Television Dialogue; Balancing Audibility, Attention and Accessibility

LAUREN WARD[1] and BEN G SHIRLEY[1]

[1]Acoustics Research Centre, University of Salford, Manchester, UK
e-mail: L.Ward7@edu.salford.ac.uk, B.G.Shirley@salford.ac.uk

Abstract

Sound effects and other non-speech broadcast elements play many roles within television and radio content, including progressing the narrative. However, accessibility strategies for hard of hearing listeners tend to reduce all non-speech elements equally, regardless of their narrative importance. This work considers what effect narratively important sound effects have on dialogue intelligibility and whether their narrative benefit outweighs their potential to mask speech for hard of hearing listeners.

This paper summarises previous work by the authors which showed the addition of relevant sound effects consistently improved keyword recognition in noise for normal hearing listeners. The current work investigates this effect with hard of hearing listeners. For unpredictable speech, this work shows that how much sound effects improve keyword recognition monotonically decreased as a listener's audiometric hearing loss, in their better hearing ear, increased. For predictable speech, inclusion of sound effects improved keyword recognition by 13.2% on average (compared with 18.7% for normal hearing listeners). However, this improvement was less consistent than for normal hearing listeners and did not display the same monotonic relationship with hearing loss severity as unpredictable speech. Other factors which may influence the narrative benefit of sound effects, including their potential to mask speech, are discussed.

Ongoing work to further characterise the relationship between sound effects, narrative benefit, and masking potential for hard of hearing listeners is described. Implications for object-based accessibility solutions for hard of hearing listeners as well as for accessibility strategies for the visually impaired like Enhanced Audio Description are also outlined.

1 Introduction

Sound effects (SFX) play many roles within television and other broadcast content including establishing location, signalling key events, and facilitating continuity between scenes [1]. In particular, diagetic SFX can often take on important roles in progressing the plot [2]. For example, the off screen sound of a car screeching to a halt, stomping footsteps and a key turning in a lock before a character enters a room informs the viewer that someone has arrived, angrily, who lives there. Such sounds could not be removed without substantially altering how effectively the narrative is conveyed [1]. The role such SFX play in carrying narrative elements is even more vital in accessibility strategies for people with visual impairments, such as *audio films* and Enhanced Audio Description [3, 4].

In the UK alone, there are estimated to be 11 million individuals with some degree of hearing impairment, and with an ageing population, this figure is likely to rise [5, 6]. Despite the narrative role many non-speech sounds are designed to play within television content, the accessibility strategies for these viewers have traditionally treated all non-speech sounds equally: as maskers. Subsequently these strategies have aimed to suppress all non-speech content whilst enhancing the dialogue [7, 8]. For legacy content, where all sound elements are mixed before broadcast and separate elements cannot be manipulated at point of service, this is a necessary approach. Object-based content however does not have this constraint as it has the flexibility for sound elements to be transmitted as separate objects, which can be rendered differently at point of service based on metadata [9]. This can allow the balance between different sound objects to be personalised by the viewer. Previous work has explored how allowing hard of hearing viewers to personalise the balance between dialogue, diagetic foreground sounds, background sounds and music can have a positive benefit on their understanding of the content [2]. The flexibility of object-based broadcasting enables the development of more nuanced and personalised approaches to accessibility for hard of hearing individuals.

In order to deliver improvements in accessibility and create practical tools for personalisation of content, a greater understanding of how different broadcast sound elements affect dialogue intelligibility for hard of hearing listeners is required [1]. The ongoing work described by this paper is endeavouring to address this need. In particular, it aims to answer the question; 'Do narratively important SFX aid dialogue intelligibility?' This paper describes prior work by this group with normal hearing listeners which has motivated the current experimental approach. The paper describes experimental results from a hard of hearing cohort followed by a discussion of these results. The implications of these results for broadcast accessibility strategies is outlined.

