Archives of Acoustics Vol. 50, No. 4, pp. 455–464 (2025), https://doi.org/10.24425/aoa.2025.154828

Research Paper

# Effect of Listener Head Position on Speech Intelligibility in an Automotive Cabin

Linda LIANG(1)*, Linghui LIAO(1), Jiahui SUN(1), Lingling LIU(1), Liuying OU(2), Xiaoyue HUANG(1)

(1) *College of Civil Engineering and Architecture, Guangxi University*
Nanning, China

(2) *Guangxi Vocational University of Agriculture*
Nanning, China

*Corresponding Author e-mail: ldliang@gxu.edu.cn

*Received November 21, 2024; revised June 9, 2025; accepted September 16, 2025; published online October 17, 2025.*

When evaluating speech intelligibility (SI) in automotive cabins, binaural measurements typically employ a fixed dummy head. However, the impact of listener head positions on SI in nonuniform cabin sound fields remains unclear. This study analyzed SI under various listener head positions in an automotive cabin. An artificial mouth was regarded as the speaker, which was placed in three passenger positions. Binaural room impulse responses were measured using a dummy head in the driver's seat with various head positions. The results show that head position significantly affects SI, with variations of up to 7 dB in octave band magnitudes, more than one just-noticeable difference in the speech transmission index, and shifts of up to 2.5 dB in the speech-reception threshold. SI variability depends on the speaker's location. Directivity patterns play a crucial role in the front-passenger position, while seat occlusion affects SI at the back-right position, causing substantial decreases below a certain height threshold. At the back-left position, head positions close to the headrest enhance SI due to distance and reflections. Minor head displacements (4 cm apart) generally have insignificant effects on SI, except near seat obstructions or reach critical thresholds.

**Keywords:** automotive cabin; speech intelligibility; head position; speech reception threshold; speech transmission index.



## 1. Introduction

In recent years, the automobile has evolved from a simple means of transportation into an essential part of everyday life, often referred to as thr third space. Consequently, acoustic comfort has emerged as a notable area of concern due to increasing consumer demands (Miqueau *et al.*, 2024). Speech intelligibility (SI) is strongly associated with the level of acoustic comfort perceived by passengers within automotive cabins (Biswas *et al.*, 2022). Thus, it plays a vital role in enhancing safety and the overall travel experience.

However, the acoustic environments within automotive cabins possess unique characteristics that distinguish them from traditional rooms, thereby rendering SI in automobiles a specific concern (Parizet, 1993). The confined dimensions and intricate boundary conditions within automotive cabins result in a notable low-frequency resonance and rapid attenuation of high-frequency sounds (Granier *et al.*, 1996; Rumsey, 2016; Meissner, 2017). Many of the reflections are early reflections (Granier *et al.*, 1996; Kleiner, Tichy, 2014; Rumsey, 2016), which are considered advantageous for SI (Bradley *et al.*, 2003; Arweiler, Buchholz, 2011; Warzybok *et al.*, 2013). Consequently, the adverse effects of reverberation on intelligibility can be disregarded (Samardzic, Novak, 2011a; 2011b; Gerrera *et al.*, 2016; Ebbitt, Remtema, 2015). Furthermore, seatbacks play a pivotal role in sound absorption within automobiles

(Parizet, 1993; Visintainer, VanBuskirk, 1997; Cao *et al.*, 2022). Seat occlusions diminish speech energy transmission from rear speakers to listeners in the front (or vice versa), significantly impairing SI (Liang *et al.*, 2021). Moreover, background noise in automotive cabins has a substantial impact on SI, as it exhibits unique and fluctuating characteristics based on speed, operating conditions, and road conditions, which are absent in traditional indoor environments (Samardzic, Novak, 2011a; 2011b; Parizet, 1992; Wang *et al.*, 2012; Samardzic, 2014). The interior environment of automotive cabins demonstrates the considerable signal-to-noise ratio (SNR) variations (Dal Degan, Prati, 1988; Ferrari *et al.*, 2023). In contrast to the quieter and more constant background noise prevalent in traditional indoor settings, the SI within automotive cabins is influenced by background noise (or SNR) rather than reverberation (Ebbitt, Remtema, 2015; Samardzic, Novak, 2011a; 2011b; Gerrera *et al.*, 2016; Liang *et al.*, 2021).

Furthermore, the extremely confined dimensions of automotive cabins place the speaker and listener within the near-field zone, which further complicates the SI variations within the cabin (Liang, Yu, 2020). Specifically, SI measurements within automotive cabins are more sensitive to factors such as speaker directivity, orientation, and position compared to typical indoor environments (Bilzi *et al.*, 2005; Liang, Yu, 2023b). Moreover, binaural listening phenomena, including binaural interactions and the head shadow effect (van Wijngaarden, Drullman, 2008), introduce an effective SNR that differs between the ears of the listener (Liang, Yu, 2020). These phenomena have a direct impact on SI in automobiles. The SI in automotive cabins is strongly influenced by the direction and distance of the speaker relative to the listener's ears. The combination of the near-field head shadow effect and the unique sound field characteristics within automotive cabins (such as the nonuniform distribution of early reflections and seatback occlusions) renders SI under binaural listening conditions in automobiles more complex than in traditional indoor environments (Liang *et al.*, 2021; Liang, Yu, 2023b). Consequently, for an accurate SI evaluation within automotive cabins, it is imperative to use binaural measurements (Samardzic, Moore, 2021) and consider the orientation of the listener's head (Liang *et al.*, 2021; Liang, Yu, 2023b). Neglecting these factors can result in substantial deviations in the SI assessment.

