SPEECH INTELLIGIBILITY FOR AUTISTIC PEOPLE IN MULTI-SPEAKER ENVIRONMENTS

Ryan Gamblin Buro Happold, UK
Benjamin West Birmingham City University, UK

1 INTRODUCTION

This study explores the speech-in-noise perception of autistic people, focusing on multi-speaker environments. It aims to utilise existing speech intelligibility and soundscape assessments to further knowledge about the auditory experience of a historically under-served group. This work is heavily supported and influenced by the emerging research area of aural diversity, an approach to music and sound research that aims to prioritise the wide range of hearing differences within the population, rather than an assumption of 'normal' hearing[1].

To understand the impact of common environments on speech intelligibility, this research utilises an auralisation approach. This uses Ambisonics recordings to create realistic reproductions of different environments, resulting in the ability to present a range of listening experiences in a way that is convenient and accessible for participants. These auralisations are then combined with speech recordings, the intelligibility of which can be measured with existing questionnaire methods, to understand how different auditory environments can impact speech comprehension for this group. The experiences of participants in these soundscapes are then assessed with data collection methods from PD ISO 12913-2:2018[2], part of the *Acoustics – Soundscape* series of standards.

It should be noted that participants anonymously gave their preference between person-first language (e.g. "person with autism") and identity-first language (e.g "autistic person"), with the vast majority stating either no preference (45.5%) or a preference for identity-first language (45.5%). This has informed the choice of language when discussing the study.

2 METHODOLOGY

2.1 Soundscape simulation

Rather than using binaural recordings over headphones, this study utilised environmental higher-order Ambisonics (HOA) recordings from the Ambisonics Recordings of Typical Environments database[3]. Four recordings were selected from the database, presented in Table 1 below.

	Soundscape	Description	SPL (dB)	SPL (dBA)	T_{30} (s)
1	Cafe	Indoor (company) café at medium occupancy before lunch, next to the wall.	71.7	66.2	1.1
2	Shopping centre	Very noisy food court in a shopping mall during lunch.	79.6	76.7	1.0
3	Dinner party	Small room with 8 people chatting over the table with background music.	72.8	68.7	0.4
4	Office	Open space office, people typing, chatting and talking on the phone.	56.7	51.4	0.2

Table 1: Details of the HOA recordings. Descriptions, SPLs and T30 measurements from Weisser et al[3].

The recordings were imported into Reaper and looped with crossfades until they were approximately three minutes long. This ensured that they were long enough to play under the speech tests, as well as leaving enough time at the end for participants to complete the soundscape assessments. These tracks were then normalised to -30 LUFS.

2.2 Presentation of speech

The Modified-Rhyme Test (MRT) was used to test speech intelligibility. The MRT presents listeners with a target word, with a multiple-choice set of six options to choose from. The multiple-choice options for the MRT provide variation between either the first or second consonant sound of the target word. For example, “led/shed/red/bed/fed/wed” for first consonant variation, or “bat/bad/back/bass/ban/bath” for second consonant variation.

To present the MRT to participants, recordings were selected for each soundscape from The Listening Talker project by the University of Edinburgh[4]. The speaker is a professional voice talent, with a Northern-influenced Received Pronunciation accent. The recordings were taken in a hemi-anechoic chamber with a Sennheiser MKH 800 P48 microphone, recorded at 48kHz. The list of multiple-choice options presented to participants were from the work of House et al.[5]

The selection of sentences from the dataset was random, with changes only made if there was not an equal split between first- and second-consonant variation. This ensured that every soundscape had 10 of each type of variation, for a total of 20 sentences per soundscape. The recordings were evenly spaced with two seconds between each track in Reaper. The recordings were then exported as one mono track for each soundscape, downsampled to 44.1kHz to match the sample rate of the HOA files, and 24-bit.

These files were then convolved with the binaural RIR for their environment (for example, speech track 1 was convolved with the RIR for Soundscape 1), using the MATLAB function *conv().* The convolved audio was then imported alongside the HOA tracks into a Reaper template, which had pre-routed outputs for presentation in the SoundLab® at STEAMHouse, Birmingham City University[6]. The speech tracks were placed on a separate mono track, output to the central speakers only, and the HOA tracks were placed on the main Ambisonics track.

