DEVELOPMENT AND PRELIMINARY RESULTS OF THE UNIVERSITY OF SALFORD MEDIA ACCESSIBILITY AND HEARING IMPAIRMENT DATABASE (U-SAID)

Lauren Ward[1] L.Ward7@edu.salford.ac.uk
Ben Shirley[1]
Bill Davies[1]
[1] Acoustics Research Centre, University of Salford, Salford, M5 4WT, UK

ABSTRACT

Recent technological advances in object-based broadcasting present the opportunity to improve broadcast accessibility, particularly for the 11 million people in the UK with hearing impairment. Taking advantage of this opportunity is important given the key social role television plays, especially for older people.

To best exploit this opportunity, a greater understanding is required of the relationship between different types of hearing impairment and the specific barriers to broadcast access they present. To begin deconstructing this relationship the University of Salford has assembled a database, which this paper describes. This database compromises of comprehensive data about individuals' hearing loss as well as quantitative and qualitative data about their experience with television speech.

This paper describes the databases' development and methodology, as well as preliminary results from the first 13 data-sets. Potential applications and plans for its open source released will also be outlined.

1 Introduction

Hearing loss is estimated to affect one in six people in the United Kingdom (UK), Europe and North America.[1–3] The most common cause of this hearing loss is age-related loss.[1,4] 2017 audience statistics indicate that this same demographic which is most at risk for hearing loss, watch the most amount of television: those over 50 years old in the United States of America and those over 55 in the UK watch more television on average than any other age demographic in their respective countries.[5,6] Furthermore as increasing amounts of communication and social interaction is conducted using audiovisual media ensuring accessibility for all, including those with hearing loss, is essential.

People with hearing loss face numerous barriers to fully engaging with television content. Television, particularly dramatic content, presents complex soundscapes to the viewer. Understanding clear speech in a television soundtrack presents a challenge for hard of hearing viewers, which may be exacerbated by unfamiliar accents, poor dialogue clarity and poor television sound reproduction among other factors.[7–10] The exact relationship between these challenges, and different types and severities of hearing loss is poorly understood. Furthermore, given recent advances in broadcast technology which facilitate personalisation of television content, a better understanding of the needs of hearing impaired individuals is required to ensure any implemented personalisation strategies improve access.

A limited number of databases exist that document the relationship between different clinical measures of hearing loss and speech intelligibility in everyday scenarios.[11,12] Only one open-access data set contains survey data which links self-reported hearing loss with experience of television speech.[10] No open-access databases exists linking clinical measures of hearing loss with individuals' experience of broadcast media. To address this deficit, the University of Salford media Accessibility and hearing Impairment Database (U-SAID) has been developed.

2 About the Database

This paper describes the development of U-SAID. Five types of quantitative data about each individuals hearing loss were collected: self-reported level of hearing loss, sensitivity to temporal fine structure,[13,14] pure tone audiometric thresholds,[15] speech reception threshold (using QuickSIN[16]) and the Speech, Spatial and Qualities of Hearing survey (SSQ49[17]). Tests were selected based on efficacy and rapid administration. Three additional data types about each individual's experience of broadcast media and new technologies were also collected: quantitative experience of TV questions (TV7), qualitative survey questions on broadcast experiences and R-SPIN[18,19] with added sound effects.[20,21]

This paper outlines each data type, the rationale for its inclusion in the database and its collection methodology. Whilst data collection is ongoing, preliminary results for the first 13 subjects are also included. Data was collected primarily in person, with all participants visiting the University of Salford campus. Some self-report data was collected from participants online prior to visiting the University.

2.1 Participant Demographics

Participants were recruited through professional organisations, community groups (such as University of the 3rd Age) and University publicity. Participants were native English speakers and either: identified as having mild to moderate hearing loss in their better hearing ear or, were over the age of 50.

13 participants took part in the first round of data collection – 11 males and 2 females. Participants ranged in age from 53 to 95, with a median age of 63. Only two participants identified as being musicians (amateur).

3 Characterisation of Hearing Impairment

A variety of clinical and research tests were selected to cover the widest range of characteristics possible. Two validated and clinically used measures were selected: pure tone audiometry and the Quick Speech in Noise (QuickSIN)[16] test. Whilst limited in its diagnostic and explanatory capabilities[22] the pure tone audiogram remains widely utilised, both clinically and in research, and it is for this reason that it is included in the database. Increasingly speech perception in noise tests are used clinically to provide a more complete measure of an individual's ability to understand speech in everyday scenarios. In addition to these clinical measures, a measure of temporal fine structure was included as many recent studies suggest a link between higher sensitivity to temporal fine structure and speech in noise performance.[23–26] Individuals were also asked to report their level of hearing loss and other salient details about their use of assistive hearing devices.

3.1 Pure tone audiogram

The pure tone audiogram determines the lowest sound pressure an individual can perceive, termed the threshold, at a given set of frequencies. This is conducted over headphones and performed individually for each ear. These thresholds are often averaged over a number of frequencies to yield a pure tone average (PTA), which is commonly used to define the severity of an individual's hearing loss.[15]

Methodology Audiograms were performed according to the BSA Recommended Procedure[15] utilising either a Kamplex r27a or Kamplex KLD21 Diagnostic Audiometer. Audiometric thresholds for the frequencies 0.25Hz, 0.5Hz, 1kHz, 2kHz, 4kHz, and 8kHz were obtained. Thresholds were tested up to 110dB – if a subject had no response at this level, that frequency was reported as a 'no response' (*NR*).