In previous studies evaluating SI within automotive cabins, binaural signals were typically captured using a dummy head in a static position (Ebbitt, Remtema, 2015; Samardzic, Novak, 2011a; 2011b; Liang *et al.*, 2021; Liang, Yu, 2023b; Samardzic, Moore, 2021). However, the nonuniform sound pressure distribution within automotive cabins is influenced by acoustic resonances and the irregular distribution of absorptive and reflective surfaces (Granier *et al.*, 1996; Rumsey, 2016). Given the interplay between the binaural effect and the unique sound field characteristics within the confined acoustic space of an automobile, it is anticipated that the variations in listener head position would result in significant differences in the sound pressure level (SPL) experienced by the ears (Ghanati, Azadi, 2020; Granier *et al.*, 1996; Rumsey, 2016). Consequently, the SI may undergo considerable fluctuations due to the uncertainty introduced by the passenger head displacement. To the best of our knowledge, this issue has not yet been thoroughly examined.

This work aims to investigate the impact of the listener head position on SI evaluations within an automotive cabin. Specifically, the primary objective is to quantify the extent to which SI discrepancies arise due to changes in listener head positions and to establish a benchmark for SI measurements in such environments. Initially, binaural room impulse responses (BRIRs) were measured with a speaker at three passenger locations: the front passenger (FP), back left (BL), and back right (BR) seats. During these measurements, an artificial mouth was used to emulate the speaker. A dummy head was placed in the driver's seat and at various spatial locations, encompassing four different heights multiplied by five horizontal positions, resulting in a total of 20 head positions. Subsequently, the magnitude spectra of the BRIRs, speech transmission indices (STIs), and speech reception thresholds (SRTs) in Mandarin Chinese were evaluated.

## 2. Methods and materials

### *2.1. BRIR measurements*

The measurements for this study were conducted within a Volkswagen Tiguan L, with dimensions of 4733 mm by 1839 mm by 1673 mm in length, width, and height, respectively. A simplified top-down view of the automobile is depicted in Fig. 1. To streamline the analysis and focus on prevalent scenarios, the listener was in the driver's seat for this study, represented by a dummy head. The dummy head used in this study is a statistical shape model-based average head model (SSMAH) (Wang, Yu, 2025) created from 100 Chinese adults (74 men and 26 women). The dummy head's primary components, including the torso, head, and shoulders, were fabricated using ABS plastic, while the pinnae were crafted from silicone rubber. To rigorously investigate the impact of head position on SI evaluation results, this study excluded the consideration of head displacement resulting from seat adjustments, which could potentially alter the sound field within the automotive cabin. To ensure stability during measurements, the seat was securely fixed in place.

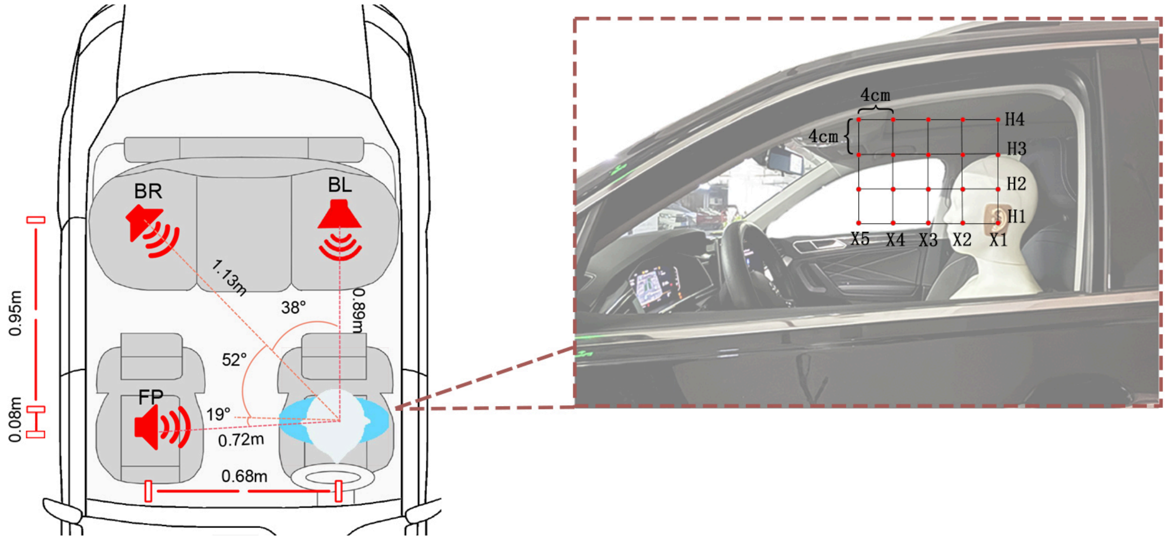


Fig. 1. Schematic of the experimental setup in the automotive cabin.

To streamline the problem and align with typical scenarios, the analysis focuses solely on the front-back and up-down dimensions of the listener's head position. A thick rectangular plastic plate with markings was laid horizontally on the driver's seat to maintain the dummy head's uniform movement in the horizontal plane. Following the measurement of one height, a 4 cm thick plastic plate was added to facilitate the dummy head's movement in the vertical direction. Consequently, the dummy head was positioned centrally in the left-right dimension of the driver's seat, facing forward. The ear canal entrance of the dummy head was systematically placed in 20 distinct locations, comprising 4 vertical levels (designated as H1 to H4, representing various heights above the seat cushion) and 5 horizontal points (designated as X1 to X5, representing different distances from the headrest). Each position was spaced 4 cm apart, as illustrated in Fig. 1. The precise location of the ear canal entrance was calibrated using a 3D laser calibrator (LSG686SPD), positioned outside the side window of the automobile. The positioning of the dummy head's head at the H1 height signifies that its ear canal entrance was 1.22 m above the ground plane.

