Applied Acoustics 193 (2022) 108782

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Effect of listener head orientation on speech reception threshold in an automotive environment

Linda Liang [a], Guangzheng Yu [a,*], Huali Zhou [b], Qinglin Meng [a]

[a] Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510641, China
[b] College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Article history:
Received 28 January 2022
Received in revised form 27 March 2022
Accepted 8 April 2022
Available online 19 April 2022

2021 MSC:
60
40–600
190–300
200–300

Keywords:
Speech intelligibility
Speech reception threshold
Binaural listening
Automotive acoustic

ABSTRACT

The head orientation of the listener significantly affects speech intelligibility (SI) in automobiles due to the effect of binaural listening and special acoustic conditions such as early reflections and seat-back occlusions. However, this issue has not been studied with subjective tests yet. This study investigates SI with various head orientations of a listener in automobiles with an subjective experiment. The sentence speech reception thresholds (SRT) in Mandarin Chinese are measured via headphones virtually in an automobile environment, and compared with results in a weak-reflective listening room. A virtual speaker is located in the front-passenger seat, the right-back seat and the left-back seat in sequence, by convolving the target speech with corresponding binaural room impulse responses measured on a dummy head in the driver's seat with five head orientations. Result shows that the SRT variations caused by head orientations are up to 5 dB in automobile, lower than that of 9 dB in listening room. Under various head orientations, the SRT in automobile decreases as the virtual speaker moves closer to the front lateral direction, which is similar to the change rule in listening room that is mainly determined by the effect of binaural listening. Overall, a lower SRT in automobile can be obtained when the listener in the driver seat turns their head inward, i.e., towards the right. In comparison with the result in listening room, early reflections improve SI in automobile by an SRT decrease of up to 3.5 dB, while seat-back occlusions reduce SI by an SRT increase of up to 5.3 dB. The early reflections play a more significant role when the listener is in an adverse position, e.g., for the head orientations and speaker locations making direct sounds difficult to reach the listener's ears. Comparison between the SRTs based on speech transmission index (STI) and STI-SI models and the SRTs measured by subjective experiment indicates that the STI-based objective method only partially expresses the variation in SI obtained from subjective experiment, and the difference between the subjective results and the objective results is mainly caused by the inapplicability of the STI-SI models derived from the traditional room in the automotive environment. The present work is relevant to understanding the combined effect of various factors on SI under such a special acoustic condition in automobile.

1. Introduction

From the occupants' perspective, the ability to hear and understand speech sounds clearly (i.e., speech intelligibility, SI) is an important indicator of the perception of a comfortable and high-quality acoustic environment inside an automobile. However, the acoustic environment inside automobiles is complex because of the small acoustic space with complicated interior reflection boundaries, the irregular interior materials with different sound-absorption coefficients, and the effect of seat-back occlusions on sound propagation between the front and rear seats [1–4]. Therefore, it is particularly challenging to predict SI inside an automobile compared to that in traditional indoor environments.

In the small acoustical space of an automobile, sound decays rapidly (with an early decay time of less than 0.15 s) because of the highly absorptive boundaries therein and the occlusions provided by the seat-backs [1]. Thus, the detrimental effect of the temporal characteristics of the transfer function between the speaker and the listener's ears (i.e., reverberation) on SI can almost be neglected [5–8]. Accordingly, almost all the reflections from the interior boundaries reach the listener's ears earlier than 50 ms, and many useful reflections arrive within 1 ms [1,4]. Thus, almost all reflections belong to early reflections, which are considered to

* Corresponding author.
E-mail address: scgzyu@scut.edu.cn (G. Yu).

https://doi.org/10.1016/j.apacoust.2022.108782
0003-682X/

be useful for SI as they can be integrated with the direct sound [9–13]. In addition, there is a loss in the transmission of speech from the front speaker to the rear listeners because the cabin is separated into two compartments by the seat-back [3], which are found as the major contributor to the interior sound absorption inside an automobile [14]. These occlusions significantly decrease SI, especially when someone in the front-row is listening to passengers in the rear [8,14]. Moreover, the SI improvement caused by the early reflections is even more important in such adverse conditions, where the direct sound is particularly weak caused by occlusions [11].