2 Prior Work

There are very few studies which quantitatively explore the effect non-speech sounds have on intelligibility for normal hearing listeners [10, 11]. For public address style speech, a 2016 study showed that preceding sounds cues can positively influence the intelligibility [10]. Related concepts have been explored in studies of knowledge transfer in multimedia learning, yielding different results; a study by Moreno in 2000 showed that, for instructional messages, additional audio elements can overload the listeners' working memory [11].
Prior work by this group [12, 13] has investigated the effects of narratively important, broadcast type SFX on the intelligibility of speech in multi-talker babble. This study was undertaken with twenty-four self-reported normal hearing, native English speakers and the remainder of this section outlines its methodology and results.

2.1 Experimental Tools and Methodology

This study used a modified version of the Revised Speech Perception in Noise (R-SPIN) test [14, 15]. The R-SPIN test has been widely adapted [16, 17, 18] to investigate the influence of different factors on speech intelligibility in noise. The original R-SPIN stimuli consist of short, phonetically balanced sentences spoken by a male speaker in American English, presented in multi-talker babble. All sentences end with a monosyllabic noun, the keyword, which participants are scored on their ability to correctly identify. The original test evaluates the effect that the predictability of the sentence has on intelligibility. This is achieved through high and low predictability 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).

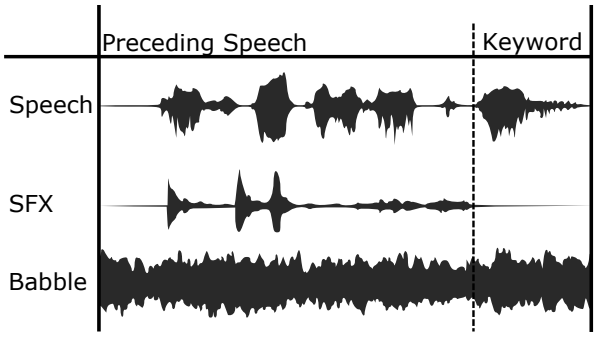

Figure 1: Example stimuli, noting alignment of the SFX and keyword

The modified version used in [12, 13] added SFX to half the stimuli, to evaluate the effect of relevant SFX on intelligibility as well as how the effect of SFX may interact with the predictability of the speech. This gave four stimuli types in the modified version: low predictability sentences, high predictability sentences, low predictability with SFX and high predictability with SFX. The modified version used only half of the 400 sentences from the original test. This gave 200 sentence stimuli which included the high and low predictability version of 100 different keywords.

The SFX selected were taken from broadcast quality SFX libraries (BBC Sound Effects Library [19] and Soundsnap [20]). They were selected to give approximately the same clues to the keyword as the preceding speech in the high predictability sentences. For example, the SFX selected for the sentence *'My son has a dog for a **pet**'* was a dog's bark. All SFX ended prior to the keyword being spoken, as seen in Figure 1. Regardless of whether the background contained babble only or babble and SFX, the loudness of the background sounds were normalised to the same level, using ITU-R BS.1770-2 [21]. The ratio of speech to background was set to -2dB and the stimuli were co-located and played from a loudspeaker at 69dBSPL.

To investigate the potential for the introduced SFX to behave as energetic maskers over the speech preceding the keyword, the signal level intelligibility was evaluated using the glimpse proportion [22]. The glimpse proportion quantifies the number of time-frequency units for which the speech survives energetic masking (i.e. has energy at least 3dBSPL greater than the masker). It reflects the local audibility of speech in noise and higher glimpse proportions correlate with greater intelligibility. Calculation of the glimpse proportion over the keyword speech also facilitated evaluation of whether all keywords, regardless of experimental condition, were energetically masked by the babble equivalently.

[12, 13] contain a complete description of the experimental method and tools used in this study.

2.2 Results

2.2.1 Perceptual intelligibility results

For low predictability sentences with no SFX the mean word recognition rate was 35.8%. High predictability sentences with no SFX improved this to a word recognition rate of 62.1%. Inclusion of the SFX to the sentences increased word recognition rate to 60.7% for low predictability sentences and 73.7% for high predictability sentences. The improvement in word recognition rate gained when SFX were present relative to stimuli without SFX were 60.7% and 18.7% for low and high predictability sentences respectively. The effects of both SFX, predictability and their interaction were all significant at the level $[p < 0.001]$ (evaluated with a two-way repeated measures ANOVA and Tukey's HSD post-hoc test).