The experiment used an artificial mouth (GRAS 44AB) as the speaker, which exhibited comparable directivity and frequency response characteristics to a human mouth. It is important to highlight that the GRAS 44AB, as initially outlined in its product documentation, was designed primarily for testing telephone mouthpieces and comparable microphones within communication systems, intended specifically for the close-proximity use. The directivity pattern of this artificial mouth might not perfectly match that of a human speaker at slightly longer distances. Nonetheless, considering that the automotive cockpit environment, which this study examines, inherently represents a unique near-field range, the influence of minor variations in directivity is expected to be relatively minor. The speaker was sequentially placed in the FP, BR, and BL locations, with its front consistently oriented towards the listener (refer to Fig. 1). The speaker was placed at a height of 1.28 m above the ground; a value determined through measurements of the mouth height of a representative sample of Chinese males with an average stature of 1.70 m. When at the FP, BR, and BL locations, the speaker was arranged at distances of 0.68 m, 1.13 m, and 0.89 m, respectively, from the listener occupying the (H1, X1) coordinate. Furthermore, the speaker was oriented at approximate angles of −19°, 52°, and 90° to the right of the listener's position.

During the measurements, all windows, doors, and the automotive air conditioning system were meticulously closed to eliminate extraneous noise. A maximum-length sequence, characterized by a 48 kHz sampling frequency, a duration of 6 s, and 24-bit quantization, served as the excitation signal. This signal was converted from digital to analog format using the Roland Studio Capture 1610 sound card and subsequently fed to the speaker. To capture the binaural signals, a pair of DPA 4060 miniature microphones were precisely positioned within the occluded ear canal entrances of the dummy head. Following this, the noise-free BRIRs were derived through deconvolution using cross-correlation between the original excitation signal and the recorded binaural signals.

### *2.2. STI calculation*

Previous studies have comprehensively established that STI can effectively predict SI within automotive cabins (Samardzic, Novak, 2011a; Liang *et al.*, 2021), despite the negligible temporal characteristics of the transfer function within these environments. The modulation transfer function represents the intelligibility interference arising from the temporal modulation reduction by the transmission system, as outlined in previous research (International Electrotechnical Commission [IEC], 2011; Houtgast *et al.*, 1980). Using the indirect method theory (Schroeder, 1981; Rife, 1992), the STI computation necessitates only the acquisition of the impulse response (refer to BRIRs in Subsec. 2.1) and the SNR.

Varying listener head positions and speaker positions can introduce significant variations in the speech levels received by each ear, potentially leading to discrepancies in the SNR perceived by the listener. To ensure a consistent transfer function, the BRIRs were employed to indirectly obtain the binaural speech signals. To create a monaural speech sample, pink noise was first generated and then filtered using the Chinese standard spectrum specified in (GB/T 7347-1987, 1987). Subsequently, the corresponding BRIRs obtained in Subsec. 2.1 were convolved with the monaural speech sample to produce the binaural speech signals. Additionally, background noise was sourced from actual measurements of binaural noise at 100 km/h within a fuel-powered vehicle. In reality, the SPLs of the binaural noise signals were very close between the left and right ears, with a difference of less than 0.3 dB(A). Using stationary noise and the binaural speech signals, the SNRs were indirectly derived for different listener head positions and speaker positions. Typically, it is necessary to measure the SPLs of both the noise and the binaural speech signal independently for determining the SNR needed to calculate the STI. In this study, both the speech signal and noise were considered virtual signals, making the SNR a relative value. This approach is primarily used to emphasize the changes in STI resulting from variations in the transfer function due to different head and speaker positions. Consequently, we select an appropriate value to observe the trend of STI changes. Then, the STI was calculated using the SNRs and BRIRs through the indirect STI approach, as specified in the IEC (2011) standard.

Actually, STI is essentially a monaural model. According to (IEC, 2011), when performing binaural STI measurements using an artificial head, the recommended approach is to use the results of STI for the better ear, i.e., selecting the better (larger) value from the pair of STIs. In practice, the better-ear STI cannot fully show the benefit of listening with two ears. To date, no standard for combining different STI measurements from the two ears has been developed, and so the advantages of binaural hearing in SI are always disregarded (van Wijngaarden, Drullman, 2008; Liang *et al.*, 2022). Nonetheless, the better-ear STI is still the most effective indicator compared with the existing binaural STI model (van Wijngaarden, Drullman, 2008), thus it is adopted in the present study.

#### *2.2.1. Subjective experiment*

In practice, the STI falls short of accurately capturing the impact of binaural hearing and low-frequency components on SI within the confined space of an automotive cabin, offering merely a partial estimation of the true SI level, as noted in (van Wijngaarden, Drullman, 2008; Huang *et al.*, 2023). To address this limitation, a supplementary subjective experiment was conducted to obtain the SRTs, defined as the SNR required for 50 % intelligibility. This subjective experiment encompassed 36 test conditions, factoring in 3 distinct speaker positions (namely, FP, BR, and BL) and 12 head positions, which were determined by 4 different heights (namely, H1–H4) and 3 horizontal coordinates (namely, X1, X3, and X5).

#### *2.2.2. Participants*

For this study, 12 volunteers were recruited, comprising an equal gender distribution with 6 males and 6 females. The study participants ranged in age from 20 to 25 years, with a mean age of 21.83 years. They were drawn from a pool of undergraduate and graduate students at Guangxi University. Each participant reported normal hearing and was a native Mandarin-speaking Chinese individual hailing from diverse geographical regions. As a token of appreciation, participants were compensated for their involvement in the study.

#### *2.2.3. Stimuli and procedure*

The Mandarin Chinese matrix sentence test served as the source of sentences comprising the target speech, as reported in (Hu *et al.*, 2018). Each sentence within the corpus adhered to a pre-established syntactic structure containing five words: *name*, *verb*, *number*, *adjective*, and *object*. This framework was both grammatically correct and semantically unpredictable. A total of 40 lists, each containing 20 sentences, was used. Employing auralization technology, the target speech was emitted from various passenger locations to the driver's position by convolving it with the corresponding BRIR. The interfering signal was the binaural noise captured in authentic automotive environments, which had previously been used for STI calculations. The interferer's level was set to approximately 60 dB(A) for each ear to ensure comfort. Prior to convolution with the BRIR, the target speech's level was adjusted to achieve different SNRs. The binaural interferer was then combined with the convolved binaural target speech to produce the final binaural signals.

