M. Evans, D. Kirk, and F. Melchior. Does environmental noise influence preference of background-foreground audio balance? In Proc. 141st Audio.Eng.Soc.Con., Los Angeles, U.S.A., 2016. AES.
102. H. Fuchs and D. Oetting. Advanced clean audio solution: Dialogue enhancement. SMPTE Motion Imaging Journal, 123(5):23–27, 2014.
103. Y. Tang, M. Cooke, B.M. Fazenda, and T.J. Cox. A metric for predicting binaural speech intelligibility in stationary noise and competing speech maskers a. J.Acoust.Soc.Am., 140(3):1858–1870, 2016.
104. Y. Tang, B.M. Fazenda, and T.J. Cox. Automatic speech-to-background ratio selection for maintaining speech intelligibility in broadcasts using an objective intelligibility metric. J. Audio Eng. Soc., 2017. In Revision.
105. M. Torcoli and C. Uhle. On the effect of artificial distortions on objective performance measures for dialog enhancement. In Proc. 141st Audio.Eng.Soc.Con. AES, 2016.
106. T. Walton, M. Evans, F. Melchior, and D. Kirk. Combining preference ratings with sensory profiling for the comparison of audio reproduction systems. In Proc. 142nd Audio.Eng.Soc.Con. AES, 2017.
107. R.H. Wilson, R. McArdle, K.L. Watts, and S.L. Smith. The revised speech perception in noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm. J.Am.Acad.Audiol., 23(8):590–605, 2012.
108. OFCOM. Code on television access services, Jan 2017. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/97040/Access-service-code-Jan-2017.pdf, [Accessed: 6/10/2017].
109. Diana Sanchez. Reading speed, 8th Aug 2014. URL: http://www.ericsson.com/broadcastandmedia/access-services/reading-speed/?platform=hootsuite, [Accessed: 6/10/2017].