Preliminary Results Two pure tone averages were calculated, PTA over which was averaged over the frequencies 0.25Hz, 0.5Hz, 1kHz, 2kHz and 4kHz and a high frequency version (HFPTA) over 2kHz, 4kHz and 8kHz. Using these averages for each individual, and the definitions of hearing loss

severity outlined by the British Society of Audiology,[15] counts for the number of participants with each severity of hearing loss can be seen in Table 1. Whether this loss was symmetric or asymmetric is also noted.

Table 1: Counts for each hearing loss severity based on low and high frequency pure tone averages, with the symmetry of loss noted.

Hearing Loss Severity	PTA (counts)	HFPTA (counts)
Normal	1	–
Asymmetric Mild	1	–
Asymmetric Moderate	–	1
Asymmetric Severe	1	1
Symmetric Mild	3	2
Symmetric Moderate	7	5
Symmetric Severe	–	4

3.2 Adaptive Frequency Test of Temporal Fine Structure

Sensitivity to temporal fine structure (TFS) refers to the ability to perceive the rapid oscillations close to the centre frequency of a complex sound, such as speech. It has been shown to play a significant role in pitch perception[27] as well as being hypothesized to play a role in speech perception in competing noise and in stream segregation.[28] It has also been shown that hearing impaired listeners have a reduced ability to utilise TFS information.[29] A measure of TFS was included in order to explore the relationship between TFS and television speech understanding.

There are a number of different ways in which sensitivity to TFS can be estimated including interaural phase descrimination (IPD), pitch perception and speech recognition.[25] The tool used in this study was the TFS-AF test developed by Füllgrabe and colleagues[13,14] which presents the low frequency IPD task developed by Hopkins and Moore[30] in an adaptive paradigm. This test determines the highest frequency at which a fixed interaural phase difference of 180° can be perceived as different from a phase difference of 0°. This was selected as it is designed to address problems encountered with the administration of other similar tests whereby some listeners, particularly older listeners, are unable to complete the tasks and results are not easily comparable across participants.[14,30,31] The method uses 2 option forced choice and an adaptive procedure with a 2-up, 1-down stepping rule which terminates after 8 reversals and takes the geometric mean of the last 6 results to estimate the 71% point on the psychometric curve. An example result can be seen in Appendix 1.

Methodology The settings utilised can be seen in Appendix 1. The level was set at 40dB above participants' hearing loss (except for one participant which was 30dB above). Level adjustment was applied for each participant based on the results of their pure tone audiometric thresholds at 0.25kHz, 0.5kHz, 1kHz and 2kHz, to ensure that all frequencies were heard at the same loudness by all participants. The starting frequency was set to 200Hz.

Preliminary Results The frequency at which participants could no longer discriminate interaural phase differences ranged from 41.8Hz (SD = 0.17) to 1687.9Hz (SD = 0.11), with a median value of 1042Hz.

3.3 QuickSIN

QuickSIN is a modified version of the Speech in Noise test developed by Killion et. al.[16] It utilises the IEEE sentences spoken by a female talker presented in four-talker babble noise. Each list contains six

sentences with five key words per sentence. The sentences are presented at pre-recorded signal-to-noise ratios (SNR) decreasing in 5-dB steps from 25 (very easy) to 0 (extremely difficult). These SNRs are designed to encompass the point of 50% recognition for normal to severely impaired people. From performance at each level SNR loss can be calculated, with a normative value of 2 dB SNR loss for normal hearing individuals.

Methodology The QuickSIN test was installed on the Kamplex r27a Diagnostic Audiometer and presented to both ears at the same level over headphones. One practise list (Practise List A Track 21) was presented first and two of the main lists (Track 2, List 4 and Track 1 List 3). The order of the two main lists was pseudo-randomised between participants. It is recommended that for improved accuracy two or more lists should be averaged[32] and for this reason participants completed two lists and the reported results are an average of the results of these lists.

A nominal presentation level of 65dB was used and during the practise list participants could request the level be increased by up to 10dB. Participants were asked to write down the whole sentence and the progression of the sentences was paused by the researcher to allow each participant adequate time to write. Participants were also given the option of having the researcher scribe for them if they preferred.

Preliminary Results SNR loss was calculated using the formula:

$$\text{SNR Loss} = 25.5\text{dB} - (\text{Total Correct})$$

the derivation of which can be found in the QuickSIN Manual[32] and is based on the Tillman-Olsen recommended method for obtaining spondee thresholds.[33]

SNR loss ranged from 1dB SNR to 22.5dB SNR, with a median value of 7.5dB SNR. The QuickSIN manual provides an interpretation of the resulting SNR loss in terms of severity. The number of individuals at each severity level is given in Table 2.