The experiment was conducted in a room with ambient noise levels below 30 dB(A). An adaptive up-down method, initiated with an SNR of 10 dB, was used to measure the SRTs, following the procedure outlined in (Brand, Kollmeier, 2002). Notably, the SNR referred to the difference between the interferer and target speech levels prior to convolution, rather than the actual SNR at the listener's ears. Stimuli were presented using Sennheiser HD650 headphones powered by a Roland Studio Capture 1610 sound card. Participants were instructed to independently mark the terms they heard and understood on a MATLAB graphical user interface (GUI) during the close-set response format assessments. Across the 36 test conditions considered, each participant completed a total of 72 runs, with each test condition repeated two times. The final SRT was obtained by averaging the SRTs from the two repetitions. Given that the total number of runs (72) exceeded the number of lists (40), some lists were used twice. However, this had no effect on the outcomes, as the corpus was designed to be semantically surprising and suitable for two uses by the listener, as noted in (Hu *et al.*, 2018). Participants were presented with random lists and test conditions in different sequences. To minimize listener fatigue, the 72 runs, each lasting 4 to 4.5 minutes, were divided into two sessions spaced at least 12 hours apart, with a 20-minute break after every eight runs. Each session began with a training period.

## 3. Results and discussion

### *3.1. Effect of listener head position on BRIR magnitude spectra*

The magnitude spectra of the BRIRs measured in Subsec. 2.1 were computed for diverse head positions, with the speaker at the FP, BR, and BL locations. Figure 2 illustrates the variations in the magnitude spectra of the measured BRIRs relative to the reference position (H3, X3), encompassing results across the 125 Hz–8000 Hz octave bands as well as the overall results. Table 1 provides the range of magnitude variations among the 20 different head positions.

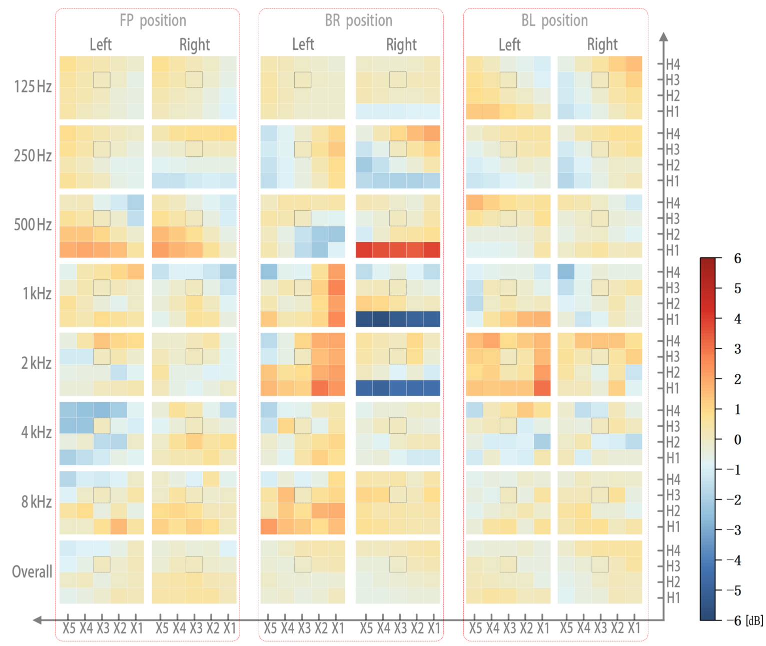


Fig. 2. Variations in the magnitude spectra (125 Hz–8000 Hz octave bands and the overall magnitude) of measured binaural room impulse responses (BRIRs) under different head positions compared to the reference position (X3, H3), when the speaker was located in the FP, BR, and BL positions.

Table 1. Magnitude variation among 20 head positions, including the 125 Hz–8000 Hz octave bands and the overall magnitude.

| Speaker | Ear | Magnitude variation [dB] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 125 Hz | 250 Hz | 500 Hz | 1 kHz | 2 kHz | 4 kHz | 8 kHz | Overall |
| FP position | Left | 1.38 | 1.55 | 3.82 | 2.08 | 2.76 | 2.61 | 3.30 | 1.03 |
| | Right | 1.63 | 2.31 | 3.25 | 2.63 | 1.49 | 2.64 | 1.96 | 1.26 |
| BR position | Left | 0.69 | 2.66 | 2.70 | 4.99 | 4.48 | 2.98 | 3.18 | 0.76 |
| | Right | 1.23 | 4.15 | 4.99 | 6.85 | 5.92 | 1.28 | 1.07 | 1.01 |
| BL position | Left | 2.28 | 1.89 | 2.41 | 3.46 | 3.12 | 3.24 | 1.43 | 1.36 |
| | Right | 2.95 | 2.58 | 1.65 | 2.91 | 2.64 | 2.05 | 1.50 | 1.40 |

As depicted in Fig. 2 and Table 1, irrespective of the speaker's location, the overall magnitude variation induced by the displacement of the listener's head falls within a range of 1 dB to 1.5 dB. Specifically, when the speaker is at the BL location, the magnitude variation is slightly more pronounced compared to other speaker positions, whereas it is minimal when the speaker is situated at the BR location. Furthermore, there are disparities in the magnitude fluctuations between the two ears across various listener head positions. Notably, the magnitude difference for the ipsilateral ear (i.e., the right ear in the case of the FP and BR speaker positions) is significantly greater than that for the contralateral ear, with a difference approaching 0.2 dB.

Regarding the outcomes observed for each octave band, the magnitude difference resulting from the displacement of the listener's head is more pronounced, occasionally approaching a value of 7 dB (as shown in Table 1). Notably, the octave bands ranging from 500 Hz to 4000 Hz exhibit larger magnitude variations compared to other frequency bands. When the speaker is at the BR location, the magnitude recorded in the right ear at the H1 height is significantly reduced in most frequency bands below 4 kHz (excluding the 500 Hz band) when compared to higher heights (H2 to H4). This reduction can be attributed to the direct obstruction of sound waves emitted by the speaker at the BR position by the driver's seatback when the listener's ear canal is at the H1 height. The opposite trend observed in the 500 Hz octave band may be due to standing wave phenomena within the automotive cabin, as the resonance zone typically falls within the 1 kHz range (Rumsey, 2016). Indeed, the standing wave phenomenon within the cabin often results in inconsistent trends in magnitude variations within the 125 Hz to 500 Hz octave bands compared to higher frequency bands (as depicted in Fig. 2).