Under such special acoustic conditions, there are many factors affecting the SI in automobile such as the locations of the speaker and the listener [15–17], various reflective interior boundaries [18], speaker directivity and orientation [19,20], and background noise [5,6]. Besides, SI in automobiles is also influenced by the effect of binaural listening, such as head shadow and binaural interaction, that is highly dependent on the listener's head orientation or the speaker's azimuth relative to the listener [21–25]. For instance, head shadow may result in an signal-to-noise ratio (SNR) that is higher at one ear than the other. In automobiles, the influence of the listener's head orientation on SI is more complex than the situation in a traditional room. On the one hand, the small dimensions of the interior of an automobile lead to near-field (i.e., when the distance is not more than 1.0 m) acoustic conditions, enhancing the head shadow [24]. On the other hand, the benefit of early reflections and the detrimental effect of the seat-back occlusions inside the automobile also vary with the listener's head orientation. The combined effect of the special acoustic conditions in the automobile and binaural listening under various head orientations influences SI considerably. However, in previous studies, the listener's head has usually been assumed to be facing forward when recording binaural signals [3,5,6]).

In our previous work [26], we just adopted the speech transmission index (STI) [27] and the Chinese STI-SI model [28] to investigate and predict SI for various listener head orientations in an automobile. However, the binaural STI (BSTI) that using the better ear may disregard the advantages of binaural hearing in SI, and the STI-SI model obtained in the large acoustical room may be unsuitable for the small acoustical room of an automobile. Admittedly, STI is a widely accepted objective evaluation metric for predicting and evaluating the loss of SI. It can reflect the influence of reverberation and SNR on SI and has a certain ability to account for nonlinear distortion [27]. However, STI is essentially a monaural model and simply adopts the "better ear" of two separated measurements, following the recommendation in IEC 60268-16 [27]. 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 [25,29]. Therefore, it maybe not accurate enough to evaluate SI based on binaural hearing with STI simply using the better ear approach. Also, the Chinese STI-SI model [28] was obtained from tests conducted in four traditional rooms with dimensions far greater than that of an automobile. However, the STI-SI model is highly dependent on the subjective and objective evaluation methods, as well as the test conditions [28,30]. Consequently, the STI-SI model derived from traditional rooms may not be applicable to the situation in automobiles. To the best of our knowledge, there has been no research on the STI-SI model in automobiles. Thus, the prediction of SI using STI-SI models in our previous work [26] may includes errors due to not only the applicability of the STI-SI model but also the difference between STI and BSTI.

Furthermore, given that the inapplicability of the STI-SI models derived from traditional room in the small room of an automobile, there is still a lack of investigation on the authentic influence of listener's head orientation on the SI based on subjective speech reception threshold (SRT, i.e., the SNR for 50% intelligibility) test. This motivates us to systematically explore how a listener's head orientation affects their SI in an automobile via authentic subjective SRT test, and to examine the particularity of the interaction between early reflections or seat-back occlusions and binaural listening on SI. Above all, we expect to examine the difference between the result from subjective experiment and from STI-based objective method in the small acoustical space of an automobile.

The SRT experiment was conducted via headphones virtually in an automobile environment, rather than SI score test with fixed SNRs, due to the large variation range of SNR between different conditions. The virtual speaker was located in either the front passenger seat (FP-seat), right-back seat (RB-seat), or left-back seat (LB-seat) by convolving the target speech with corresponding binaural room impulse responses (BRIRs) measured on a dummy head, that is in the driver's seat (D-seat) and with five head orientations. To further distinguish the effect of special acoustic conditions and the binaural listening on SI, the SRT experiment was repeated in a virtual listening room with negligible reflection conditions using the same position configurations as in the automobile. The psychoacoustic experiment SRT results were also compared with those from objective metrics, i.e., based on five typical STI-SI models in traditional rooms [28,30–33], and their performance in automotive SI assessment was discussed.

2. Methods

2.1. BRIR measurement

To distinguish the effect of reflections and the binaural listening on SI, BRIRs were measured under two separate conditions: (i) in an automobile and (ii) in a listening room with negligible reflections. These BRIRs were used for the synthesis of binaural test signals for subjective experiment in Section 2.2 and the calculation of the objective metrics in Section 2.3.

Measurements (i) were carried out in a left-hand-drive automobile [Toyota Highlander, with dimensions of 489 cm × 193 cm × 172 cm (length × width × height)] in the quiet environment of a university campus. Measurements (ii) were conducted in a listening room of size 8 m × 6 m × 3.5 m, with background noise lower than 30 dBA and negligible reflections (the reverberation time was less than 0.15 s after sound absorption and isolation), in compliance with the International Electronic Standard. The speaker was placed on a pallet fixed on a tripod and covered with sound-absorbing cotton to reduce reflection. An appropriate time window was used for the BRIR to further eliminate the influence of room reflections from the tables and shelves in the listening room.