2.2.2 Objective intelligibility analysis

The glimpse proportion was calculated separately over the speech preceding the keyword and the keyword itself. There was no significant difference between the glimpse proportion over the keyword for any of the experimental conditions, having a mean GP = 13.19%. This indicates that the keywords in each condition had, on average, equivalent levels of energetic masking from the babble. However for the preceding speech, which was overlapped by the SFX in half the conditions, the glimpse proportion differed significantly despite all the non-speech elements having been normalised to the same loudness levels. In conditions without SFX the preceding speech had a mean GP = 18.72% whilst, when SFX were present, this was reduced to GP = 9.96% (significantly different at the level $[p < 0.001]$). This reduction in available glimpses of the target speech is likely to have had the most effect on the condition with high predictability sentences, as the SFX may have interfered with the listeners' ability to fully utilise the clues to the keyword in the preceding speech.

2.3 Conclusions

From this study it is clear that the effect which narratively important SFX have on intelligibility in noise for normal hearing listeners is positive, large and consistent across listeners. Furthermore it appears that the perceptual benefit of the SFX outweighs any energetic masking or distracting effects it may have had (at the speech to background ratio used in this study).

3 Hard of hearing study

The above study was replicated with a hard of hearing cohort in order to determine whether the perceptual benefit SFX have for normal hearing listeners is also present for hard of hearing listeners.

3.1 Cohort

Fourteen predominantly older native English speakers took part. Audiometric thresholds over the frequencies 0.25Hz, 0.5Hz, 1kHz, 2kHz, 4kHz, and 8kHz were obtained for all participants, using a Kamplex r27a Diagnostic Audiometer. The mean pure tone average, at speech frequencies (0.5-4kHz), across the cohort was 36dBSPL (standard deviation = 21dBSPL) and 49dBSPL (standard deviation = 27dBSPL) for their better and worse hearing ears respectively. The cohort had significant variation in their hearing impairments, ranging from normal hearing thresholds with tinnitus or Ménière's disease to severe loss (as defined by the British Society of Audiology [23]). The majority of the cohort had symmetric hearing loss (12 out of 14).

3.2 Alterations to Methodology

A number of alterations had to be made to the normal hearing implementation of the experiment to make it suitable for the hard of hearing cohort. Rather than a single speech to background ratio, the ratio was calibrated for each participant to ensure that they could hear the speech. This was achieved by using a set of unused sentences (without SFX) from the normal hearing implementation as calibration sentences, starting at the -2dB speech to background used by normal hearing listeners. The speech to background ratio was altered in 1dB increments until the participant expressed that they could understand approximately half of the sentences. This resulted in a wide range of speech to background ratios, from -2dB, the same as the normal hearing listeners, up to +12dB. Participants were also allowed to make small modifications to the overall reproduction level (between +4dBSPL and -2dBSPL from the original 69dBSPL level).

Only half the sentences from the normal hearing implementation of the experiment were used: 100 sentences with 50 different keywords. This was to ensure that the total length of the experiment, inclusive of the audiogram and calibration procedure, did not induce listener fatigue. Participants who had been fitted with a hearing aid were encouraged to wear it during the test if they usually wore it whilst watching television.

Table 1: Correlation between pure tone average (PTA, 0.5-4kHz) in better and worse ears, speech to background ratio and improvement in word recognition rate when SFX are included for low and high predictability sentences, using Spearman's two-tailed rank correlation.