### *3.2. Effect of listener head position on STI results*

Figure 3 illustrates the better-ear STIs recorded using a dummy head at various head positions, encompassing the results obtained when the speaker was situated at the FP, BR, and BL positions. When the speaker is at the FP location, the overall variation in the STI value resulting from the listener's head position remains within 0.024, which is below the just noticeable difference (JND, 0.03) as reported in (Bradley *et al.*, 1999). When the head is at the horizontal coordinate X3, the STI value tends to be higher at the same vertical level compared to other head positions (see Fig. 3a). This is primarily because in the X3 coordinate, the listener's head is positioned nearest to the principal axis direction of the artificial mouth (speaker), as well as is closer to it. Furthermore, the radiation characteristics of the artificial mouth dictate that the radiation intensity peaks in the direction of its principal axis (Liang, Yu, 2023a).

When the speaker is at the BR location, the overall variation in the STI value due to the changes in listener head position is as high as 0.043, exceeding 1 JND. This variation is significantly larger compared to the scenario where the speaker is located at the FP position. The fluctuation in STI is primarily evident in the substantial difference between the heights of H1 and H2. Conversely, the differences among the heights of H2, H3, and H4 remain within 0.01 (as illustrated in Fig. 3b). This observation aligns with the BRIR magnitude results depicted in Fig. 2, which is attributed to the direct obstruction caused by the driver's seatback.

When the speaker is positioned directly behind the listener, specifically in the BL position, the overall variation in the STI value is 0.033, exceeding 1 JND. This variation is slightly lower than that observed in the BR position but slightly higher than that in the FP position. For head positions situated closer to the headrest, such as X1, there is a tendency for larger STI values, particularly at the H1 and H2 heights, as depicted in Fig. 3c. This phenomenon may be attributed to the proximity of the listener to the speaker or the fact that the reflection area generated by the left rear window is situated closer to the headrest. Relevant insights can also be gleaned from previous results for the magnitude spectra, specifically the magnitude observed at the left ear in Fig. 2.

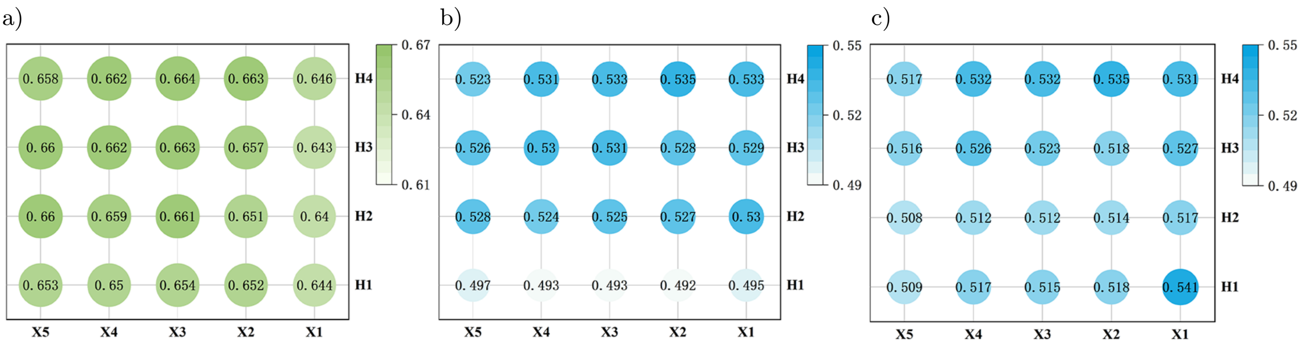


Fig. 3. Speech transmission index (STI) results obtained through a dummy head positioned at various locations when the speaker was situated in three different seats: a) the front passenger (FP); b) the back right (BR); c) the back left (BL) positions.

In general, the STI values observed when the speaker is in the FP location, ranging from 0.64 to 0.664, surpass those observed in the BR position (0.492 to 0.535) and the BL position (0.508 to 0.541). This disparity is predominantly due to the fact that, when the speaker is in the FP position, the radiated sound waves are able to reach the listener's ipsilateral (right) ear without any impediments. Conversely, for speakers in the rear (the BR and BL positions), the sound energy received at both ears is substantially diminished as a result of obstructions posed by seatbacks, as well as the listener's head and external ear (pinna) (Liang *et al.*, 2021).

### *3.3. Effect of listener head position on SRT results*

Figure 4 illustrates the Chinese SRT results obtained when the speaker is at the FP, BR, and BL locations, along with the corresponding average values and the standard error of the mean (SEM). When the speaker occupies the FP position, the SRT value, averaging −14.9 dB, is consistently lower than that of rear-position speakers, which average more than −8.8 dB. Specifically, the SRT typically attains its highest level when the speaker is at the BL location, averaging as high as −6.9 dB, except in instances where it is occasionally lower than the BR position at the H1 height.

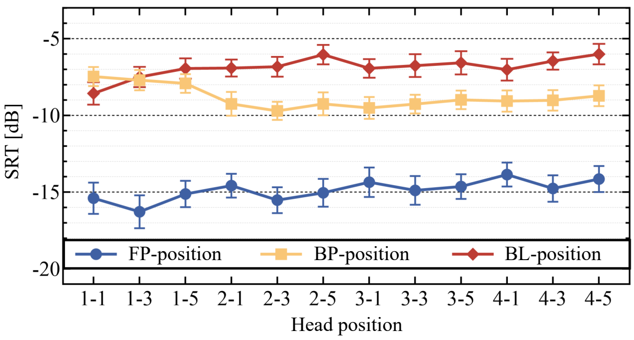


Fig. 4. Speech reception threshold (SRT) results (mean ±standard error of the mean, SEM) from measurements with the dummy head with various head positions when the speaker was in the front passenger (FP), the back right (BR), and the back left (BL) positions.