The speaker-to-listener positional relationships were consistent in the two measurements and were determined by the configuration of the seats in the automobile, as shown in Fig. 1. The listener was fixed in the D-seat and horizontally rotated his/her head by various amounts, corresponding to head orientations (denoted by φ) of -60°, -30°, 0°, 30°, and 60°. The negative or positive head orientation corresponds to the head turning to the right or the left, respectively. The speaker was arranged in each of the three passenger seats in sequence. When the speaker was placed in the FP-seat, i.e., directly to the right of the listener, the distance from the head center of the listener was 77.9 cm. When the speaker was on the RB-seat, the distance from the head center of the listener was 107.2 cm, and the angle to the right of the listener was 43°. When the speaker was on the LB-seat, i.e., directly behind the listener, the distance from the head center of the listener was 81.3 cm. The speaker always faced forward with respect to the listener. The speaker and the two ears of the listener were arranged on the same

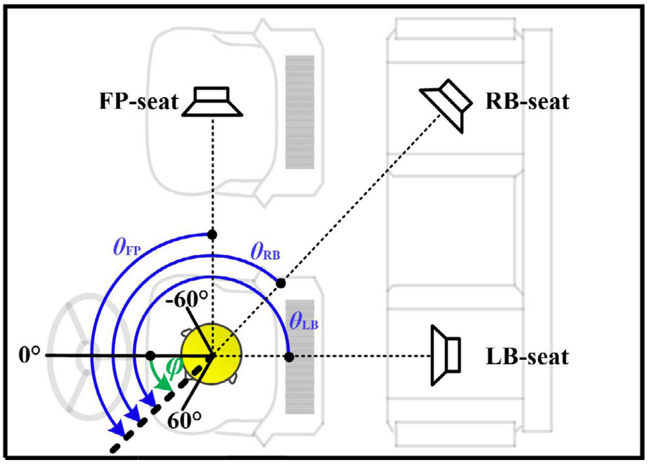

Fig. 1. Speaker-to-listener positional relationship in the measurements in the automobile and the listening room. The speakers are arranged in each of the three passenger seats in sequence, i.e., FP-seat, RB-seat, and LB-seat respectively. The gray diagrammatic presentations are only appropriations for the automobile environment. Negative or positive head orientation (φ) expresses the head turning to the right or the left, respectively. θ_{FP}, θ_{RB}, and θ_{LB} are the azimuths of the speaker relative to the listener's head with various head orientations when the speaker is in the FP-seat, RB-seat, and LB-seat, respectively (modified appropriately based on the Fig. 1 of Ref. [26]).

plane at a height of 115.1 cm above the automobile floor. This was set with respect to the height of the ear-canal entrance of the average Chinese male (170 cm tall).

To facilitate discussion of the effect of binaural listening, we introduced another additional variable, that is, the azimuth of the speaker relative to the listener's head (θ), as indicated by the blue line in Fig. 1. Azimuths of $0^\circ, 90^\circ, 180^\circ$, and 270° represent the front, right, rear, and left directions, respectively. In fact, θ varied with the head orientation φ, as the actual azimuth θ is the sum of the azimuth of the speaker relative to the listener's head when the listener's head is facing forward (i.e., $90^\circ, 133^\circ$ and 180° for speakers in the FP-seat, RB-seat, and LB-seat, respectively) and the specific head orientation value of the listener. Thus, the azimuth θ ranged from $90 \pm 60^\circ(\theta_{FP})$, $133 \pm 60^\circ(\theta_{RB})$, and $180 \pm 60^\circ(\theta_{LB})$ with various head orientations φ when the speaker was in the FP-seat, RB-seat, and LB-seat, respectively (Fig. 1).

