frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 27 November 2025
DOI 10.3389/frvir.2025.1691731

OPEN ACCESS

EDITED BY
Xianmin Wang,
Guangzhou University, China

REVIEWED BY
Zahra Jeddi,
Shiraz University of Medical Sciences,
Shiraz, Iran
Halide Cetin Kara,
Istanbul University Cerrahpasa, Türkiye

*CORRESPONDENCE
Mali Harshada,
harshada17.aiish@gmail.com
K. V. Nisha,
nishakv@aiishmysore.in

RECEIVED 24 August 2025
REVISED 30 September 2025
ACCEPTED 08 October 2025
PUBLISHED 27 November 2025
CORRECTED 03 December 2025

CITATION
Harshada M, Pitchaimuthu AN and Nisha KV (2025) Midlife challenges in speech perception in spatial noise under virtual reverberant environments.
Front. Virtual Real. 6:1691731.
doi: 10.3389/frvir.2025.1691731

Midlife challenges in speech perception in spatial noise under virtual reverberant environments

Mali Harshada*, Arivudai Nambi Pitchaimuthu and K. V. Nisha*

Centre for Hearing Sciences, Department of Audiology, All India Institute of Speech and Hearing, Mysore, India

Introduction: Speech recognition in noisy, reverberant environments is challenging, particularly with aging. Subtle spatial auditory deficits emerging in midlife may precede measurable hearing loss and impair communication. Real-world studies face challenges in control and replication, whereas virtual reality (VR) simulations offer an alternative. This study examines how age and noise location influence speech recognition in virtual reverberant environments.

Methods: Sixty normal-hearing adults participated: 30 young (18–40 years, M = 25.19, SD = 5.23) and 30 middle-aged (41–60 years, M = 55.79, SD = 4.57). Participants completed sentence recognition tasks in virtual acoustic simulations with three reverberation levels (anechoic, short: 0.8 s, long: 3.0 s) and three noise locations (0°, 60° right, 60° left). Sentences were presented at 0° amidst spatial noise. A generalized linear mixed model (GLMM) analyzed sentence recognition scores, with fixed effects for age, reverberation, and noise location, and random effects for participant variability.

Results: GLMM results showed middle-aged adults had poorer sentence recognition than young adults ($p < 0.05$). Both groups exhibited SRM in anechoic and short reverberation conditions, but middle-aged adults showed no spatial release from masking in long reverberation. Significant age-reverberation interactions indicated greater deficits in middle-aged adults under challenging acoustics.

Discussion: Findings suggest that middle-aged adults may experience subtle speech perception difficulties in noisy and reverberant environments, even with clinically normal hearing. However, generalization to hearing-impaired populations remains limited.

KEYWORDS

spatial release from masking, reverberation, speech perception, age, virtual reality

1 Introduction

Understanding speech in real-world listening environments laden with background noise, such as classrooms, offices, and busy public places, is a complex auditory task that requires listeners to extract meaningful information from a mixture of target and background sounds (Bronkh et al., 1990). Parsing speech from background noise relies on using a combination of acoustic cues and spatial cues: temporal envelope fluctuations, voice pitch differences, and the onset/offset timing of sounds along with interaural time and level differences of sound sources (Blauert, 1996; Bregman and McAdams, 1994; Bronkhorst, 2015; Middlebrooks and Green, 1991; Shamma et al., 2011). These cues enable the auditory system to segregate the target speech from competing sources. In configurations where the target and noise are presented from the same location (typically front-facing at 0° azimuth), listeners often experience greater difficulty due to overlapping binaural cues and energetic masking (Viswanathan et al., 2016). Spatial separation between the speech and noise sources, i.e., the angular difference between the target and noise, can facilitate speech recognition, which is commonly termed as spatial release from masking (SRM) (Zenke and Rosen, 2022). This benefit from spatial separation declines with age due to altered neural encoding of interaural phase cues, and this further results in diminished speech-in-noise performance (Grose et al., 2019; Grose and Mamo, 2010).

In a closed space, reverberation tends to further degrade speech intelligibility by smearing temporal and spectral cues (Nam and Park, 2025). Reverberation introduces a trailing "echo" of the sound owing to reflections from the surfaces. This overlap in time masks phonemic transitions and disrupts speech cues that are critical for intelligibility, in addition to the presence of a masker (Nabelek and Pickett, 1974; Steeneken and Houtgast, 1980). Viveros Muñoz et al. (2019) found that SRM persists in young (mean: 24.3 years) and older adults (mean: 69.5 years), indicating robust binaural unmasking, but it is compromised as the reverberation increases. These effects are well-documented in younger adults (Deroche et al., 2017); however, middle-aged listeners have also shown diminished sensitivity to temporal fine structure and interaural phase cues (Grose and Mamo, 2010), impaired speech-in-noise performance in collocated conditions, and reduced SRM under reverberant simulations (Marrone et al., 2008). Srinivasan et al. (2017) reported that while early reflections are typically beneficial for speech perception, their effectiveness diminishes with age, indicating deficits in spatial–temporal integration. Despite emerging evidence of early auditory–cognitive changes in spatial and temporal resolution, the literature on middle-aged adults remains limited. This is particularly notable considering that nearly two-thirds of India's population is of working age (15–64 years), representing a broad span from young to middle-aged adulthood. Given their demographic weight, the communication needs of middle-aged adults must be prioritized and systematically addressed. This knowledge gap is hence critical as spatial hearing decline is expected to begin in midlife, coinciding with reductions in temporal processing, working memory, and attentional control, all of which can exacerbate difficulties in complex listening environments (Anderson et al., 2013). Because the effects of reverberation and spatial segregation on speech perception are challenging to study under controlled conditions, immersive virtual reality simulation can offer a valuable approach by bridging the gap between laboratory-based acoustics and real-world listening complexity (Doggett et al., 2021; Serafin et al., 2023). Hence, in the present study, we compared speech recognition in young and middle-aged adults under three noise locations (collocated, right-only, and left-only) and reverberation conditions (anechoic, short, and long) to examine early-onset processing declines in midlife before age-related changes appear on the audiogram.

2 Materials and methods

2.1 Participants

The study employed a mixed design comprising both between-subject young and middle-aged adult groups and within-subject [three reverberation levels: anechoic, short, long; three noise locations: collocated at 0°, segregated at left 60° (L60°), and segregated at right 60° (R60°)] factors to investigate speech recognition scores (SRSs).