	Better Ear PTA		Worse Ear PTA		Speech to Background Ratio		SFX Improvement: Low Predictability
Speech to Background Ratio	0.647	*	0.629	*	—		—
SFX Improvement:							
Low Predictability	-0.857	***	-0.709	**	-0.707	**	—
High Predictability	-0.045		0.057		-0.103		-0.045

$*p < .05, **p < .01, ***p < .001$

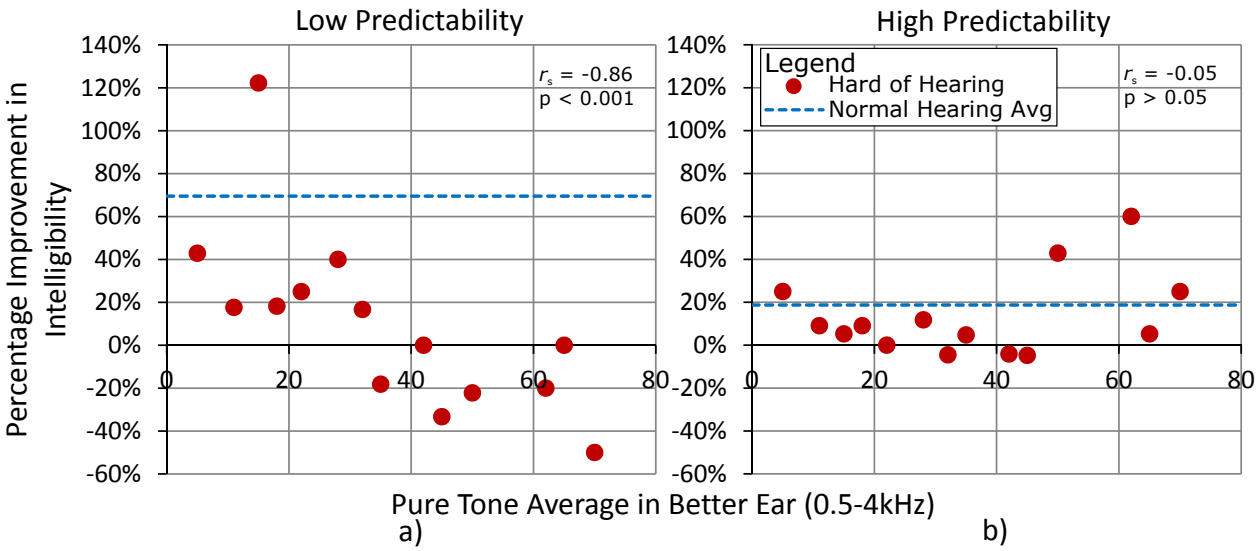

Figure 2: Scatter plot of pure tone average (0.5-4kHz) in their better hearing ear against improvement in word recognition rate when SFX were included for a) low predictability sentences and b) high predictability sentences. Average improvement for normal hearing listeners and Spearman's rank correlation coefficient is also shown.

3.3 Results

3.3.1 Perceptual intelligibility results

Preliminary results from this experiment are described as part of the doctoral work outlined in [24].

As each participants' stimuli had a different speech to background ratio, absolute word recognition rates could not be calculated. Instead the mean improvement in word recognition rate was calculated, relative to the low predictability with no SFX (control) condition for each participant. The mean improvement between the low predictability sentences and the high predictability sentences was 91.8%. There was large variation in this value, having a standard deviation of 63.0%. The benefit was positive for all listeners, except one for whom the high predictability sentences made no difference. This benefit compares closely with previously reported results for hard of hearing listeners, where high predictability sentences increase word recognition rates from 28% to 70% (at 80dBSPL and -1dB speech to background ratio) [14].

The mean improvement when SFX were added to the low predictability sentences was 9.9%, much smaller than for normal hearing listeners who exhibited a mean improvement of 69.5%. There was also a large amount of variation in this result, with a standard deviation of 42.4%. Furthermore, the SFX either degraded or had no effect on word recognition rates for some of the participants. The addition of SFX to the high predictability sentences also offered only a small mean improvement of 13.18%. However, this had a smaller standard deviation of 19.0% and was of a similar magnitude to the improvement exhibited by normal hearing listeners of 18.7% .