Table 2: Severity levels of SNR loss and number of participants at each level

SNR Loss	Count	SNR Loss Severity
0-3dB	3	Near Normal
3-7dB	2	Mild
7-15dB	6	Moderate
>15dB	2	Severe

3.4 Speech, Spatial and Qualities of Speech Survey and Self-reported Hearing Loss

The Speech, Spatial and Qualities of Hearing scale (SSQ)is a survey designed to capture information about an individual's perceptions of how their hearing (dis)ability impacts upon them in the complex listening situations encountered in everyday life.[17] It has been widely used as both a clinical assessment of the benefit of different hearing interventions (including but not limited to cochlear implants, bilateral hearing aids and bone-anchored hearing aids[34]) as well as in research investigations of younger and older listeners.[24]

The original, full form of the survey has 50 items, though one question applies only to hearing aid users and is often omitted.[24,35] The 49 item version is abbreviated here as SSQ49. SSQ49 consists of three categories of questions: Speech Hearing (14 items e.g. *"Can you have a conversation in the presence of someone whose voice is the same pitch as that of the person you're talking to?"*), Spatial Hearing (17 items e.g. *"Do you have the impression of sounds being exactly where you would expect them to be?"*) and Qualities of Hearing (18 items e.g. *"Do you find it easy to recognize different*

people you know by the sound of each one's voice?"). Each item asks the respondent to rate their ability to perform an auditory based 'activity' on an 11 point Likert scale from 0 (= complete disability) to 10 (= complete ability). It can be self-administered or administered by a clinician or researcher. The test-retest reliability of the SSQ49 when administered by a clinician is 0.83 and reduces to 0.65 when self-administered.[36] Internal reliability, using Cronbach's alpha (for which high values, $\geq$0.9, indicate excellent internal consistency), has been shown to be high (0.96-0.97[36] and 0.96[37]), when administered by interview. Although lower when self-administered, 0.88-0.93,[36] this value is still defined as 'good' (values $\geq$0.8).

In addition to the SSQ49, participants were asked whether they had hearing aids fitted and their hearing aid usage patterns when they attended the University.

Methodology The SSQ49 was delivered online to the participants, allowing them to complete it in their own time beforehand. Whilst the test-retest reliability and Cronbach's alpha of the SSQ49 is lower when self-administered, this allowed more tests to be conducted in person during test sessions. Included in the online delivery were the survey questions from 4.2 on broadcast media experiences.

Preliminary Results 8 participants had hearing aids fitted in at least one ear (5 bilateral, 2 unilateral). Of those, half identified as using these devices regularly. Participants have had assistive hearing devices fitted for periods ranging from 2 to 16 years and utilised a variety of aids. Five participants reported suffering from Tinnitus.

The results from the SSQ49 can be seen in Figure 1. All participants rated their ability to understand speech lower than their general quality of hearing and their localisation ability. On average, qualities of hearing and spatial hearing were rated similarly and much higher than average speech understanding.

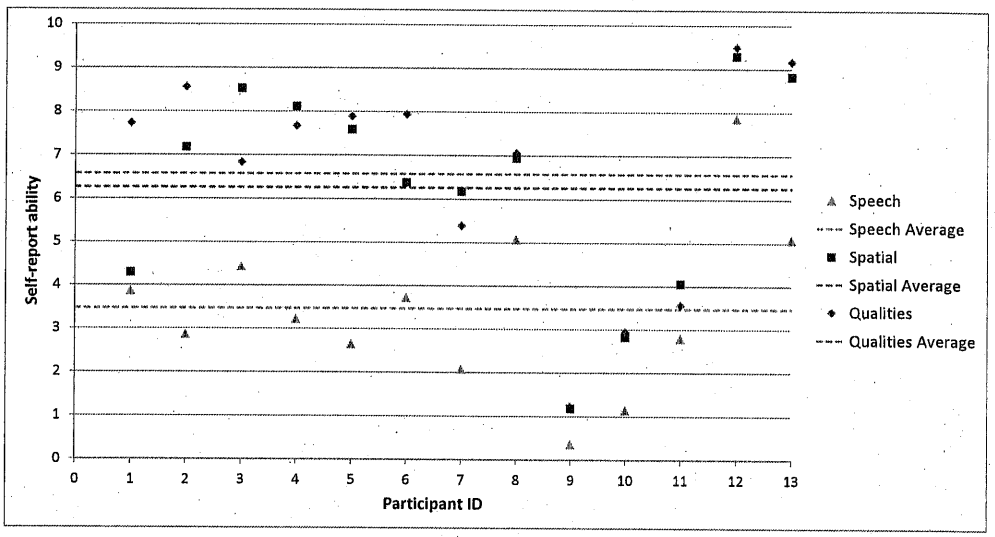

Figure 1: Individual Participant results and averages for the Speech, Spatial and Qualities of Hearing sections of the SSQ49

4 Media needs

To complement the data collected about each individual's hearing loss, a number of measures of television speech understanding were gathered: closed and open-ended qualitative survey items, quantitative self-report measures and quantitative laboratory measures of speech in television style maskers. Qualitative survey items were utilised to capture the inherent subjective nature of quality of television

experience. Quantitative self-report measures, in the same style as the SSQ49, were utilised to capture data about difficulty of speech understanding in scenarios specific to television. This allowed for direct comparison with the SSQ49 measures, which cover a broader range of listening scenarios.

Television speech understanding presents a unique speech in noise challenge. The competing sounds are not only varying and covering a wide array of sounds,[10] but many of these sounds are not strictly noise and play a meaningful, narrative role in conveying the story.[38] No dedicated objective quantitative measures exist which capture this unique scenario. In order to obtain an objective measure of speech understanding in television style scenarios comparable to other clinical speech in noise measurements, a modified version of the Revised Speech Perception in Noise (R-SPIN) was used. This has added broadcast sound elements to make the measure more ecologically valid to television listening and has been previously used in a number of studies.[20,21,39]

4.1 Revised Speech Perception in Noise Test

The Revised Speech Perception in Noise (R-SPIN) test is a widely used research test[18,40–44] The original R-SPIN stimuli were developed to evaluate both top-down and bottom-up processes involved in understanding speech in noise.[18] This was achieved by controlling the predictability of the sentences, either giving the listeners clues to the keyword (e.g. *'Stir your coffee with a **spoon**'* and termed *high predictability*) or no clues (e.g. *'Bob could have known about the **spoon**'* and termed *low predictability*) (keywords are noted in **bold**). Recognition of the keyword in these low predictability (LP) sentences relies entirely on receiving the acoustic signal of the keyword correctly. The high predictability (HP) stimuli differ in that the surrounding sentences allows for the use of top-down processing: any ambiguity in the keyword's acoustic signal can be resolved using an individuals knowledge of the English language and the contextual information provided by the sentence. In this way a base line speech in noise measurement can be obtained (LP) as well as a measure of the individual's ability to leverage the context of speech (HP), which is likely to occur in television speech style scenarios.