The sample size was determined using the R package pwr (Champely, 2020) for a two-sample t-test comparing SRSs between young and middle-aged groups under anechoic conditions, with a significance level of $\alpha = 0.05$ and power of 0.8. Based on pilot data, the estimated Cohen's d was 0.851 (young-adult mean SRS = 15.7, middle-aged adult mean SRS = 14.2, and pooled SD = 2.21), yielding a required sample size of approximately 23 participants per group. The 95% CI for the mean SRS difference is [0.35, 2.65]. A total of 30 participants per group were recruited to ensure sufficient power, accounting for potential within-subject variability and multiple comparisons.

Sixty individuals with normal hearing comprising 30 young adults (12 male, 18 female; age range: 18 years–40 years, M = 25.19, SD = 5.23) and 30 middle-aged adults (19 male, 11 female; age range: 41 years–60 years, M = 49.79, SD = 4.57) participated in the study. Participants with pure-tone audiometric thresholds ≤15 dB HL at 250 Hz–8,000 Hz were included. To confirm symmetrical hearing across ears, the thresholds of all 60 participants in the audiometric frequency range (250–8,000 Hz) were subjected to mixed analysis of variances (ANOVA). This revealed no significant main effects [for ears: $F\,(1.58) = 0.12$, $p = 0.73$; for frequencies: $F\,(7.406) = 1.45$, $p = 0.19$] or interactions effects between ears and frequencies [$F\,(7.406) = 0.89$, $p = 0.51$]. Participants with otological complaints and hearing loss were not included in the study. The Neuropsychological Evaluation Screening Test (NEST) (Chopra et al., 2018) and Screening checklist for Central Auditory Processing—Adults (SCAP-A) (Vaidyanath and Yathiraj, 2014) were administered to screen for normal cognition and auditory processing, respectively. Participants who were native speakers of the Kannada language and had received formal education in Kannada at least through secondary school and those who scored ≥7 on the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) were included in the study. Participants who scored less than 2 on NEST and less than 3 on SCAP (A) and those with no history of neurological or otological disorders were recruited in the study. Demographic, audiometric, and LEAP-Q assessment details of both groups (young and middle-aged adults) are provided in Table 1. Eligible individuals meeting the inclusion criteria were identified and invited to participate through non-random purposive sampling from individuals who had visited, studied, or worked at our institute. Ethical committee approval was obtained from the Institutional Review Board (no. SH/IRB/M.1-20-2024-25), and all the participants provided written informed consent prior to inclusion in the study.

2.2 Stimuli

Reverberation and spatial separation were simulated using Audio 3D, a Windows application (available at: https://www.phon.ucl.ac.uk/resource/audio3d/) that enables users to construct dynamic virtual auditory environments. It combines real-time binaural processing with customizable head-related transfer

TABLE 1 Demographic and auditory characteristics of group I (young adults) and group II (middle-aged adults).

Characteristic		Group I (young adults) n = 30	Group II (middle-aged adults) n = 30
Sex (M/F)		(18/12)	(11/19)
Age (years)	Mean	25.19	49.79
	SD	5.23	4.57
	Range	18–40	41–60
PTA left (dB HL)	Mean	11.2	14.5
	SD	3.1	4.5
	Range	5–15	10–15
PTA right (dB HL)	Mean	10.9	12.5
	SD	2.9	2.5
	Range	5–15	10–15
SIS (%)	Mean	99	98
	SD	0.8	1.8
	Range	97–100	94–100
NEST score	Mean	1.2	1.5
	SD	0.6	0.5
	Range	0–2	0–2
SCAP (A) score	Mean	0	0
	SD	0	0
	Range	0	0
LEAP-Q score	Mean	9.03	9
	SD	0.7	0.3
	Range	8–10	8–10

The table summarizes the distribution of sex, age, and pure-tone average (PTA) thresholds for the left and right ears (in dB HL), speech identification scores (SIS, in %), NEST scores, SCAP (A), and LEAP-Q scores across the two groups. Values are presented as the mean, standard deviation (SD), and range.
NEST stands for Neuropsychological Evaluation Screening Test; SCAP-A stands for Screening checklist for Central Auditory Processing–Adults; LEAP-Q stands for Language Experience and Proficiency Questionnaire.

TABLE 2 Comparison of seven models based on the degrees of freedom (df), Akaike information criterion (AIC), and Bayesian information criterion (BIC).

Model	df	AIC	BIC
Model 1	21	−1,276.679	−1,186.948
Model 2	22	−1,279.681	−1,185.678
Model 3	22	−1,290.341	−1,196.337
Model 4	25	−1,279.842	−1,173.020
Model 5	25	−1,347.655	−1,240.833
Model 6	28	−1,290.727	−1,171.086
Model 7	37	−1,347.462	−1,189.365

Lower values of AIC and BIC indicate better model fit, with model 5 and model 7 showing the lowest AIC and BIC values, respectively.

function (HRTF) data and allows recording of the combined simulated effects of the direct path, first reflections, and early and late reverberation, along with the settings for room reverberations at each sound source. Room dimensions of 4.5 m width, 7 m depth, and 3 m height were utilized for the simulation (Opochinsky et al., 2025; Tillery et al., 2013). Three reverberation conditions were selected for the study based on the reverberation time in seconds (RT60): (1) *anechoic* condition (RT60 = 0 s); (2) short reverberation (RT60 = 0.8 s); (3) long