Correlation analysis between the experimental factors was performed and is seen in Table 1. The aim of this

Table 2: Partial correlation between SFX improvement for low predictability sentences and pure tone averages (PTA, 0.5-4kHz) in better and worse ears and speech to background ratio, using Spearman's two-tailed rank coefficient

	Better Ear PTA		Worse Ear PTA	Speech to Background Ratio
SFX Improvement: Low Predictability	-0.671	*	-0.252	-0.304

*$p < .05$

analysis was twofold. Firstly to investigate whether the selected speech to background ratio was related to the participants' pure tone averages. Secondly, to determine whether the degree to which SFX were beneficial could be explained by how audible the SFX was (given the selected speech to background ratio and the participants' degrees of hearing loss). Normality of the variables was first assessed using the Anderson-Darling test for normality. As some of the variables did not meet the normality criterion, Spearman's rank correlation coefficient was used to evaluate the relationship between the different variables.

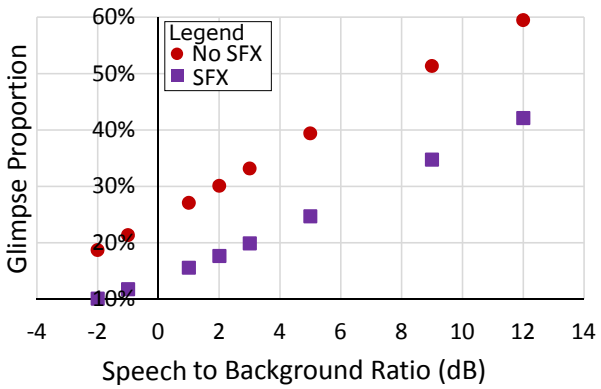

Figure 3: Mean glimpse proportion of the preceding speech for conditions with and without SFX, for each selected speech to background ratio

Table 1 indicates that the selected speech to background ratio is dependent on the pure tone average in the participant's better and worse hearing ears. The degree of improvement (or degradation) which the SFX had on word recognition rate for low predictability sentences is strongly correlated with the participants' better ear hearing. It is also correlated, though less strongly, with the worse hearing ear and selected speech to background ratio. Figure 2a) shows a scatterplot of the pure tone average in the participant's better hearing ear against the SFX improvement for low predictability sentences . It can be seen that there is a monotonically decreasing relationship between the SFX improvement and better ear hearing. In order to determine whether better ear hearing alone was a predictor for the benefit of SFX inclusion in low predictability sentences, partial correlation analysis was also performed and can be seen in Table 2. It can be seen that when the effects of the worse hearing ear and the speech to background ratio are controlled for, the participant's pure tone average in their better hearing ear remains a predictor for how beneficial SFX are to word recognition rate in low predictability speech.

Table 1 also shows that the degree to which SFX interact with low and high predictability speech in an uncorrelated manner. Figure 2b) shows a scatterplot of better ear hearing and SFX improvement for high predictability sentences. Unlike for the low predictability sentences there is no clear monotonic relationship across the range of hearing abilities. A monotonically decreasing relationship, similar to the one present for low predictability speech, does appear to exist in the region where the participants' pure tone averages are less than 50dBSPL. Correlation analysis on participants with pure tone averages below 50dBSPL was performed to determine the size and significance of this. This showed a significant montonically decreasing relationship, similar to the relationship seen for low predictability speech, with $[r_s = -0.784, p < 0.05]$.

3.3.2 Objective intelligibility analysis

The glimpse proportion for the keyword and the preceding speech were calculated separately, as for the normal hearing stimuli. A three-way ANOVA between the two experimental conditions: predictability and presence of SFX, as well as speech to background ratio was performed for both the keyword and preceding speech. This allowed for the effect of the speech to background ratio to be partialled out, as changes in this produced the most significant differences between glimpse proportion scores. For the keyword, conditions with SFX exhibited a slight difference in mean glimpse proportion, though this difference was only weakly significant $[F = 5.5, p < 0.05]$. For the preceding speech, Figure 3 shows the range of glimpse proportion values at each speech to background ratio for conditions with and without SFX. It can be seen that at all speech to background ratios there is a large, and strongly significant, difference between the glimpse proportions when SFX are present and absent $[F = 1468.3, p < 0.001]$. Having controlled for the effect of difference speech to background ratios, these results mirror those seen for the normal hearing stimuli.