2.3 Soundscape assessment

An adaptation of the sound source identification questions from PD ISO 12913:2-2018, Annex C.3.1.2[2], was chosen to better represent the range of sound sources in these environments: the office sound sources developed by Acun and Yilmazer[7] and utilised in West et al.[8] While the present study is not focused on office acoustics, it is similarly using indoor environments, and therefore these sound sources are more appropriate and useful for analysis. The source options presented to participants from this research are: Sound from human beings; mechanical and electronic sounds; outdoor sounds; and music. The Perceived Affective Quality (PAQ) rating scale is unchanged from PD ISO 12913:2-2018, Annex C.3.1.3.[2]

3 PARTICIPANT TESTING

Eleven autistic adults took part (age range = 19 to 23, median age = 21), all native English speakers. Participants were recruited via either posts on the Birmingham City University audio degrees message board, or word of mouth. Five identified as non-binary, 4 as male, and 2 as female. While noteworthy that almost half (45%) of participants identified outside the gender binary, this study does not seek to explore the occurrence of gender-diverse identities amongst autistic people.

Before taking part, participants received an information sheet detailing the test methodology. Participants were also reminded before beginning that they could take breaks or stop the test entirely at any time with no consequence, and that all data was being collected anonymously. No participants had a diagnosis or suspicion of a learning or cognitive disability, and therefore were able to provide informed consent. This project received ethics approval from Birmingham City University's Computing, Engineering and Built Environment Faculty Ethics Committee.

Participants were all either diagnosed or self-diagnosed as autistic, and reported no physical hearing loss. All testing was undertaken in the SoundLab® at STEAMHouse, Birmingham City University. This environment contains a third-order Ambisonics speaker array (a total of sixteen speakers). The participants were sat in the centre of the room (shown in Figure 1).

Figure 1: Photograph of the test setup.

The signal-to-noise ratio (SNR) between the speech and the soundscapes was -10dB; i.e., the normalised speech was 10dB lower than the normalised soundscapes. This aligns with the highest SNR used in Callejo et al[9], and the midpoint of the range in Irwin et al.[10], both studies into speech intelligibility for autistic people. The total level in the centre of the room was approximately 60LAeq, based on the levels of the two studies referred to for the SNR.

Test responses were recorded anonymously in Microsoft Forms, with the forms presented to participants on an iPad. Audio was presented from within Reaper, with numbered markers at the start of each soundscape.

To ensure that they understood the test procedure, participants were presented with a training test set, which presented them with six MRT questions, sound source identification and PAQ scales. Following this, participants were asked if they had been able to complete this test successfully, and if they had any questions they wanted to ask before beginning.

Each soundscape was assessed in the same manner. The first twenty questions were the multiple-choice MRT. This stage of the testing took approximately 2 minutes for each soundscape. Following this were the sound source identification and PAQ scales. Participants were initially given 1 minute with the soundscape playing (without the addition of speech) to fill out these assessments, with the option to request more time if needed.

Questionnaires were presented to participants on an iPad through Microsoft Forms. MRT questions were tagged as non-mandatory, enabling participants to leave questions blank if they felt they could not understand a word. Soundscapes questions were tagged as mandatory (i.e., the form could not be submitted until they were completed), to ensure that results were not skewed by human error, such as a person forgetting to select an option for one row. The MATLAB function randperm() was used to generate a random order of the tracks 1-4 for each participant, to control for the ordering effect.

After each soundscape, participants were prompted to consider if they would like to take a break, as reviews of the literature[11] have found some tendencies towards hyposensitivity – sometimes being less aware of one's own internal cues. Having this prompting enabled participants had the time to reflect on whether they felt close to being overwhelmed at any point, and reduced listener fatigue.

4 RESULTS

All statistical analysis was undertaken in MATLAB, aside from the ISO Pleasantness/Eventfulness plotting, which utilised the Python library Soundscapy.[12]

The impact of each soundscape was evaluated using one-way Analysis of Variance (ANOVA) testing, which showed that there was a statistically significant ($F(3, 40) = [6.87]$, $p = 0.0007$) variance between soundscapes. The results of post-hoc Tukey's HSD testing is shown in Table 2.

Group 1	Group 2	Mean difference	Mean diff (lower)	Mean diff (upper)	p
1	**2**	0.27273	-1.9752	2.5206	0.99
1	**3**	-1.5455	-3.7933	0.70243	0.27
1	**4**	-3.0909	-5.3388	-0.84302	0.0036 **
2	**3**	-1.8182	-4.0661	0.42971	0.15
2	**4**	-3.3636	-5.6115	-1.1157	0.0014 **
3	**4**	-1.5455	-3.7933	0.70243	0.27

Table 2: Tukey's HSD results for MRT score variation between soundscapes. ** $p < 0.01$.