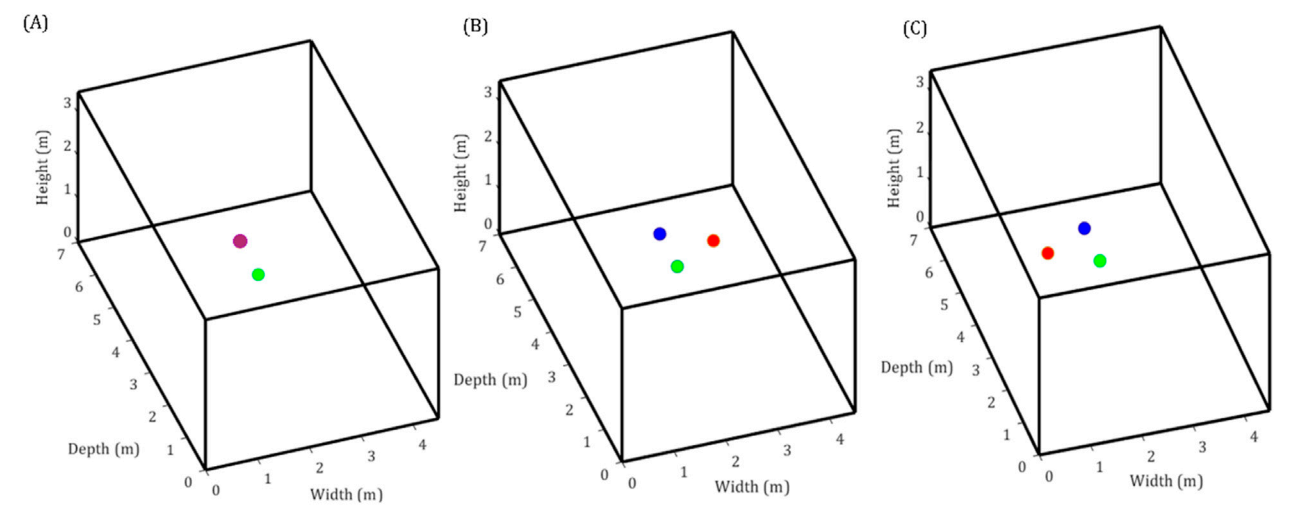

FIGURE 1
Three-dimensional representation of the simulated room setup. The green dot indicates the listener position, the blue dot indicates the target speech source, and the red dot indicates the noise source. **(A)** Collocated condition (speech and noise at 0°). **(B)** R60° condition (noise at right 60°) and **(C)** L60° condition (noise at left 60°).

reverberation (RT60 = 3.0 s). The reverberation time (RT60) was estimated using Sabine's formula (Kuttruff, 2016), and the wall, floor, and ceiling reflection coefficients were derived accordingly from the calculated absorption values. These settings were applied uniformly across all surfaces under the assumption of equal reflectivity. The binaural room impulse responses were calculated using the image-source method (Allen and Berkley, 1979), which simulates the effect of multiple room reflections based on the geometrical relationship between the source and listener positions, room dimensions, and surface reflectivity. The software independently models the direct path, early reflections, and late reverberation. Early reflections were explicitly calculated from six first-order image sources (walls, ceiling, and floor), while late reverberation was derived from residual impulse responses with the direct and early components removed.

The HRTFs were based on the Center for Image Processing and Integrated Computing database and implemented via a principal components analysis approach to enable efficient binaural rendering (Algazi et al., 2001). All sources were rendered with full directional filtering for spatial realism at the listener's position. The listener's position was maintained in the center of the room at a height of 1.2 m. The target location and masker location were 1 m away from the listener, as shown in Figure 1.

A total of 45 low-predictability Kannada sentences (Geetha et al., 2014) that are characterized by minimal semantic constraints on the final word were used in the experiment. These were distributed across three spatial conditions with an 8-talker Kannada babble: (referred as noise in this article hereafter) 15 sentences in the co-located condition, 15 sentences with a masker at 60° left (L60°), and 15 sentences with a masker at 60° right (R60°). Within each of these spatial conditions, five (20 keywords) sentences were simulated under each reverberation condition (anechoic, short, and long) and were saved offline as .wav files. The sentence stimuli were presented at 65 dB SPL, while the competing 8-talker babble (henceforth referred as noise) was delivered at 70 dB SPL. Figure 1 shows the simulated room setup, including the positions of the listener, target, and noise sources, while Figure 2 presents the corresponding spectrograms and energy decay curves of the signals under each reverberation condition.

2.3 Stimulus presentation and response recording

A custom MATLAB script randomized the order of sentence presentation across three reverberation levels and conditions (anechoic, short, and long) and three noise locations collocated, R60° and L60° Stimuli were delivered through calibrated closed-back headphones (Sennheiser HD 569, Wedemark, Germany). After each sentence was presented, participants were prompted to verbally repeat the sentence, and their responses were recorded and saved as individual .wav files. An Excel file was used to log the sequence of presented stimuli, allowing us to correlate the stimulus condition with the participant's responses involved in the subsequent scoring and analysis phase.

2.4 Scoring procedure

Each sentence consisted of four keywords. The responses were manually scored. Each correctly identified keyword was awarded 1 point, resulting in a maximum score of 4 points per sentence. Incorrect or omitted keywords received a score of 0. The SRSs across conditions (reverberation and noise location) were computed. With a total of nine conditions comprising 45 sentences, each containing four keywords, the maximum possible composite score was 180.

2.5 Statistical analyses

The Shapiro–Wilk test of normality was conducted using R software (version 4.5.1) with the "stats" package (R Core Team, 2025).