The original sentence recordings had a male speaker with an American accent and were recorded on magnetic tape.[45] The new recordings of the sentences used here were re-recorded in a high quality digital format (48kHz; 32 bit) by BBC R&D in Received Pronunciation. This re-recording is referred to here as the R^2-SPIN. The methodology for re-recording can be found in.[19] The 12 talker babble noise from the original recording was still used.

In addition to the HP and LP conditions, a third condition with HP sentences and added broadcast style sound effects was included. The sound effects selected were taken from broadcast quality sound effects libraries (BBC Sound Effects Library[46] and Soundsnap[47]). They were chosen to have equivalent contextual information as the language cues in the HP sentences. For example, the HP sentence for the keyword *pet* is *'My son has a dog for a pet'*, which utilises the assumed knowledge that children often keep pet dogs. As such, the sound effect selected for this keyword was a dog's bark. These overlapped the sentences preceding the keyword but never overlapped the keyword.

The stimuli were reordered into a multiple signal to noise ratio (SNR) paradigm to allow determination of the 50% point on the psychometric curve to be determined. It also addresses problems faced by Ward previously utilising a single SNR paradigm for hard of hearing listeners which had restricted comparability between subjects.[21] The range of SNRs used was based on those selected by participants in the previous study,[21] in the range of +14dB to -1dB. 3dB steps were chosen and 6 sentences for each condition were selected for each SNR. To smooth the psychometric curve, a similar methodology to that used by Wilson et. al.[42] was taken to determine which sentence/masker combinations had a higher average intelligibility and which had a lower average intelligibility. In contrast to Wilson's method which established this using human responses an objective intelligibility measure was used: the glimpse proportion.[48] Once the SNRs for each sentence were selected, the conditions were randomised into three lists containing 6 sentences at each SNR, with two from each condition.

Methodology The stimuli were co-located and presented from a Genelec 8030A Studio Monitor mounted at a height of 1.1 m. The listener was situated 1.1 m from the speaker. Half the participants

completed the task in a listening room meeting the ITU-R BS.1116-1 standard for listening tests[49] whilst the other half completed it in an acoustically treated but not isolated room. The stimuli were presented at a sound pressure level of 69 dB(A), measured at the listening position (as in similar studies[39,41,50]). The three main lists were presented to the participants in a pseudo random order. A practice list, with higher SNRs were given to them first to familiarise themselves with the procedure.

Preliminary Results To calculate the SNR of the 50% was the following equation was used:

$$50\%\ \text{SNR} = 15.5\text{dB} - \left(\frac{\text{Total Correct}}{2}\right)$$

The total correct was averaged as there are there are 6 keywords for each 3dB step, unlike in the QuickSIN equation (as QuickSIN has only 5 keywords per 5dB step). The results can be seen in Figure 2. It can be seen that for the majority of participants the LP condition requires 3dB or more SNR than the other conditions which is expected as the LP sentences present the most difficult stimuli. Most participants exhibit similar SNRs for the sounds effects (SFX) and HP conditions, with 7 participants having a lower SNR for HP and 8 having a lower SNR for SFX. This suggests both types of contextual information provide a benefit but the magnitude of this benefit differs between subjects. Two participants, 7 and 9, have all conditions very close to the maximum possible SNR which likely indicates a larger range of SNRs may be necessary.

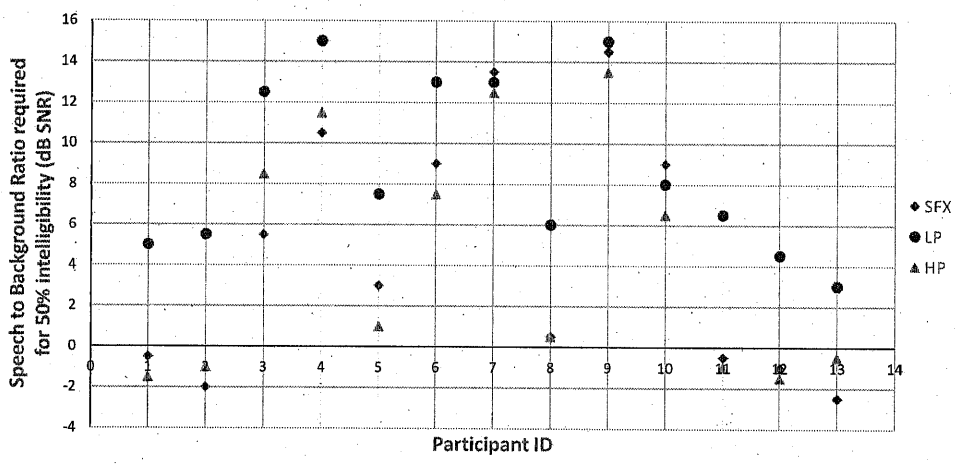

Figure 2: Speech to background ratio required to achieve 50% intelligibility for each participant in the three experimental conditions

4.2 Survey

Both closed- and open-ended qualitative questions were used to fully characterise the experiences of viewers and the problems they encounter with television speech, in addition to quantitative Likert scale questions in the same form as the items in the SSQ49. These questions can be seen in Appendix 2 in addition to the self-report measure of hearing loss.