The eight PAQ scores were transformed into the two dimensions of pleasantness *P* and eventfulness *E*, in accordance with Formulas 1 and 2 from PD ISO 12913-3:2019, Appendix A.3.[13] These ISO Pleasantness and Eventfulness scores were plotted to visualise the overall perception of each soundscape.

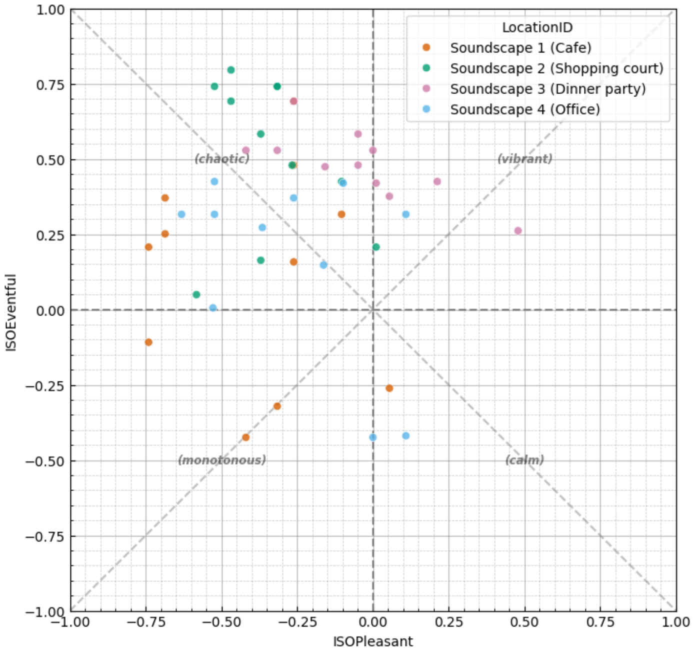

Figure 2: Two-dimensional PAQ scatter plots for each soundscape.

MRT score was evaluated to view the correlations between it and ISO Pleasantness, ISO Eventfulness, and the identified sound sources. Both ISO Pleasantness and ISO Eventfulness were evaluated for their relationships with the identified sound sources. Table 3 shows the ρ values of the Spearman's rank correlations.

	Pleasantness	Eventfulness	Human	Mechanical	Outdoor	Music
MRT score	0.12	- 0.3 *	0.10	-0.16	- 0.19	- 0.06
Pleasantness			0.07	-0.42 **	- 0.13	0.42 **
Eventfulness			0.35 *	-0.3 *	- 0.1	0.33 *

Table 3: Spearman's rank correlation ρ values. * $p < 0.05$, ** $p < 0.01$.

5 DISCUSSION

The moderate negative correlation between eventfulness and MRT score is of particular interest. One possibility is that this leads on from the theories of earlier work into speech intelligibility for autistic people, such as Alcántara et al. (2003)[14], which found indications that autistic people may be less able than allistic people to take advantage of spectral and temporal dips in noise when trying to understand speech. The implication of the present study is therefore that not only are autistic people less able to take advantage of more varied background sounds, they may in fact hindered by it.

Linked to this is research into autistic people's differences in processing. Some studies indicate that autistic people have slower processing speeds than their allistic counterparts,[15] and this may make eventful soundscapes more challenging, as the increased eventfulness results in more information to process at once. Similarly, hyperacusis is often explored as part of decreased sound tolerance (DST) in autism,[16] which results in auditory events being perceived as significantly louder than they actually are.

Additionally, a potential factor in this is the fact that research interviews[17] find autistic people believe they have more difficulty 'tuning out' background noise than allistic people. Therefore, if the surrounding auditory environment is more eventful, this then may have more of an impact on processing than one that is monotonous. If a person is already trying to cope with significant interference from background sounds, then increased variation and complexity in that sound may make it even more difficult, both as a distractor and by increasing listener fatigue.

Another factor worth noting here is that there was considerable overlap between the ISO Eventfulness of each soundscape, as well as a wide range in responses between participants. This is illustrated in Figure 3.