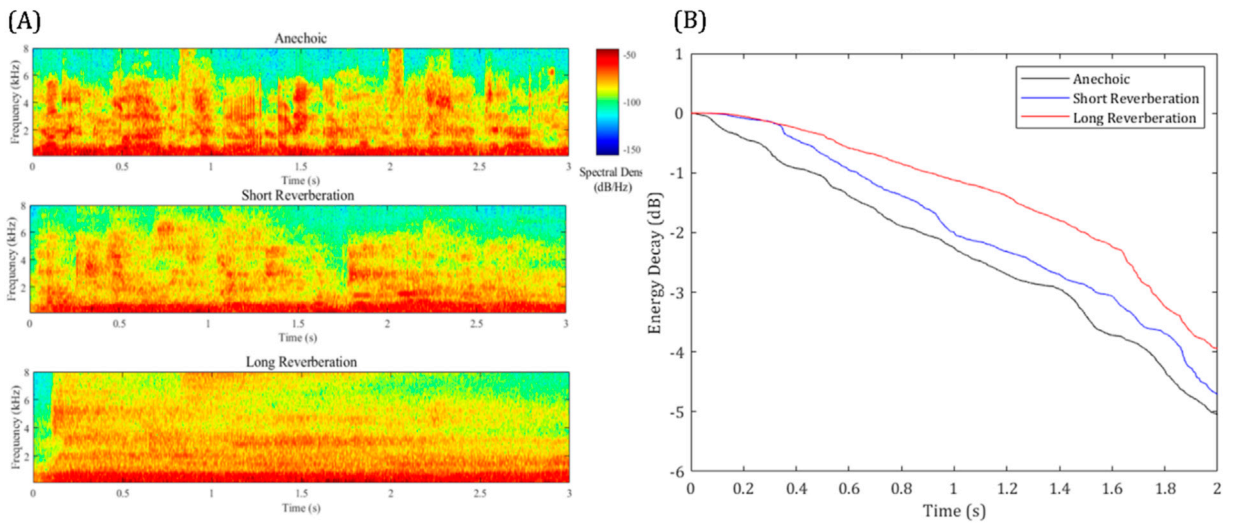

FIGURE 2
(A) Spectrograms for anechoic, short reverberation, and long reverberation environments, with frequency (Hz) on the y-axis and time (s) on the x-axis using a color scale to represent the spectral density (dB). **(B)** Energy decay curves for anechoic (black), short reverberation (blue), and long reverberation (red) environments, with energy decay (dB) on the y-axis and time (s) on the x-axis, showing how sound energy dissipates over time.

The distribution of SRS (n = 540; range: 2–20) was assessed using the Shapiro–Wilk test, confirming a non-normal distribution ($p < 0.05$). SRSs were normalized to proportions (SRS/20, range: 0.1–1.0) and adjusted to the open interval (0, 1) using (SRS/20 × (n-1) + 0.5)/n, where n = 20 (the maximum achievable score per condition), to fit a Beta distribution (Smithson and Verkuilen, 2006). The probability density function of normalized SRSs revealed a multimodal, bounded distribution. A generalized linear mixed model with a Beta distribution and logit link was fitted using the glmmTMB package in R (v4.5.1):

$$y \sim \text{Beta}(\mu, \varphi), \text{where logit}(\mu) = \log(\mu/(1\text{-}\mu)),$$

where y is the normalized speech recognition score, μ is its mean, and ϕ is the precision (inverse dispersion) parameter. Diagnostics for hierarchical (multi-level/mixed) regression models using residual analysis were performed to ensure model fit (Hartig, 2024). The model included fixed effects of the group, location, reverberation, and all interactions, with a random intercept for Subject_ID (n = 60) to account for individual variability: SRS ~ group × noise location × reverberation + (1|Subject_ID). This model was further subjected to type-II Wald chi-square tests for interpretation of the fixed effects (car package). *Post hoc* pairwise comparisons were conducted on the estimated marginal means with Bonferroni-adjusted t-values to control for multiple comparisons across the three noise locations and reverberation conditions.

3 Results

Altogether, with 30 participants in each of the two groups and nine conditions per participant (three reverberation × three spatial locations), a total of 540 data points were obtained. The distribution of normalized SRSs exhibited multimodality and bounded characteristics. Initial DHARMa diagnostics revealed issues with the model fit. DHARMa showed non-uniform residuals (Kolmogorov–Smirnov test: D = 0.069, p = 0.011) and under-dispersion (dispersion = 0.803, p = 0.008). Levene's test further confirmed heteroscedasticity [F (18, 521) = 7.287, $p < 0.001$]. To address these issues, 10 extreme outliers (1.85%) with scaled residuals <0.005 or >0.995 were removed, yielding a cleaned dataset (n = 530).

Seven models with varying dispersion structures were compared using Akaike information criterion (AIC) and Bayesian information criterion (BIC): variance modeled by (1) group; (2) noise location; (3) reverberation; (4) group + noise location; (5) group + reverberation; (6) noise location + reverberation; (7) group + noise location + reverberation. Model 7, with a dispersion submodel for the three-way interaction, provided the best fit (AIC = −1,347.5, BIC = −1,189.4). Post-adjustment DHARMa diagnostics on Model 7 showed uniform residuals (D = 0.046, p = 0.211) and no significant dispersion issues (dispersion = 0.876, p = 0.152), confirming adequate model fit, as shown in Table 2. However, heteroscedasticity persisted [Levene's test: F (18, 511) = 7.629, $p < 0.001$], suggesting residual variance differences across groups that may require further exploration in future analyses.

Figure 3 illustrates the mean SRS patterns, showing the highest scores in anechoic followed by short reverberation conditions and the lowest scores in long reverberation condition. Young adults outperformed middle-aged adults across all conditions, and spatially separated maskers (L60° and R60°) yielded better performance than co-located maskers (0°).

3.1 Effect of group

Type-II Wald chi-square tests (ANOVA, car package) revealed significant main effects of *group* [χ^2 (1) = 41.484, $p < 0.001$]. Pairwise

TABLE 3 Pairwise comparisons of speech recognition performance between young and middle-aged adults across reverberation times and noise locations.

Reverberation	Location of the noise	Group contrast	Odds ratio	z	p
Anechoic	0°	Young vs. Middle-aged	0.688	−2.596**	0.009
	L60°		0.515	−3.226***	0.001
	R60°		0.434	−3.089**	0.002
Short	0°		0.579	−5.710***	<0.001
	L60°		0.495	−6.076***	<0.001
	R60°		0.324	−7.847***	<0.001
Long	0°		1.161	1.170	0.242
	L60°		0.829	−1.487	0.1371
	R60°		0.666	−3.030**	0.002

Odds ratios (OR) < 1 indicate better performance by young adults, whereas OR > 1 indicate better performance by middle-aged adults ** is less $p < 0.05$, whereas *** is $p < 0.001$.