Closed-ended qualitative questions were used to characterise a respondent's television watching habits by amount of content watched and most commonly watched genres. Participants were also asked how often they use subtitles, watch sign-interpreted programming and programs not in their native language. The selection of programming genres used were informed by the categorisation of programmes utilised by BBC iPlayer. One question queried the participants perception of how helpful different broadcast elements are to following the plot of a programme by asking respondents to recall a recent drama they have watched. The categories of broadcast elements used are based on

those utilised in the study by Shirley et. al.,[38] in turn based on work categorising broadcast objects by Woodcock et. al.[51] Drama was selected as it is a genre in which all these objects types are regularly present. The final closed-ended qualitative question aimed to to determine in which genres respondents find speech easiest to understand.

There were seven quantative Likert scale questions termed TV7 (questions 1, 4, 5 and 7 – 10 in the 'Your experience of television' section of Appendix 2. These begin by determining generally how difficult they find understanding speech on television. The following questions investigated the difficulty participants experienced in different scenarios typical to a variety of genres such as a new anchor reporting from a quiet studio or speech of panel show contestants against audience laughter. Participants were also asked to quantify the effort needed to understand speech on television.

Two open-ended qualitative items were also included to allow individuals to explore ideas beyond the structured scope of the survey and relate any other problems or experiences that they wished to express. One of these questions focused on previous experiences of problems with speech intelligibility on television whilst the other asked respondents to envisage what measures would improve their experience of television sound.

Methodology Along with the questions in Section 3.4, these questions were part of the survey conducted online by participants in their own time.

Preliminary Results The types of content most commonly watched by participants were 'News and Current Affairs' - 11 participants, 'Documentary' - 10 participants and 'Drama and Soaps' - 9 participants. The amount of hours per day spent watching television can be seen in Table 3. When asked which genre of content do they find the easiest to understand the speech in, the most commonly identified genre, by 10 participants was 'News and Current Affairs', followed by Documentary, identified by 8 participants. Only one participant identified 'Drama and Soaps'. When asked about which sounds help to follow a plot in a drama, almost all participants identified dialogue (12 participants) and only three participants identified foreground sounds. The other two categories of sound were not identified at all.

Table 3

SNR Loss	Number of Participants
Less than 1 hr	0
1-2hrs	5
3-4hrs	6
More than 4hrs	2

All except one participant reported using subtitles, with the majority using them regularly (on average 6.6 out of ten when asked to rate how regularly they use subtitles). Only one participant reported watching signed content and they reported watching it rarely.

Participants reported a range of difficulties when asked how hard they generally find television speech to understand (Appendix 2 Question 1), spanning almost the whole range from 0/10 - Very Difficult to 9/10 Easy. The mean rating was 4.7 and the median was 5.0. Of the different scenarios in the TV7 questions, participants on average rated the panel shows scenario hardest (mean rating = 3.2) and news anchor scenario easiest (mean rating = 7.1). This corresponds with the responses from the qualitative questions. When averaged over the 7 questions, the difficulty ratings ranged from 0.9 to 7.9, with a mean of 4.4 and median of 4.6.

A preliminary thematic analysis of the qualitative responses was undertaken. When asked what could be improved to make speech easier to understand the most common responses were related to the balance between speech levels and background sound levels (reported by 8 participants) and delivery of speech, including pace, clarity and people talking one at a time (7 participants). Other themes identified by two or more than participants were the quality/availability of subtitling (4 participants),

accents (3 participants), speakers facing forward to allow lipreading (2 participants) and the need for familiarity with the context of the speech (2 participants).

5 Future Work

This paper reports on the first cohort of participants collected for U-SAID. Once participant data collection has been completed, the database will be released open source for researchers to use. Analysis of the data-set will be undertaken, including principle component analysis of the quantitative data (PTA, QuickSin, TFS-AF, R^2-SPIN, SSQ49 and TV7) to investigate the relationship between different measures of hearing loss and experience of television. Directed Content analysis will be performed on the open-ended qualitative survey responses to elucidate the key problems and areas of improvement for television speech.

It is envisaged that this data may be used for scientific studies investigating the relationship between different characteristics of hearing loss and level of (dis)ability in everyday scenarios. Additionally, it may be leveraged in engineering applications to inform personalisation systems for next generation audio. In particular, the authors plan to utilise the database to develop a calibration procedure for the audio personalisation technology described in [52].

6 Conclusions

This paper reports on the design and development of the University of Salford media Accessibility and hearing Impairment Database (U-SAID). This database aims to address the lack of comprehensive and objective data available on relationship between different characteristics of hearing loss and speech understanding in television content. At date of submission, data had been collected for 13 participants for whom their preliminary results are outlined. Work is ongoing to expand the database and utilise it to understand how television speech can be made more accessible for those with hearing impairments.