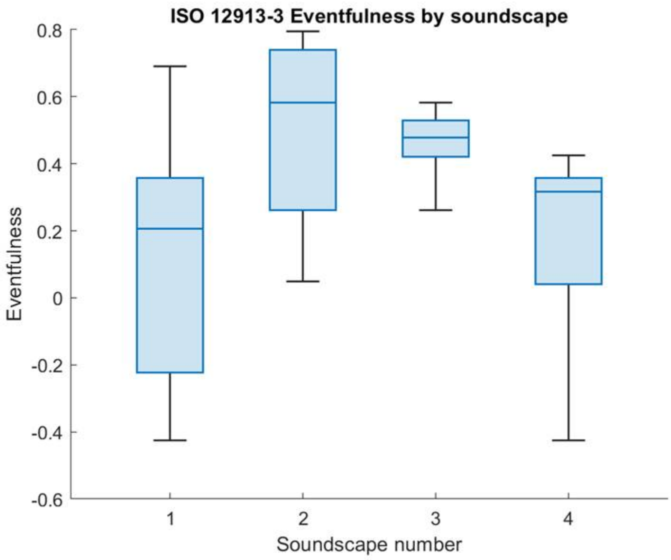

Figure 3: Box plot of ISO Eventfulness by soundscape, excluding outliers (calculated using interquartile range).

This implies that there are factors that may considerably alter how eventful a soundscape is perceived as by each person that uses it. This may be linked to differing experiences of DST. For example, some participants may have much more exacerbated hyperacusis than others. This could then result in them perceiving auditory events as more prominent, and therefore providing more extreme ratings. Similarly, a participant who experiences a misophonia response to certain sounds may be more aware of these if they occur.

While ISO Eventfulness appeared to have some correlation with speech intelligibility, ISO Pleasantness did not. This is contrary to how many participants perceived their performance, as verbal feedback generally indicated that the soundscapes they "hated" (as one participant stated about Soundscape 1) were also the ones that they perceived to be the most difficult.

For example, it was common for participants to state that they found the increased mechanical noises in Soundscapes 1 and 4 made them more difficult. This correlates with the pleasantness scores calculated from the questionnaires, and also further shows that the perceived pleasantness of the environments does not necessarily correlate with overall intelligibility; Soundscape 4 had a notably *higher* MRT score than several of the other environments.

Regardless, the importance of pleasantness in the auditory environment should not be understated due to these findings. It is worth reiterating that these conditions were purposely selected based on existing findings that autistic people find multi-speaker environments to be difficult experiences; indeed, the median score for all soundscapes in the present study was negative (Figure 4). Therefore, this relative lack of variation means that an underlying correlation between intelligibility and pleasantness may have not been uncovered.

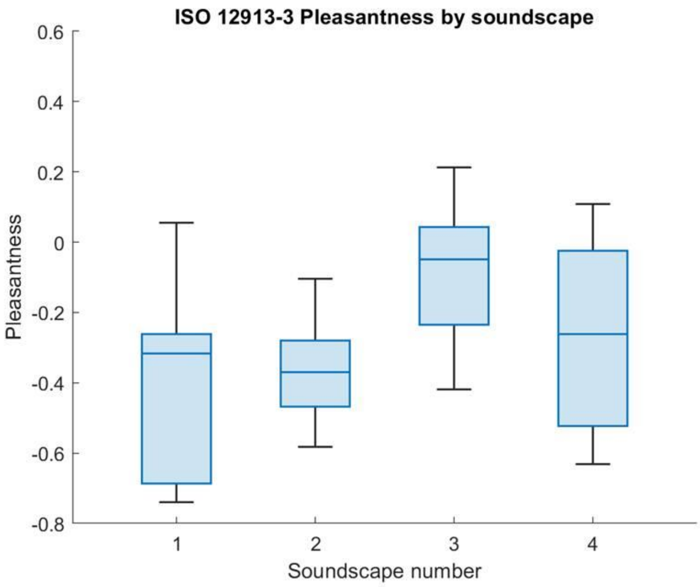

Figure 4: Box plot of ISO Eventfulness by soundscape, excluding outliers (calculated using interquartile range).

Perhaps more importantly, however, if future work were to affirm that there is no link between pleasantness and speech intelligibility for autistic people, it should not be dismissed when autistic people find an auditory environment to be unpleasant. Speech intelligibility is only one aspect of what makes a space usable; the fact that a soundscape is experienced as difficult or distressing is significant in itself, and should still inform a holistic understanding of whether an environment is accessible.