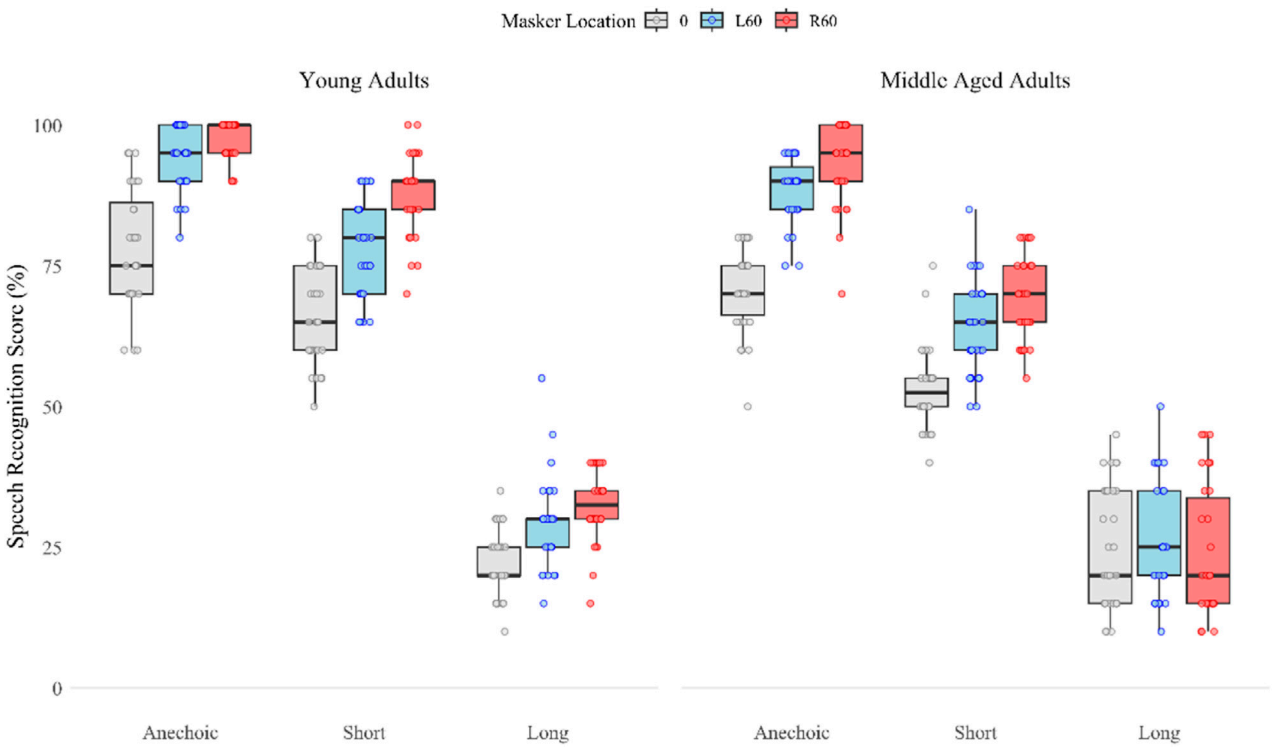

FIGURE 3
Boxplots showing speech recognition scores (%) across reverberation conditions (anechoic, long, and short) for two age groups—young adults (left panel) and middle-aged adults (right panel). Each reverberation condition is further divided by masker location: collocated (0°, gray), spatially separated to the left (L60°, blue), and right (R60°, red). Spatial separation improved the scores across both age groups, with young adults exhibiting higher overall performance and greater resilience to reverberation compared to middle-aged adults. Data shown are after the removal of statistical outliers (identified using scaled residuals from the beta regression model; cutoff <0.005 or >0.995).

comparisons of the SRSs across the conditions between the two groups with Bonferroni-adjusted p-value are given in Table 3. Young adults outperformed middle-aged adults in all combinations of levels of noise location and reverberation, except in the long reverberation condition where the distracter is located at L60° and 0°. On the other hand, young adults performed better under the R60° condition compared to middle-aged adults.

The *Group* × *Noise location* interaction [χ^2 (2) = 29.731, $p < 0.001$] revealed that young adults benefited more from spatial separation (L60° and R60°) than middle-aged adults, particularly in anechoic and short reverberation conditions. The *Group* × *Reverberation* interaction was significant [χ^2 (2) = 56.920, $p < 0.001$], wherein the SRSs of young adults was most pronounced in the short reverberation condition. The three-way interaction

TABLE 4 Pairwise comparisons of speech recognition performance across spatial conditions.

Reverberation	Group	Noise location contrast	Odds ratio	z	p
Anechoic	Young adults	0° vs. L60°	0.249	−6.815***	<0.001
		0° vs. R60°	0.125	−9.57***	<0.001
		L60° vs. R60°	0.501	−2.714**	0.02
	Middle-aged adults	0° vs. L60°	0.333	−9.531***	<0.001
		0° vs. R60°	0.198	−8.284***	<0.001
		L60° vs. R60°	0.594	−2.54**	0.033
Short	Young adults	0° vs. L60°	0.541	−5.35***	<0.001
		0° vs. R60°	0.282	−8.949***	<0.001
		L60° vs. R60°	0.521	−4.232***	<0.001
	Middle-aged adults	0° vs. L60°	0.634	−12.586***	<0.001
		0° vs. R60°	0.505	−16.632***	<0.001
		L60° vs. R60°	0.796	−4.866***	<0.001
Long	Young adults	0° vs. L60°	0.682	−5.623***	<0.001
		0° vs. R60°	0.596	−9.803***	<0.001
		L60° vs. R60°	0.873	−2.005	0.134
	Middle-aged adults	0° vs. L60°	0.955	−0.328	1.00
		0° vs. R60°	1.039	−0.249	1.00
		L60° vs. R60°	1.087	0.574	1.00

Odds ratios (OR) across distracter locations (0° = co-located, L60° = left 60°, and R60° = right 60°). For each contrast pair (e.g., 0° vs. L60°), OR > 1 indicates better performance for the first condition (e.g., 0°) than for the second (e.g., L60°), while OR < 1 indicates better performance for the second condition than for the first; *p*-values are Bonferroni-corrected, with non-significant results ($p \geq 0.05$) adjusted to 1 ** is less $p < 0.05$, whereas *** is $p < 0.001$.

[*Group* × *Reverberation* × *Noise location*] was not statistically significant [χ^2 (4) = 1.307, $p > 0.05$].

3.2 Effect of noise location

Type-II Wald chi-square tests also revealed that noise location significantly influenced SRSs [χ^2 (2) = 637.315, $p < 0.001$], with Noise location (L60° and R60°) improving the performance compared to collocated noise. Pairwise comparisons of SRSs across spatial conditions are given in Table 4. For young adults, the noise location benefited speech recognition performance, which was evident across all reverberation conditions. For middle-aged adults, noise location had no effect in the long reverberation condition (0° vs. R60°: OR = 1.039, p = 1.000) but was significant in anechoic and short reverberation conditions. Significant *noise location* × *Reverberation* interaction effect [χ^2 (4) = 140.731, $p < 0.001$] was also seen; with anechoic and short reverberation conditions, all noise–location pairs differed significantly for both groups. Under long reverberation conditions, young adults showed significant differences for 0° versus L60° and 0° versus R60°, whereas middle-aged adults did not show significant differences between noise locations (see Figure 3).