Acknowledgements

The collection of this data was funded by an Institute of Acoustics Bursary. Lauren Ward is funded by the General Sir John Monash Foundation. The authors would like to acknowledge the support of additional organisations: the University of Manchester Centre for Audiology and Deafness for providing audiometry equipment and the TFS-AF test, the Department of Speech & Hearing Science at University of Illoinis in providing the R-SPIN sentences and the Audio Team at BBC R&D in re-recording the sentences. This work was also supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1).

Lauren Ward and Ben Shirley would like to gratefully acknowledge the hard work of Will Bailey, Lara Harris, Zuza Podwińska and Georgia Shirley in collecting the data. Additionally they would like to recognise the assistance and phenomenal organisational prowess of Philippa Demonte.

REFERENCES

1. Action on Hearing Loss. Hearing matters report, 2015.
2. 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.
3. B. Shield. Evaluation of the social and economic costs of hearing impairment. Oct 2006.
4. T. N. Roth, D. Hanebuth, and R. Probst. Prevalence of age-related hearing loss in europe: a review. Eur. Arch. Otorhinolaryngol., 268(8):1101–1107, 2011.
5. The Nielsen Company (US). The total audience report Q1, 2017. 2017.
6. Broadcasters Audience Research Board. Trends in television viewing 2017, Feb 2018.
7. D. Cohen. Sound matters, BBC College of Production, March 2011.
8. M. Armstrong. BBC white paper WHP 324: From clean audio to object based broadcasting, Oct, 2016.
9. P. Mapp. Intelligibility of cinema & tv sound dialogue. In 141st Audio Eng. Soc. Convention, Sept 2016.
10. O. Strelcyk and G. Singh. Tv listening and hearing aids. PLOS One, 13(6), 2018.
11. P. T. Johannesen, P. Pérez-González, S. Kalluri, J. L. Blanco, and E. A. Lopez-Poveda. The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise for hearing-impaired listeners. Trends in hearing, 20:2331216516641055, 2016.
12. V. Summers, M. J. Makashay, S. M. Theodoroff, and M. R. Leek. Suprathreshold auditory processing and speech perception in noise: hearing-impaired and normal-hearing listeners. Journal of the American Academy of Audiology, 24(4):274–292, 2013.
13. C. Füllgrabe and B. C. J. Moore. Evaluation of a method for determining binaural sensitivity to temporal fine structure (tfs-af test) for older listeners with normal and impaired low-frequency hearing. Trends in hearing, 21:2331216517737230, 2017.
14. C. Füllgrabe, A. J. Harland, A. P. Sęk, and B. C. J. Moore. Development of a method for determining binaural sensitivity to temporal fine structure. International journal of audiology, 56(12):926–935, 2017.
15. B.S.A. Recommended Procedure: Pure-tone air-conduction and bone conduction threshold audiometry with and without masking. Technical report, 2017.
16. M. C. Killion, P. A. Niquette, G. I. Gudmundsen, L. J. Revit, and S. Banerjee. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am., 116(4):2395–2405, 2004.
17. S. Gatehouse and W. Noble. The speech, spatial and qualities of hearing scale (SSQ). Int. J. Audiol., 43(2):85–99, 2004.
18. 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.
19. L. Ward, M. Paradis, and C. Robinson. R^2spin: Re-recording the revised speech perception in noise test. 2018. [In Preparation].
20. L. Ward, B. Shirley, Y. Tang, and W. J. Davies. The effect of situation-specific acoustic cues on speech intelligibility in noise. In Interspeech 2017, pages 2958–2962, Stockholm, Sweden, Aug. 2017. ISCA.
21. L. Ward and B. Shirley. Television dialogue; balancing audibility, attention and accessibility. In Conference on Accessibility in Film, Television and Interactive Media, York, UK, Oct. 2017.
22. M. C. Killion and P. A. Niquette. What can the pure-tone audiogram tell us about a patient's snr loss. Hear J, 53(3):46–53, 2000.
23. J. Swaminathan, C. R. Mason, T. M. Streeter, V. Best, E. Roverud, and G. Kidd. Role of binaural temporal fine structure and envelope cues in cocktail-party listening. Journal of Neuroscience, 36(31):8250–8257, 2016.
24. 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. Ageing Neurosci., 6, 2014.
25. I. J. Moon and S. H. Hong. What is temporal fine structure and why is it important? Korean journal of audiology, 18(1):1, 2014.
26. O. Strelcyk and T. Dau. Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. Jour. Acoust. Soc. Am., 125(5):3328–3345, 2009.