4 Discussion

From these results it can be seen that for low predictability speech, the participants' pure tone averages in their better hearing ear is the strongest predictor for how beneficial narratively relevant sounds are. As the stimuli was reproduced monaurally with the babble, speech and SFX co-located, it is reasonable that performance would be dominated by the participant's better hearing ear. This relationship appears to approach the level of benefit exhibited by normal hearing listeners, as pure tone averages approach 0dBSPL. Whether better ear hearing remains a strong predictor of SFX utility when the content is reproduced in stereo or using spatialised reproduction methods needs to be further investigated.

Interestingly, for the low predictability speech not only did the utility of the SFX decrease for participants with higher pure tone averages but the presence of SFX degraded word recognition rates below that of the control condition for some participants. Given that the level of the SFX was tied to the level of the babble, for participants who selected higher speech to background ratios (predominantly those with higher pure tone averages), the SFX were presented at a lower volume. The reasons that the SFX actively degraded intelligibility may be linked to this reduced audibility, as more of the listener's attention was required to identify the quieter sound and subsequently make use of it. Furthermore, given that the preceding speech did not relate to the SFX, the process of switching attention between the speech and the SFX may have resulted in increased cognitive load [25]. This increased load potentially impaired parsing of the speech and SFX compared with when the cognitive resources are mostly dedicated to the speech alone (in the babble only conditions). A similar hypothesis was proposed in [11], where the addition of music and SFX was shown to reduce knowledge transfer in multimedia content. Whilst [11] only studied normal hearing listeners, it is possible that this effect is more prominent in those with higher degrees of hearing loss. However, given that on average the keywords of conditions with SFX had slightly higher glimpse proportions and subsequently slightly more energetic masking, it is also possible that this was having a greater impact on those with higher pure tone averages. This may have contributed to the degradation in intelligibility for these participants.

For high predictability speech it appears that for hard of hearing listeners with a pure tone average below 50dBSPL, better ear hearing remains a useful predictor of SFX benefit. As with low predictability speech, it appears as pure tone averages approach 0dBSPL, this relationship approaches the level exhibited by normal hearing listeners. However given the small size of the cohort and the large variability in their hearing impairments, it is possible that this trend may not be generalisable. As for low predictability speech, some hard of hearing listeners found SFX degraded intelligibility. In addition to the possible distraction effects of the SFX, for high predictability speech this degradation in intelligibility may be due to the SFX energetically masking the clues from the preceding speech (as indicated by the significantly reduced glimpse proportion when SFX were present). However, unlike for low predictability speech, for listeners with pure tone averages about 50dBSPL the SFX did not degrade intelligibility. It is possible in this condition that the preceding high predictability speech aided the listener in identifying the SFX, rather than the other way around. The overall effect being that, despite listeners' difficulty in identifying the SFX, the SFX still acted as redundant information for determining the keyword. It is however evident that a more complex relationship between better ear hearing and SFX utility exists for high predictability speech than for low predictability, which warrants further investigation.

5 Implications for Accessibility

The results of this paper highlight that accessibility strategies which treat hard of hearing listeners' needs as homogeneous are unlikely to be broadly effective. These results are particularly significant as the majority of those with hearing loss in the UK (91.7%) have mild to moderate loss [5] and the results given here indicate that this listener group varies broadly in how SFX affect intelligibility.

These results indicates that there is subset of hard of hearing listeners for whom narratively important SFX will aid intelligibility. This is consistent with previous subjective work where, when hard of hearing listeners were given the opportunity to alter the volume of different object categories (dialogue, diagetic foreground SFX, background SFX and music) for what they personally felt gave the greatest understanding of on screen action, a subset of participants consistently set the foreground SFX higher than other non-speech objects (4 out of 15) [2]. Such a subset is also mirrored by the results of an ongoing survey of television experience and hearing[1]. When asked to consider a recent drama they have watched, only 20% of hard of hearing respondents (to date) reported that they felt foreground SFX aided their understanding of the dialogue ($n = 15$). For object-based broadcasting methods, which give the potential for end-users to personalise the balance between different sound elements for intelligibility, the results here begin to define possible predictors for different user groups.