3.3 Effect of reverberation

Type-II Wald chi-square tests showed that reverberation had a statistically significant effect on SRSs [χ^2 (2) = 3402.987, $p < 0.001$]. Pairwise comparisons confirmed poorer performance in long reverberation conditions compared to anechoic and short reverberation conditions across all groups and noise locations, as shown in Table 5. The *Group* × *Reverberation* [χ^2 (2) = 56.920, $p <$ 0.001] interaction showed that the difference in the young adults' performance and that of middle-aged adults diminished in long reverberation conditions.

4 Discussion

The findings provide insights into speech recognition in spatial noise between young adults and middle-aged adults in reverberant environments. By examining SRSs across varying noise locations and reverberation conditions, the study reveals distinct patterns of auditory performance, particularly how young adults consistently outperform middle-aged adults in less reverberant environments and spatially separated conditions, while these advantages were diminished under long reverberation conditions.

TABLE 5 Pairwise comparisons of speech recognition performance across reverberation conditions.

Location	Group	Reverberation time contrast	Odds ratio	z	*p*
0°	Young adults	Long vs. Anechoic	0.081	−21.811	<0.001
		Long vs. Short	0.146	−24.555	<0.001
		Anechoic vs. Short	1.81	4.611	<0.001
	Middle-aged adults	Long vs. Anechoic	0.136	−16.046	<0.001
		Long vs. Short	0.292	−11.668	<0.001
		Anechoic vs. Short	2.152	10.69	<0.001
L60°	Young adults	Long vs. Anechoic	0.029	−19.449	<0.001
		Long vs. Short	0.116	−19.956	<0.001
		Anechoic vs. Short	3.931	7.012	<0.001
	Middle-aged adults	Long vs. Anechoic	0.047	−23.202	<0.001
		Long vs. Short	0.194	−16.722	<0.001
		Anechoic vs. Short	4.095	14.485	<0.001
R60°	Young adults	Long vs. Anechoic	0.017	−21.286	<0.001
		Long vs. Short	0.069	−20.699	<0.001
		Anechoic vs. Short	4.089	6.254	<0.001
	Middle-aged adults	Long vs. Anechoic	0.026	−17.014	<0.001
		Long vs. Short	0.142	−16.541	<0.001
		Anechoic vs. Short	5.485	9.125	<0.001

Odds ratios (ORs) across reverberation. For each contrast pair (e.g., long vs. Anechoic), OR < 1 indicates better performance for the second condition (anechoic) than for the first (long), while OR > 1 indicates better performance for the first condition than for the second; *p*-values are Bonferroni-corrected, with non-significant results ($p \geq 0.05$) adjusted to 1.

4.1 Effect of age on speech recognition in reverberant environments

Young adults outperformed middle-aged adults in anechoic and short reverberation conditions (see Table 3). Younger adults possess more robust temporal and spectral resolution, which are critical for parsing speech in complex acoustic environments (Pichora-Fuller and Souza, 2003). A recent review by Windle et al. (2023) reported a modest decline in binaural processing efficiency in middle-aged adults compared to young adults, which was attributed to early age-related declines in auditory nerve function and central auditory processing. Grose et al. (2019) suggested that as individuals age, the brainstem's response time may become slow, potentially impacting how temporal information is processed. Similarly, Grose and Mamo (2012) proposed that a decline in the brain's ability to effectively encode binaural cues might contribute to diminished temporal resolution, which is a key factor in challenging listening conditions. Additionally, Grose and Mamo (2010) indicated that the aging process may impair the processing of fine temporal structures, further complicating speech perception and providing a broader context for these age-related effects.

The interaction between age and reverberation further suggests that the young adults' advantage is most pronounced in anechoic and short reverberation conditions. In these environments, the auditory system can better resolve fine-grained speech features, such as formant transitions, which are critical for intelligibility (Grose and Mamo, 2012).

4.2 Effect of noise location on speech perception in reverberant environments

The results showed that speech recognition was better when the noise was spatially segregated from the incoming frontal speech, particularly on the right as compared to the left (Table 4). Although all participants exhibited symmetric hearing, this asymmetry may arise from hemispheric specializations. Noise from the right side primarily reaches the right ear, projecting contralaterally to the left hemisphere, where language-dominant processing can actively disambiguate speech from noise via linguistic cues, enhancing stream segregation and resolving informational masking beyond spatial mechanisms alone (Ruggles and Shinn-Cunningham, 2011). In this scenario, the left hemisphere's engagement allows for robust integration of noisy input with the speech signals, effectively prioritizing speech (Scott and Johnsrude, 2003).

Conversely, noise from the left side projects more to the right hemisphere, which excels in spatial cue processing but lacks equivalent linguistic support, leading to poorer segregation (Zatorre et al., 2002). This interpretation is supported by neuroimaging and behavioral evidence of asymmetries in auditory processing (Tervaniemi and Hugdahl, 2003).

4.3 Effect of reverberation time on speech perception in reverberant environments

Table 5 shows the impact of reverberation on both groups. This underscores the role of temporal smearing in degrading speech cues (Hiscock and Kinsbourne, 2011). In long reverberation conditions, reflections overlap with direct sound, blurring temporal and spectral information, which reduces the effectiveness of spatial cues for both young and middle-aged adults (Nabelek and Pickett, 1974). This explains why group differences diminish in long reverberation conditions with collocated noise as both groups struggle equally with distorted speech signals (Table 4). Reverberation disproportionately affects speech intelligibility when spatial separation is limited as the reflections mask binaural cues (Xia et al., 2018).