As depicted in Fig. 4a, when the speaker is located at the FP position, the SRT value fluctuates around −15 dB. Specifically, the SRT reaches a minimum of −16.3 dB at the coordinates (H1, X3) and a maximum of −13.9 dB at the coordinates (H4, X1), yielding a variation range of 2.4 dB. Furthermore, at a constant height, the SRT values recorded at the horizontal coordinate X3 are consistently lower than those at other horizontal coordinates, exemplified by the coordinates (H1, X3), (H2, X3), (H3, X3), and (H4, X3). This observation aligns with the STI results, primarily attributed to the directional characteristics of the speaker (artificial mouth), as illustrated in Fig. 3.

When the speaker is at the BR location, the SRT value ranges from −7 dB to −8 dB at the H1 height, marking a significant increase compared to other head positions. For the remaining head positions, the SRT value fluctuates around −9 dB, with a narrow variation range of approximately 1 dB (refer to Fig. 4b). This finding is in accordance with the previously presented magnitude spectra and STI results (Figs. 2 and 3), which are attributed to seat occlusion. Specifically, the SRT attains its minimum value of −9.7 dB at the coordinates (H2, X3) and its maximum value of −7.4 dB at the coordinates (H1, X1), resulting in an overall variation range of 2.3 dB.

When the speaker is at the BL location, the SRT value attains a minimum of −8.5 dB at the coordinate (H1, X1) and a maximum of −6 dB at the coordinate (H4, X5), encompassing an overall variation range of 2.5 dB (Fig. 4c). Additionally, at a fixed height, the SRT value increases as the head position moves further away from the headrest (or towards the front), suggesting a corresponding decrease in intelligibility. This pattern is consistent with the STI results, primarily due to distance factors and the arrangement of reflection areas generated by the rear side window. Overall, irrespective of the speaker position, certain displacement of the listener's head results in an SRT difference of at least 2 dB, signifying a substantial variation in SI. The SRT distinction for the FP and BL speaker positions is primarily manifested in horizontal displacement, which is influenced by the speaker's directional patterns and the distribution of reflected sound. Conversely, the BR position exhibits a primary difference due to vertical displacement, attributed to seat obstruction.

A two-way analysis of variance (ANOVA) with repeated measures indicated that the SRT results in Mandarin Chinese were significantly influenced by both the speaker position and the head position, with statistical significance demonstrated by $F(2, 22) = 422.9$ and $F(11, 121) = 5.64$, respectively (both $p < 0.0001$). Furthermore, a significant interaction was observed between these two factors for the SRT value, as evidenced by $F(22, 242) = 11.64$ ($p < 0.0001$). This suggests that the impact of head position on SRT varies depending on the speaker's position, and conversely, the influence of speaker position on SRT also varies with changes in head position.

Post-hoc pairwise comparisons, employing Bonferroni corrections, were conducted to assess differences in SRT changes across various head positions. Results indicated that when the speaker was at the FP location, statistically significant differences ($p < 0.05$) were observed between the SRT values at the coordinate (H1, X3) and other head positions, except for the coordinate (H1, X1) and (H2, X3). These differences ranged from 1.16 dB to 2.43 dB. Conversely, at heights H3 and H4, the SRT values across various head positions exhibited nonsignificant differences ($p > 0.05$), with variations not exceeding 1 dB. When the speaker was at the BR location, a significant difference was observed solely between the H1 height and other heights. When the speaker was at the BL location, the significant differences emerged between the SRT values of the head position at the (H1, X1) coordinate and other coordinates.

### 3.4. Implications and limitations

From the above results and analysis, there are considerable variations in SI caused by listener's head positions in automotive cabins, especially when the speaker is in the rear. Small head displacements do not cause significant changes in SI. Significant changes in SI occur only when head displacement reaches a critical threshold or is obstructed by the seatback. Differences in SI caused by driver's heights can be ignored unless their height difference exceeds a certain limit.

These insights offer the reference value for follow-up studies and other researchers involved in binaural SI measurement. For the acoustic design of the automotive cabin, the relevant conclusions in this study emphasize the significance of optimizing seat occlusion for verbal communication between front and rear passengers. From the perspective of the front row speaker having a higher SI level for the listener in the driver seat than that of a speaker in the rear, or from the perspective of the impact of seat occlusion on head position changes, a design with a certain gap between the backrest and headrest may be more advantageous.

This study has inherent limitations, including the use of a single vehicle model and a monolingual participant group, which may restrict the generalizability of the results. Different vehicle models feature varying seat structures, which may lead to different results. Additionally, the interior space dimensions and interior materials also affect the in-cabin acoustic environment. To address these limitations, our future research plans include testing multiple vehicle models, conducting cross-lingual studies, and conducting dynamic driving scenario experiments. It is expected that these conclusions will be like those for other car models, as issues such as seat distribution and seat obstruction in cars share strong similarities. Changes in body shape and details should have a minimal impact on the conclusions drawn from this study.

## 4. Conclusions

To investigate the variations in speech intelligibility (SI) within an automotive cabin as a function of the listener's head position under various speaker locations, this study conducted an exhaustive analysis encompassing BRIR magnitude spectra, STI, and SRT results across various experimental conditions. The research results demonstrate that the SI within the automotive cabin was markedly influenced by the displacement of the listener's head. Across different head positions, notable differences were observed, including octave band magnitudes varying by approximately 7 dB, STI discrepancies exceeding 1 JND, and SRT shifts as substantial as 2.5 dB.