[1] Take the survey at conducted by this research group http://bit.ly/soundTV

Such predictors could be utilised to determine optimal preset volume balances for content between broadcast sound objects based on the end-users' degree of hearing impairment.

Interestingly, results from the ongoing survey indicate that the proportion of normal hearing respondents who reported foreground SFX aid their understanding was also small, 44.4% ($n = 37$). Given that for normal hearing listeners SFX were consistently beneficial this indicates that what is beneficial in terms of intelligibility may not be what is considered preferential by listeners. As such, accessibility strategies based around characterising user needs should still maintain the ability for the listener to adjust any calibrated levels based on their preferences as well as needs.

The way SFX interact with speech intelligibility also has implications for accessibility strategies for people with visual impairments. The provision of audio description, where the visual modality is compensated for with greater amounts of speech, increases the possibility of speech overlapping with SFX. The potential for this increases further for Enhanced Audio Description, which may have also have greater amounts of SFX [3]. Furthermore, as hearing impairment becomes more prevalent with age, as does vision loss [26]. The potential for masking from the SFX and degradation of intelligibility should be considered when SFX overlap speech, in particular for content which may be targeted towards an older audience.

5.1 Ongoing Work

Ongoing work by this group aims to further characterise the relationship between audibility of SFX, attentional factors, masking and intelligibility. One of the limitations of the current experimental method is the individually calibrated speech to background ratios. For this reason, and utilising the adaptation of the R-SPIN test by Wilson et. al [18], ongoing work will utilise a multiple speech to background ratio paradigm. This approach, based on the results of ongoing pilots, will remove the need to calibrate individual ratios and allow all participants to utilise the same stimuli. Furthermore, this will allow the determination of a 50% speech reception threshold under each experimental condition for each listener, which may facilitate better comparison between these results and other speech in noise studies.

Another limitation of the current method was that the level of the SFX was tied to the level of the babble masker. This approach has meant the results give an insight into the effect of the SFX on intelligibility of legacy content where all non-speech elements are likely to be reduced in volume together. However, as the level of the SFX is free to be altered within object-based content future experimental work needs to accommodate this additional degree of freedom. Experimental work currently being piloted will begin to explore this through two conditions: SFX at -6dB below the speech level and SFX and speech equally loud. Exploration of attentional effects is also planned through the use of self-report measures as well as modification of stimuli to mimic reduced cognitive load conditions.

6 Conclusions

This paper begins to address the lack of quantitative study into the effect of narratively relevant sounds on speech intelligibility. It has demonstrated that the inclusion of narratively relevant SFX can aid keyword recognition in noise for some hard of hearing listeners. Furthermore, the strongest predictor of whether SFX give perceptual benefit for a particular listener is the severity of hearing loss in the listeners' better hearing ear (if SFX, low predictability speech and masker are co-located). When speech is highly predictable, the presence of SFX gives a mean 13.2% improvement in keyword recognition for hard of hearing listeners. For those with pure tone average hearing loss below 50dBSPL, their better ear hearing is also a predictor for how beneficial SFX are likely to be for high predictability speech. These results give the basis for developing personalised accessibility strategies for hard of hearing listeners using object-based broadcasting methods. Further characterisation of the relationship between narratively relevant sounds and intelligibility at different speech to background ratios is still required.

7 Acknowledgements

Lauren Ward is supported by the General Sir John Monash Foundation.

References

[1] M. Armstrong. (Oct, 2016) BBC white paper WHP 324: From clean audio to object based broadcasting. BBC. [Online]. Available: http://www.bbc.co.uk/rd/publications/whitepaper324

[2] B. G. Shirley, M. Meadows, F. Malak, J. S. Woodcock, and A. Tidball, "Personalized object-based audio for hearing impaired TV viewers," *Journal of the Audio Engineering Society*, vol. 65, no. 4, pp. 293–303, 2017.