For young adults, the persistence of some spatial benefit in long reverberation with separated noise suggests that their auditory system retains partial resilience, possibly due to stronger top–down cognitive compensation, such as selective attention, which aids in segregating speech streams (Duquesnoy and Plomp, 1980). The interaction between noise location and reverberation further explains how acoustic environments modulate auditory performance. In anechoic and short reverberation conditions, spatial separation enhances the performance, particularly for young adults, because clear spatial cues facilitate source segregation. However, in long reverberation conditions, spatial benefits are minimal as reflections degrade interaural cues; this aligns with the existing literature, which noted that reverberant energy disrupts binaural processing (Grose and Mamo, 2012). Consistent with our findings, Missoni et al. (2025) reported reduced SRM under both short (0.6 s) and long (3.0 s) reverberation conditions in Mandarin and German listeners, which was assessed via language-specific matrix sentence tests over headphones. The reduction stemmed primarily from impairment of binaural cues (e.g., ILDs and ITDs) brought about by reverberation and the degradation of monaural cues such as temporal fine structure and envelope modulations, with a more pronounced effect in Mandarin due to the reliance on pitch cues for tonal languages. The cross-linguistic nature of their results supports the view that reverberation-related SRM degradation reflects universal auditory processing constraints rather than language-specific factors, aligning with the present study's observation that the impact of reverberation persists across listener groups and spatial configurations. However, Figure 3 reveals a wide spread in speech recognition scores among middle-aged adults under long reverberation conditions, though these differences are not statistically significant. This variability may result from individual differences, potentially reflecting compensatory cognitive strategies or experience-driven adaptations in some middle-aged adults (Hu et al., 2025). This variability may also reflect differences in auditory processing or attention that was not captured by group averages, highlighting a limitation in generalizing the performance across this age group (Wingfield and Grossman, 2006).

Future research should first establish the behavioral reductions in spatial processing observed in middle-aged adults and subsequently investigate the underlying neural mechanisms using techniques such as electroencephalography (EEG) to assess auditory evoked potentials and attentional modulations to spatial and reverberant stimuli. The study's controlled laboratory conditions, while necessary for isolating variables, limit ecological validity as real-world environments often involve dynamic noise sources and variable reverberation. Another key limitation of the current simulation is the assumption of uniform surface reflectivity, which reduces ecological validity by neglecting the non-uniform absorption and scattering in real environments (Kuttruff, 2016). This simplification may exaggerate the group differences as middle-aged and older adults are more susceptible to reverberant degradations in speech intelligibility, independent of hearing loss (Anderson et al., 2012; Gordon-Salant and Fitzgibbons, 1999). Future work should incorporate varied surface reflectivity to improve realism and isolate age-related vulnerabilities. The sample population was restricted to listeners with normal hearing to understand the baseline trend. This should be translated to populations with hearing impairment to assess the generalizability potential adaptations.

5 Conclusion

This study demonstrates that young adults achieved higher SRS than middle-aged adults, particularly when speech is spatially separated from the noise and reverberation was low. Long reverberation conditions degraded the performance in both groups; however, spatial release from masking was not observed in middle-aged adults, but was preserved in young adults. The results suggest early age-related declines in stream segregation despite normal audiograms, with reverberation acting as a major limiting factor. Clinically, these findings underscore the need for proactive interventions, such as auditory training programs or assistive listening devices, to mitigate challenges in everyday settings such as noisy classrooms where reverberation can obscure the teachers' instructions and impair learning or bustling workplaces, where spatial noise segregation is essential for effective communication and productivity. Addressing these declines requires both acoustic optimization and listener-specific strategies, especially in complex listening environments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the AIISH Institute Review Board (IRB), All India Institute of Speech and Hearing, Mysore. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MH: Writing – review and editing, Writing – original draft. AP: Writing – review and editing, Formal Analysis. KN: Writing – review and editing, Conceptualization, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the All India Institute of Speech and Hearing (AIISH) Research Fund (ARF), Mysuru, India, through the Article Processing Charges (APC) support scheme (Ref. SH/RDC/ARF-APC-NKV/2025-26, dated 16 October 2025). The funding was provided solely to cover publication-related expenses. The funder had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Acknowledgements

The authors sincerely thank the Director of the All India Institute of Speech and Hearing and the head of the Department of Audiology and the Department of Centre of Hearing Sciences for their support and for providing resources for the testing process. Heartfelt appreciation is also extended to all the participants for their valuable time and kind cooperation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

A correction has been made to this article. Details can be found at: 10.3389/frvir.2025.1755815.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE workshop on the applications of signal processing to audio and acoustics (Cat No01TH8575)* (New Platz, NY: IEEE).

Allen, J. B., and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65 (4), 943–950. doi:10.1121/1.382599

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). Aging affects neural precision of speech encoding. *J. Neurosci.* 32 (41), 14156–14164. doi:10.1523/jneurosci.2176-12.2012

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2013). Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *J. Speech Lang. Hear Res.* 56 (1), 31–43. doi:10.1044/1092-4388(2012/12-0043)

Blauert, J. (1996). *Spatial hearing: the psychophysics of human sound localization*. 2nd ed. The MIT Press. Available online at: https://direct.mit.edu/books/book/4885/Spatial-HearingThe-Psychophysics-of-Human-Sound.

Bregman, A. S., and McAdams, S. (1994). Auditory scene analysis: the perceptual organization of sound. *J. Acoust. Soc. Am.* 95 (2), 1177–1178. doi:10.1121/1.408434

Bronkhorst, A. W., and Plomp, R. (1990). A clinical test for the assessment of binaural speech perception in noise. *Int. J. Audiol.* 29 (5), 275–285. doi:10.3109/00206099009072858

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* 77 (5), 1465–1487. doi:10.3758/s13414-015-0882-9

Champely, S. (2020). pwr: basic functions for power analysis. Available online at: https://CRAN.R-project.org/package=pwr.

Chopra, S., Kaur, H., Pandey, R., and Nehra, A. (2018). Development of neuropsychological evaluation screening tool: an education-free cognitive screening instrument. *Neurol. India* 66 (2), 391. doi:10.4103/0028-3886.227304

Deroche, M. L. D., Culling, J. F., Lavandier, M., and Gracco, V. L. (2017). Reverberation limits the release from informational masking obtained in the harmonic and binaural domains. *Atten. Percept. Psychophys.* 79 (1), 363–379. doi:10.3758/s13414-016-1207-3

Doggett, R., Sander, E. J., Birt, J., Ottley, M., and Baumann, O. (2021). Using virtual reality to evaluate the impact of room acoustics on cognitive performance and well-being. *Front. Virtual Real* 2, 620503. doi:10.3389/frvir.2021.620503

Duquesnoy, A. J., and Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *J. Acoust. Soc. Am.* 68 (2), 537–544. doi:10.1121/1.384767

Geetha, C., Kumar, K., Manjula, P., and Pavan, M. (2014). Development and standardisation of the sentenceidentification test in the Kannada language. *J. Hear Sci.* 4 (1), 18–26. doi:10.17430/890267