27. A. J. Oxenham, J. G. W. Bernstein, and H. Penagos. Correct tonotopic representation is necessary for complex pitch perception. Proceedings of the National Academy of Sciences, 101(5):1421–1425, 2004.
28. K. Hopkins and B. C. J. Moore. The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. The Journal of the Acoustical Society of America, 125(1):442–446, 2009.
29. K. Hopkins, B. C. J. Moore, and M. A. Stone. Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. The Journal of the Acoustical Society of America, 123(2):1140–1153, 2008.
30. Kathryn Hopkins and Brian CJ Moore. Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies. International Journal of Audiology, 49(12):940–946, 2010.
31. B. C. J. Moore and A. Sek. Development of a fast method for determining sensitivity to temporal fine structure. International Journal of Audiology, 48(4):161–171, 2009.
32. Etymotic Research Inc. Quicksin: Speech-in-noise test. 1.3, 2006.
33. T. W. Tillman and W. O. Olsen. Speech audiometry. Modern developments in audiology, 2:37–74, 1973.
34. M. A. Akeroyd, F. H. Guy, D. L. Harrison, and S. L. Suller. A factor analysis of the SSQ (speech, spatial, and qualities of hearing scale). International journal of audiology, 53(2):101–114, 2014.
35. W. Noble, N. S. Jensen, G. Naylor, N. Bhullar, and M. A. Akeroyd. A short form of the speech, spatial and qualities of hearing scale suitable for clinical use: The SSQ12. Int. Jour. Audiology, 52(6):409–412, 2013.
36. G. Singh and M. K. Pichora-Fuller. Older adults' performance on the speech, spatial, and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of interview and self-administration methods. International Journal of Audiology, 49(10):733–740, 2010.
37. K. Demeester, V. Topsakal, J. Hendrickx, E. Fransen, L. van Laer, G. Van Camp, P. Van de Heyning, and A. Van Wieringen. Hearing disability measured by the speech, spatial, and qualities of hearing scale in clinically normal-hearing and hearing-impaired middle-aged persons, and disability screening by means of a reduced SSQ (the SSQ5). Ear and hearing, 33(5):615–616, 2012.
38. 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.
39. L. Ward, B. Shirley, and W. J. Davies. Turning up the background noise; the effects of salient non-speech audio elements on dialogue intelligibility in complex acoustic scenes. In Proc. of Institute of Acoustics 32nd Reproduced Sound Conf., Southampton, Nov. 2016. IOA.
40. B. Spehar, S. Goebel, and N. Tye-Murray. Effects of context type on lipreading and listening performance and implications for sentence processing. Journal of Speech, Language, and Hearing Research, 58(3):1093–1102, 2015.
41. B. Shirley. Improving Television sound for people with hearing impairments. PhD thesis, University of Salford, 2013.
42. 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.
43. A.R. Carmichael. Evaluating digital "on-line" background noise suppression: Clarifying television dialogue for older, hard-of-hearing viewers. J. Neuropsychol. Rehab., 14(1-2):241–249, 2004.
44. 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.
45. R.C. Bilger, J.M. Nuetzel, W.M. Rabinowitz, and C. Rzeczkowski. Standardization of a test of speech perception in noise. J.Speech.Lang.Hear.Res., 27(1):32–48, 1984.
46. BBC. BBC Sound Effects Library CDs 1-60.
47. Tera Media and CRG. Soundsnap.com.
48. M. Cooke. A glimpsing model of speech perception in noise. J. Acoust. Soc. Am., 119(3):1562–1573, 2006.
49. ITU Recommendation. ITU-R BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. 1997.
50. L. E. Humes, B. U. Watson, L. A. Christensen, C. G. Cokely, D. C. Halling, and L. Lee. Factors associated with individual differences in clinical measures of speech recognition among the elderly. J. Speech, Lang. Hear. Res., 37(2):465–474, 1994.
51. J. S. Woodcock, W. J. Davies, T. J. Cox, and F. Melchior. Categorization of broadcast audio objects in complex auditory scenes. J. Audio Eng Soc., 2016.
52. L. Ward, B. Shirley, and J. Francombe. Accessible object-based audio using hierarchical narrative importance metadata. In 145th Audio Eng. Soc. Convention. Audio Engineering Society, 2018.