[3] M. Lopez, "Perceptual evaluation of an *audio film* for visually impaired audiences," in *Proc. 138th*

Audio Engineering Society Convention. Warsaw, Poland: AES, May 2015.

[4] M. Lopez and G. Kearney, "Enhancing audio description: sound design, spatialisation and accessibility in film and television," in *Proc. of Institute of Acoustics 32nd Reproduced Sound Conf.* Southampton, U.K.: IOA, Nov. 2016.

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

[6] Office for National Statistics. (Oct, 2015) National population projections: 2014-based statistical bulletin. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2015-10-29#older-people

[7] M. Armstrong. (Jan, 2011) BBC white paper WHP 190: Audio processing and speech intelligibility: a literature review. BBC. [Online]. Available: http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP190.pdf

[8] H. Fuchs and D. Oetting, "Advanced clean audio solution: Dialogue enhancement," in *IBC2013 Conference Proceedings*. Amsterdam, Netherlands: IET, Sept. 2013.

[9] J. Popp, M. Neuendorf, H. Fuchs, C. Forster, and A. Heuberger, "Recent advances in broadcast audio coding," in *Proc. 9th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*. London: IEEE, 2013, pp. 1–5.

[10] N. Hodoshima, "Effects of urgent speech and preceding sounds on speech intelligibility in noisy and reverberant environments," in *Proc. Interspeech 2016: 17th Annual Conf. of International Speech Communication Association*. San Francisco, U.S.A.: ISCA, 2016, pp. 1696–1699.

[11] R. Moreno and R. Mayer, "A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages." *Journal of Educational Psychology*, vol. 92, no. 1, p. 117, 2000.

[12] L. Ward, B. Shirley, Y. Tang, and W. Davies, "The effect of situation-specific acoustic cues on speech intelligibility in noise," in *Proc. Interspeech 2017: 18th Annual Conf. of International Speech Communication Association*. Stockholm, Sweden: ISCA, Aug. 2017, pp. 2958–2962.

[13] 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: IOA, Nov. 2016.

[14] D. Kalikow, K. Stevens, and L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1337–1351, 1977.

[15] R. Bilger, *Speech recognition test development, In: E. Elkins ed. Speech recognition by the hearing impaired*, 1984, vol. 14, pp. 2–15.

[16] 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*, vol. 58, no. 3, pp. 1093–1102, 2015.

[17] S. Sheldon, M. K. Pichora-Fuller, and B. A. Schneider, "Priming and sentence context support listening to noise-vocoded speech by younger and older adults," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 489–499, 2008.

[18] R. Wilson, R. McArdle, K. Watts, and S. Smith, "The revised speech perception in noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm," *Journal of American Academy of Audiology*, vol. 23, no. 8, pp. 590–605, 2012.

[19] BBC, "BBC Sound Effects Library CDs 1-60."

[20] Tera Media and CRG. Soundsnap.com. [Online]. Available: http://www.soundsnap.com/

[21] ITU Recommendation, "ITU-R BS. 1770-2, Algorithms to measure audio programme loudness and true-peak audio level," 2011.

[22] M. Cooke, "A glimpsing model of speech perception in noise," *Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1562–1573, 2006.

[23] 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. [Online]. Available: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf

[24] L. A. Ward, "Accessible broadcast audio personalisation for hard of hearing listeners," in *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. Hilversum, Netherlands: ACM, June 2017, pp. 105–108.

[25] J. B. Fritz, M. Elhilali, S. V. David, and S. A. Shamma, "Auditory attention—focusing the searchlight on sound," *Current Opinion in Neurobiology*, vol. 17, no. 4, pp. 437–455, 2007.

[26] A. L. Pelletier, L. Rojas-Roldan, and J. Coffin, "Vision loss in older adults." *American family physician*, vol. 94, no. 3, pp. 219–226, 2016.