Gordon-Salant, S., and Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *J. Speech Lang. Hear Res.* 42 (2), 300–311. doi:10.1044/jslhr.4202.300

Grose, J. H., and Mamo, S. K. (2012). Frequency modulation detection as a measure of temporal processing: age-related monaural and binaural effects. *Hear Res.* 294 (1–2), 49–54. doi:10.1016/j.heares.2012.09.007

Grose, J. H., and Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear Hear* 31 (6), 755–760. doi:10.1097/aud.0b013e3181e627e7

Grose, J. H., Buss, E., and Elmore, H. (2019). Age-related changes in the auditory brainstem response and suprathreshold processing of temporal and spectral modulation. *Trends Hear* 23, 2331216519839615. doi:10.1177/2331216519839615

Hartig, F. (2024). DHARMa: residual diagnostics for hierarchical (Multi-Level/mixed) regression models. Available online at: https://CRAN.R-project.org/package=DHARMa.

Hiscock, M., and Kinsbourne, M. (2011). Attention and the right-ear advantage: what is the connection? *Brain Cogn.* 76 (2), 263–275. doi:10.1016/j.bandc.2011.03.016

Hu, H., Biberger, T., Kollmeier, B., Vickers, D., Chen, F., and Ewert, S. D. (2025). Speech intelligibility and spatial release from masking in anechoic and reverberant rooms: comparison between German-speaking and Mandarin-speaking listeners. *J. Acoust. Soc. Am.* 158 (1), 259–275. doi:10.1121/10.0036943

Kuttruff, H. (2016). *Room acoustics*. 6th ed. Boca Raton, MA: CRC Press.

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear Res.* 50 (4), 940–967. doi:10.1044/1092-4388(2007/067)

Marrone, N., Mason, C. R., and Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *J. Acoust. Soc. Am.* 124 (5), 3064–3075. doi:10.1121/1.2980441

Middlebrooks, J. C., and Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.* 42 (1), 135–159. doi:10.1146/annurev.psych.42.1.135

Missoni, F., Poole, K., Picinali, L., and Canessa, A. (2025). Effects of auditory distance cues and reverberation on spatial perception and listening strategies. *arXiv*. doi:10.1038/s44384-025-00027-4

Nabelek, A. K., and Pickett, J. M. (1974). Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. *J. Speech Hear Res.* 17 (4), 724–739. doi:10.1044/jshr.1704.724

Nam, H., and Park, Y. H. (2025). Coherence-based phonemic analysis on the effect of reverberation to practical automatic speech recognition. *Appl. Acoust.* 227, 110233. doi:10.1016/j.apacoust.2024.110233

Opochinsky, R., Moradi, M., and Gannot, S. (2025). Single-microphone speaker separation and voice activity detection in noisy and reverberant environments. *EURASIP J. Audio Speech Music Process* 2025 (1), 18. doi:10.1186/s13636-025-00404-7

Pichora-Fuller, M. K., and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42 (Suppl. 2), S11–S16. doi:10.3109/14992020309074638

R Core Team (2025). *R: a language and environment for statistical computing*. Vienna, Austria. Available online at: https://www.R-project.org/.

Ruggles, D., and Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. *J. Assoc. Res. Otolaryngol. JARO* 12 (3), 395–405. doi:10.1007/s10162-010-0254-z

Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26 (2), 100–107. doi:10.1016/s0166-2236(02)00037-1

Serafin, S., Adjorlu, A., and Percy-Smith, L. M. (2023). A review of virtual reality for individuals with hearing impairments. *Multimodal Technol. Interact.* 7 (4), 36. doi:10.3390/mti7040036

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34 (3), 114–123. doi:10.1016/j.tins.2010.11.002

Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* 11 (1), 54–71. doi:10.1037/1082-989x.11.1.54

Srinivasan, N. K., Stansell, M., and Gallun, F. J. (2017). The role of early and late reflections on spatial release from masking: effects of age and hearing loss. *J. Acoust. Soc. Am.* 141 (3), EL185–EL191. doi:10.1121/1.4973837

Steeneken, H. J. M., and Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *J. Acoust. Soc. Am.* 67 (1), 318–326. doi:10.1121/1.384464

Tervaniemi, M., and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res. Rev.* 43 (3), 231–246. doi:10.1016/j.brainresrev.2003.08.004

Tillery, A. D., Varjas, K., Roach, A. T., Kuperminc, G. P., and Meyers, J. (2013). The importance of adult connections in adolescents' sense of school belonging: implications for schools and practitioners. *J. Sch. Violence* 12 (2), 134–155. doi:10.1080/15388220.2012.762518

Vaidyanath, R., and Yathiraj, A. (2014). Screening checklist for auditory processingin adults (SCAP-A): development and preliminary findings. *J. Hear Sci.* 4 (1), 27–37. doi:10.17430/890788

Viswanathan, N., Kokkinakis, K., and Williams, B. T. (2016). Spatially separating language masker from target results in spatial and linguistic masking release. *J. Acoust. Soc. Am.* 140 (6), EL465–EL470. doi:10.1121/1.4968034

Viveros Muñoz, R., Aspöck, L., and Fels, J. (2019). Spatial release from masking under different reverberant conditions in young and elderly subjects: effect of moving or stationary maskers at circular and radial conditions. *J. Speech Lang. Hear Res.* 62 (9), 3582–3595. doi:10.1044/2019_jslhr-h-19-0092

Windle, R., Dillon, H., and Heinrich, A. (2023). A review of auditory processing and cognitive change during normal ageing, and the implications for setting hearing aids for older adults. *Front. Neurol.* 14, 1122420. doi:10.3389/fneur.2023.1122420

Wingfield, A., and Grossman, M. (2006). Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. *J. Neurophysiol.* 96 (6), 2830–2839. doi:10.1152/jn.00628.2006

Xia, J., Xu, B., Pentony, S., Xu, J., and Swaminathan, J. (2018). Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners. *J. Acoust. Soc. Am.* 143 (3), 1523–1533. doi:10.1121/1.5026788

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6 (1), 37–46. doi:10.1016/s1364-6613(00)01816-7

Zenke, K., and Rosen, S. (2022). Spatial release of masking in children and adults in non-individualized virtual environments. *J. Acoust. Soc. Am.* 152 (6), 3384–3395. doi:10.1121/10.0016360