6 CONCLUSION

The contribution of ISO Eventfulness to decreased speech intelligibility is the key finding from this research, and goes some way to informing how soundscape design can be made more accessible for autistic people. Further work is needed to corroborate these results, with a larger sample size and a wider range of environments. Additionally, future work should explore environments with more variation across pleasantness-unpleasantness, to better understand the impact of this dimension.

This work has demonstrated how an Ambisonics environment can be utilised to present more immersive listening studies to participants, whilst still making use of previously validated methodologies for assessment. Assessment procedures were in accordance with ISO standards where relevant, and produced several findings that were statistically significant enough to warrant discussion.

Alongside the data-based conclusions, the outlook and findings of this project can be seen as helping to continue the changing tone of autism research, by demonstrating that work led and informed by autistic people is an important step in producing a grounded understanding of both autistic listening and autism more broadly.

7 REFERENCES

1. A. Hugill and J. L. Drever, Eds., Aural Diversity. New York: Routledge, 2022.
2. British Standards Institute, PD ISO/TS 12913-2:2018 Acoustics - Soundscape. Part 2: Data collection and reporting requirements (2018).
3. A. Weisser et al., 'The Ambisonic Recordings of Typical Environments (ARTE) Database', Acta Acust. united Acust. 105(4), 695–713 (2019).
4. C. Valentini-Botinhao, C. Mayo, and M. Cooke, 'Hurricane natural speech corpus - higher quality version'. Available: [10.7488/ds/2482](10.7488/ds/2482) (2019).
5. A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, 'Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set', J. Acoust. Soc. Am. 37(1), 158–166 (1965).
6. STEAMHouse, 'Digital Labs'. Available: https://steamhouse.org.uk/home/buildingandequipment/digital-labs/ (2024).
7. V. Acun and S. Yilmazer, 'A grounded theory approach to investigate the perceived soundscape of open-plan offices', Applied Acoustics, vol. 131, 28–37 (2018).
8. B. West, A. Deuchars, and I. Ali-MacLachlan. 'Office Soundscape Assessment: A Model of Acoustic Environment Perception in Open-Plan Offices'. J. Acoust. Soc. Am. 156(5), 2949–59 (2024).
9. D. Callejo, J. Wouters, and B. Boets. 'Speech-in-Noise Perception in Autistic Adolescents with and without Early Language Delay'. Autism Research 16(9), 1719–27 (2023).
10. J.R. Irwin, L.A. Tornatore, L. Brancazio, and D. H. Whalen. 'Can Children With Autism Spectrum Disorders "Hear" a Speaking Face?' Child Development 82(5), 1397–403 (2011).
11. D. DuBois, S. H. Ameis, M.-C. Lai, M. F. Casanova, and P. Desarkar, 'Interoception in Autism Spectrum Disorder: A review', Int. J. Dev. Neurosci. vol. 52, 104–111 (2016).
12. A. Mitchell, F. Aletta, and J. Kang, 'How to analyse and represent quantitative soundscape data', JASA Express Letters. 2(3), 037201 (2022).
13. British Standards Institute, PD ISO/TS 12913-3:2019 Acoustics. Soundscape. Part 3: Data analysis (2019).
14. José I. Alcántara, Emma J.L. Weisblatt, Brian C.J. Moore, and Patrick F. Bolton. 'Speech-in-Noise Perception in High-Functioning Individuals with Autism or Asperger's Syndrome'. J. Child Psychol. Psychiatr., 45(6), 1107–1114 (2004).
15. S. M. Haigh, J. A. Walsh, C. A. Mazefsky, N. J. Minshew, and S. M. Eack, 'Processing Speed is Impaired in Adults with Autism Spectrum Disorder, and Relates to Social Communication Abilities', J Autism Dev Disord, vol. 48, 2653–2662 (2018).
16. Z. J. Williams, J. L. He, C. J. Cascio, and T. G. Woynaroski, 'A review of decreased sound tolerance in autism: Definitions, phenomenology, and potential mechanisms', Neurosci. Biobehav. Rev. vol. 121, 1–17 (2021).
17. A. Sturrock, H. Guest, G. Hanks, G. Bendo, C. J. Plack, and E. Gowen, 'Chasing the conversation: Autistic experiences of speech perception', Autism & Developmental Language Impairments. vol. 7, (2022).