A notable interaction effect on SI was observed between the speaker position and the head position. Specifically, the magnitude of the influence that the head position exerts on SI is contingent on the speaker's position. Conversely, the effect of speaker position on SI also varies in response to alterations in the listener head position. When the speaker was at the FP position, the directivity pattern of the speaker significantly influenced the results. Horizontal coordinates that aligned closely with the speaker's principal axis direction exhibited increased the BRIR magnitude and STI values, coupled with lower SRT values, collectively indicating superior SI. At elevated head heights, the variations in SI were minimal, with SRT changes limited to within 1 dB. In scenarios where the speaker was at the BR position, a substantial decrease in SI was observed when the listener's head was positioned below a certain height threshold. This decrease was primarily attributed to direct obstruction by the seatback, resulting in a decrease in STI by 0.035 and an increase in SRT by more than 2 dB. Conversely, at unobstructed heights, the STI variations remained below 0.01, and SRT changes were generally limited to less than 1 dB. Hence, for speakers at the BR position, the influence of head position on SI was predominantly governed by seat occlusion, particularly in the vertical dimension. When the speaker was located at the BL position, head positions closer to the headrest yielded higher SI, primarily due to the combined effects of distance and reflections from the rear side window. Subjective results revealed insignificant differences among various listener head positions, i.e., the SRT variation did not surpass 1 dB, except for those at the lowest height positions.

Overall, except for head positions at lower heights when the speaker was at the BR position, the differences in STI and SRT values between adjacent measurement points (spaced 4 cm apart) were minor. This suggests that during binaural measurements for SI assessment, minor head displacements do not elicit significant changes. Significant alterations in SI only occur when the head displacement reaches a critical threshold or is obstructed by the seat. This study has systematically analyzed the impact of head position on SI, and its findings offer significant value as a benchmark for future binaural assessments of SI within automotive environments.

Fundings

This work was supported by the Middle-aged and Young Teachers' Basic Ability Promotion Project of

Guangxi (2024KY0028) and the College Student Innovation and Entrepreneurship Training Program Project (S202410593082).

## Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Authors' contributions

Linda Liang designed the overall research framework, reviewed and revised the manuscript, secured research funding, and provided overall guidance. Linghui Liao collected and preprocessed experimental data, conducted field/laboratory work, and organized original data/reports. Jiahui Sun and Lingling Liu led in-depth data analysis and result interpretation, verified research hypotheses, and drafted the Sec. 3. Liuying Ou advised on the research methodology and tool selection, and assisted in manuscript revision (language and citation standardization). Xiaoyue Huang completed literature review and theoretical foundation construction, and summarized research background and existing studies.

## Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Acknowledgments

We would like to extend our heartfelt gratitude to all the participants who contributed to this study. We also thank Prof. Guangzheng Yu of South China University of Technology for providing the Chinese head and torso model.

## References

1. Arweiler I., Buchholz J.M. (2011), The influence of spectral characteristics of early reflections on speech intelligibility, *The Journal of the Acoustical Society of America*, **130**(2): 996–1005, https://doi.org/10.1121/1.3609258.
2. Bilzi P., Bozzoli F., Farina A. (2005), Influence of artificial mouth's directivity in determining speech transmission index, *Journal of the Audio Engineering Society*, http://www.aes.org/e-lib/browse.cfm?elib=13345.
3. Biswas R., Nathwani K., Hafiz F., Swain A. (2022), Optimal speech intelligibility improvement for varying car noise characteristics, *Journal of Signal Processing Systems*, **94**: 1429–1446, https://doi.org/10.1007/s11265-022-01815-x.
4. Bradley J.S., Reich R., Norcross S.G. (1999), A just noticeable difference in $C_{50}$ for speech, *Applied Acoustics*, **58**(2): 99–108, https://doi.org/10.1016/S0003-682X(98)00075-9.
5. Bradley J.S., Sato H., Picard M. (2003), On the importance of early reflections for speech in rooms, *The Journal of the Acoustical Society of America*, **113**(6): 3233–3244, https://doi.org/10.1121/1.1570439.
6. Brand T., Kollmeier B. (2002), Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, *The Journal of the Acoustical Society of America*, **111**(6): 2801–2810, https://doi.org/10.1121/1.1479152.
7. Cao Y. *et al.* (2022), The influence of seat occlusion on driver's binaural signal in automobile based on FEM-RTM simulation, *Journal of Physics: Conference Series*, **2185**(1): 012025, https://doi.org/10.1088/1742-6596/2185/1/012025.
8. Dal Degan N., Prati C. (1988), Acoustic noise analysis and speech enhancement techniques for mobile radio applications, *Signal Processing*, **15**(1): 43–56, https://doi.org/10.1016/0165-1684(88)90027-8.
9. Ebbitt G.L., Remtema T.M. (2015), Automotive speech intelligibility measurements, *SAE International Journal of Passenger Cars – Mechanical Systems*, **8**(3): 1120–1127, https://doi.org/10.4271/2015-01-2337.
10. Ferrari C., Cheer J., Mautone M. (2023), Investigation of an engine order noise cancellation system in a super sports car, *Acta Acustica*, **7**: 1–9, https://doi.org/10.1051/aacus/2022060.
11. GB/T 7347-1987 (1987), *The standard spectrum of Chinese speech*, National Standard of China.
12. Gerrera G.C., Donoso-Garcia P.F., Medeiros E.B. (2016), Intelligibility in low-cost automotive audio systems, *Journal of the Audio Engineering Society*, **64**(5): 320–331, http://www.aes.org/e-lib/browse.cfm?elib=18137.
13. Ghanati G., Azadi S. (2020), Decentralized robust control of a vehicle's interior sound field, *Journal of Vibration and Control*, **26**(19–20): 1815–1823, https://doi.org/10.1177/1077546320907760.
14. Granier E. *et al.* (1996), Experimental auralization of car audio installations, *Journal of the Audio Engineering Society*, **44**(10): 835–849, http://www.aes.org/e-lib/browse.cfm?elib=7882.
15. Houtgast T., Steeneken H.J.M., Plomp R. (1980), Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics, *Acustica*, **46**(1): 60–72.
16. Hu H., Xi X., Wong L.L.N., Hochmuth S., Warzybok A., Kollmeier B. (2018), Construction and evaluation of the Mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise, *International Journal of Audiology*, **57**(11): 838–850, https://doi.org/10.1080/14992027.2018.1483083.
17. Huang W., Peng J., Xie T. (2023), Study on Chinese speech intelligibility under different low-frequency

characteristics of reverberation time using a hybrid method, *Archives of Acoustics*, **48**(2): 151–157, https://doi.org/10.24425/aoa.2023.145229.