APPENDIX 1

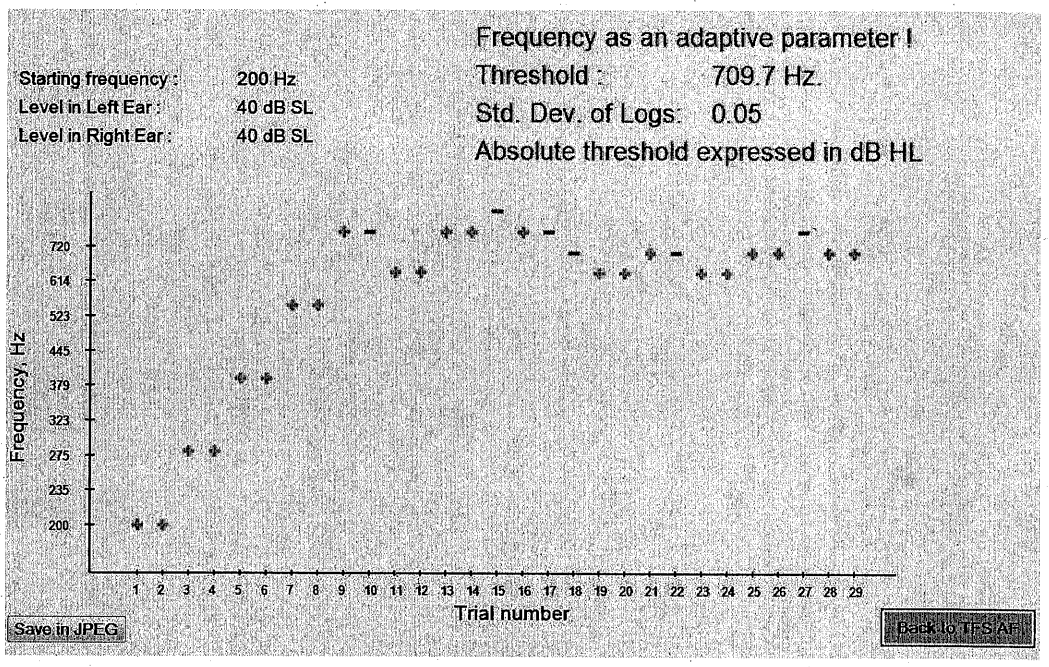

Figure 3: Example output from TFS-AF test

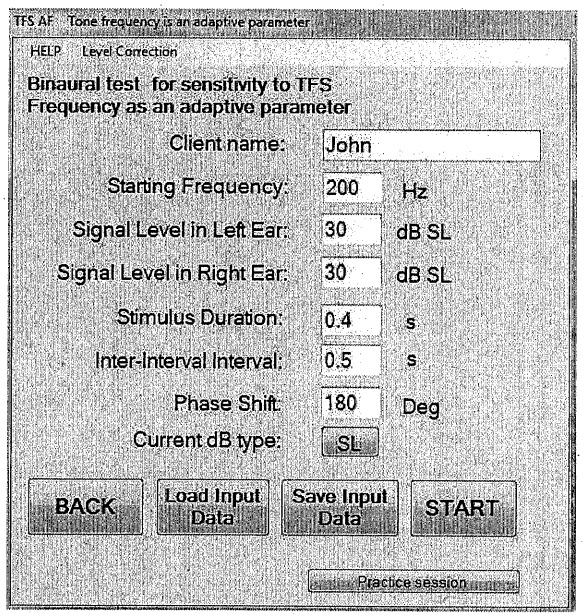

Figure 4: Settings utilised for the TFS-AF test

APPENDIX 2

About you

1. What is your age?
2. Where do you live? (country of residence)
3. What is your native language?
4. How many hours a day do you watch television, on average?
   (a) Less than 1 hr
   (b) 1 - 2hrs
   (c) 3 - 4 hrs
   (d) More than 4 hrs
5. What type of programming do you mostly watch? (you may select more than one option)
   (a) News and Current Affairs programming
   (b) Drama and Soaps
   (c) Sports
   (d) Documentary
   (e) Comedy
   (f) Lifestyle, Music and Food
   (g) Films
6. Are you a musician?
   (a) No
   (b) Yes, I am a professional musician
   (c) Yes, I am an amateur musician
7. Do you identify as Hard of Hearing?
   (a) Yes
   (b) No
   (c) Prefer not to say

About your hearing

1. What degree of hearing loss do you have?
   (a) Mild
   (b) Moderate
   (c) Severe
   (d) Profound
   (e) I do not have hearing loss
   (f) I'm not sure
2. Do you suffer from Tinnitus?
   (a) No
   (b) Yes
3. Do you use any assistive hearing devices?
   (a) No
   (b) Yes, only in my right ear

(c) Yes, only in my left ear

(d) Yes, in both ears

4. If you answered yes above, what sort of assistive device do you use? (leave blank if not applicable)
5. Do you regularly wear this device? (leave blank if not applicable)
   (a) Yes
   (b) No
6. How many years have you had this device fitted? (leave blank if not applicable)

Your Experience of Television

1. Generally, how difficult do you find it to understand speech on television?

```
Very                                          Very
Difficult (0)                                 Easy (10)
0    1    2    3    4    5    6    7    8    9    10
```

2. Thinking about a recent drama you have watched on television. Which of the following sounds helped you personally to follow the plot? (Select as many as you think apply)
   (a) Dialogue
   (b) Foreground sounds (e.g.the main character slamming a door in anger or other sounds that the characters can hear)
   (c) Background sounds (e.g. sounds of birds in the countryside, background chatter in a pub scene)
   (d) Music
3. Which of the following types of television content do you find the dialogue/speech easiest to understand whilst watching? (Select as many as you think apply)
   (a) News and Current Affairs
   (b) Drama and Soaps
   (c) Sports
   (d) Lifestyle, Music and Food
   (e) Documentary
   (f) Comedy
   (g) Films
4. A character is speaking but they are not on screen. How easily can you understand the speech without seeing the character's face?

```
Not
at all (0)                                      Perfectly (10)
0    1    2    3    4    5    6    7    8    9    10
```

5. You are watching a panel show and one of the panellists is speaking whilst the studio audience laughs and cheers. How easily are you able to understand the panellist's speech?

```
Not
at all (0)                                      Perfectly (10)
0    1    2    3    4    5    6    7    8    9    10
```

6. How often do you use subtitles?

```
Never (0)                                       Always (10)
0    1    2    3    4    5    6    7    8    9    10
```

7. A news presenter is reporting from a quiet studio. Without using subtitles, how easily can you understand the speech?

```
Not
at all (0)                                          Perfectly (10)
0    1    2    3    4    5    6    7    8    9    10
```

8. You are watching a scene on television which has the sound of clinking glasses, music and people talking in the background. Can you make out the different sounds?

```
Not
at all (0)                                          Perfectly (10)
0    1    2    3    4    5    6    7    8    9    10
```

9. You are watching a nature documentary. The narrator is speaking with the constant sound of a waterfall in the background. Can you follow what the narrator is saying?

```
Not
at all (0)                                          Perfectly (10)
0    1    2    3    4    5    6    7    8    9    10
```

10. How much effort do you require to hear what is being said in a television drama?

```
A lot
of Effort (0)                                          No Effort (10)
0    1    2    3    4    5    6    7    8    9    10
```

11. How often do you watch television programs which are not in your native language?

```
Never (0)                                          Always (10)
0    1    2    3    4    5    6    7    8    9    10
```

12. When sign-interpretation is available, how often do you watch sign language interpreted programming?

```
Never (0)                                          Always (10)
0    1    2    3    4    5    6    7    8    9    10
```

13. What measures do you feel would make television speech easier for you to understand?
14. Are there any specific examples of times when you have found television speech hard to understand that you would like to tell us about?