18. International Electrotechnical Commission (2011), Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index (IEC Standard No. 60268-16:2011), https://webstore.iec.ch/en/publication/1214.
19. Kleiner M., Tichy J. (2014), *Acoustics of Small Rooms*, CRC Applied Mathematics Research Press, Boca Raton, FL.
20. Liang L., Yu G. (2020), Binaural speech transmission index with spatialized virtual speaker in near field: Distance and direction dependence, *The Journal of the Acoustical Society of America*, **148**(2): EL202–EL207, https://doi.org/10.1121/10.0001808.
21. Liang L., Yu G. (2023a), The combined effect of source directivity and binaural listening on near-field binaural speech transmission index evaluation, *Applied Acoustics*, **203**: 109198, https://doi.org/10.1016/j.apacoust.2022.109198.
22. Liang L., Yu G. (2023b), Effect of speaker orientation on speech intelligibility in an automotive environment, *Applied Acoustics*, **205**: 109269, https://doi.org/10.1016/j.apacoust.2023.109269.
23. Liang L., Yu G., Zhou H., Meng Q. (2022), Effect of listener head orientation on speech reception threshold in an automotive environment, *Applied Acoustics*, **193**: 108782, https://doi.org/10.1016/j.apacoust.2022.108782.
24. Liang L., Yu L., Zhao T., Meng Q., Yu G. (2021), Speech intelligibility for various head orientations of a listener in an automobile using the speech transmission index, *The Journal of the Acoustical Society of America*, **149**(4): 2686–2697, https://doi.org/10.1121/10.0004265.
25. Meissner M. (2017), Acoustics of small rectangular rooms: Analytical and numerical determination of reverberation parameters, *Applied Acoustics*, **120**: 111–119, https://doi.org/10.1016/j.apacoust.2017.01.020.
26. Miqueau V., Parizet E., Germes S. (2024), Influence of glazing on sound quality in the car: Validation of auralizations obtained from SEA calculations, *Acoustics Australia*, **52**: 145–159, https://doi.org/10.1007/s40857-024-00315-1.
27. Parizet E. (1992), The influence of speech importance function upon articulation index computation in cars, *Noise Control Engineering Journal*, **38**(2): 73–79, https://www.ingentaconnect.com/content/ince/ncej/1992/00000038/00000002/art00003?crawler=true.
28. Parizet E. (1993), Speech intelligibility mappings in a car compartment, *International Journal of Vehicle Design (IJVD)*, **14**(2/3): 132–144, https://doi.org/10.1504/IJVD.1993.061830.
29. Rife D.D. (1992), Modulation transfer function measurement with maximum length sequences, *Journal of the Audio Engineering Society*, **40**(10): 779–790, http://www.aes.org/e-lib/browse.cfm?elib=7032.
30. Rumsey F. (2016), Automotive audio: They know where you sit, *Journal of the Audio Engineering Society*, **64**(9): 705–708, http://www.aes.org/e-lib/browse.cfm?elib=18378.
31. Samardzic N. (2014), The applicability of the objective speech intelligibility metrics for vehicle interior speech intelligibility evaluation, considering different listening configurations and background noise spectra, *SAE International Journal of Passenger Cars – Mechanical Systems*, **7**(1): 434–438, https://doi.org/10.4271/2014-01-9126.
32. Samardzic N., Novak C. (2011a), In-vehicle speech intelligibility for different driving conditions using the speech transmission index, *Noise Control Engineering Journal*, **59**(4): 397–407, https://doi.org/10.3397/1.3598380.
33. Samardzic N., Novak C. (2011b), In-vehicle application of common speech intelligibility metrics, *International Journal of Vehicle Noise and Vibration (IJVNV)*, **7**(4): 328–346, https://doi.org/10.1504/IJVNV.2011.043193.
34. Samardzic N., Moore B.C.J. (2021), Binaural speech-to-noise loudness ratio at the speech reception threshold in vehicles, *Noise Control Engineering Journal*, **69**(2): 173–179, https://doi.org/10.3397/1/376917.
35. Schroeder M.R. (1981), Modulation transfer functions: Definition and measurement, *Acta Acustica united with Acustica*, **49**(3): 179–182.
36. van Wijngaarden S.J., Drullman R. (2008), Binaural intelligibility prediction based on the speech transmission index, *The Journal of the Acoustical Society of America*, **123**(6): 4514–4523, https://doi.org/10.1121/1.2905245.
37. Visintainer A.P., VanBuskirk J.A. (1997), Effects of sound absorption on speech intelligibility in an automotive environment, *SAE*, Technical Paper 971881, https://doi.org/10.4271/971881.
38. Wang D.F., Tan G.P., Chen S.M., Jiang J.G., Su L.L. (2012), Research on speech intelligibility of sound field inside passenger car, *Advanced Materials Research*, **403–408**: 5214–5219, https://doi.org/10.4028/www.scientific.net/AMR.403-408.5214.
39. Wang Y., Yu G. (2025), Typicality analysis on statistical shape model-based average head and its head-related transfer functions, *The Journal of the Acoustical Society of America*, **157**(1): 57–69, https://doi.org/10.1121/10.0034746.
40. Warzybok A., Rennies J., Brand T., Doclo S., Kollmeier B. (2013), Effects of spatial and temporal integration of a single early reflection on speech intelligibility, *The Journal of the Acoustical Society of America*, **133**(1): 269–282, https://doi.org/10.1121/1.4768880.