Excitation was induced by a logarithmic sweep signal at a sampling frequency of 44.1 kHz and 24-bit quantization for 3 s over a frequency range of up to 20 kHz. This was fed to a loudspeaker (8010A; Genelec, Finland) after passing through a digital-to-analog converter in a sound card (Fireface UC; RME Audio, Germany). Binaural recordings of the logarithmic sweep signals were then deconvolved with the inverse of the original logarithmic sweep signal to obtain noise-free BRIRs. Details of the signal processing method can be found in Majdak [34]. Past work has indicated that binaural responses recorded with a Knowles Electronic Manikin for Acoustic Research (KEMAR) can achieve SI comparable to that of recordings of an individual subject [35]. Therefore, the KEMAR was regarded as the virtual "listener". The binaural signals were recorded by a pair of miniature microphones (DPA 4060; Danish Pro Audio, Denmark; frequency range: 20 Hz to 20 kHz) inserted into blocked entrances of the KEMAR ear canals. Finally, we obtained BRIRs under 30 conditions, consisting of 2 measured conditions $\times$ 3 speaker locations $\times$ 5 head orientations.

2.2. SRTs from subjective experiment

Subsequently, a subjective experiment measuring the SRTs for 50% intelligibility was administered. The measured BRIRs were used to generate speakers at different seats with different head orientations in an automobile and a listening room.

2.2.1. Participants

Ten participants (four male and six females, aged 22–26 years with a mean age of 24 years) were recruited from the graduate students in South China University of Technology. All of them self-reported normal hearing, and all are native Mandarin speakers who were born and grew up in different provinces of China. The listeners were paid for their participation. SRT measurements under all of 30 conditions were performed for each subject.

2.2.2. Stimuli

Sentences from the Mandarin Chinese matrix sentence test (CMNmatrix) [36] were used as the target speech. All sentences have a fixed syntactical structure with five words (*name-verb-number-adjective-object*), which is semantically unpredictable and grammatically correct. The 20-sentence lists were used, of which there are 40 lists in total. The target speech was virtually located at different speaker locations (FP-seat, RB-seat, LB-seat) by using Auralization technology, i.e., convolving the target speech with the corresponding BRIR obtained in Section 2.1.

The interferer was obtained by generating a pink noise signal and then filtering and adjusting the spectrum according to GB/T 7347 [37], the standard spectrum of Chinese speech. The interferer was processed through audio signal decorrelation so that the noise field in the automobile (from the engine, road excitation, wind noise, etc.) would be diffuse and contain no spatial auditory information [3,38]. If the same interferer signals were fed to the ears directly, it would affect the head center location or synthesize a frontal sound image. However, the introduction of spatial auditory information from the interferer can lead to the spatial unmasking phenomenon, which is not conducive to analyzing the effect of head orientation on SI alone. Therefore, we must eliminate the spatial auditory effect of the interferer as much as possible. Audio signal decorrelation creates two or more replicas of the input signal that have different waveforms, but are perceived as being similar to the original signals in most aspects except some spatial auditory effects [39]. Xie [40] proposed an audio signal decorrelation method based on reciprocal-maximal length sequence (MLS) filters, taking advantage of the uniform power spectrum and low-valued cross-correlation of reciprocal MLSs [41]. This method creates a pair of output signals with low cross-correlation, but almost identical magnitude spectra to the input signal. In the current study, this approach was adopted to obtain decorrelated binaural interferer signals through the convolution of single-channel interferer with a pair of reciprocal MLSs, which sound like they have no spatial auditory information but produce a certain sense of envelopment.

The level of the interferer was fixed at around 60 dBA in each ear, which ensured a comfortable level. The level of the target speech was adjusted before convolution with a BRIR to obtain different SNRs. The binaural signals were obtained by adding the decorrelated binaural interferer to the convolved binaural target speech.

2.2.3. Procedure

The experiment was conducted in a sound-treated room with a background noise level lower than 30 dBA. Stimuli were presented via headphones (Sennheiser HD800) driven by an RME sound card (Germany, FIRFACE UC). SRTs were measured using an adaptive procedure [42] and the initial SNR was set to 10 dB. It is worth mentioning that the SNRs in current work do not represent the actual ones at the ear canal entrances of the listener, but the difference in level between the target speech before convolution and the interferer, and the same is true for the acquisition of SRT from objective metric in Section 2.3. Tests were conducted in an open-set response format, in which participants were instructed to orally repeat the sentence and guess if necessary, and an experimenter marked the correctly repeated words on a graphical user interface in MATLAB.

There were 30 test conditions in total, and each condition was tested with two repetitions, making a total of 60 runs for each participant. The 30 conditions included two measured conditions (in the automobile and in the listening room), three speaker locations (FP-seat, RB-seat, and LB-seat), and five listener head orientations ($-60^{\circ}, -30^{\circ}, 0^{\circ}, 30^{\circ}$, and 60°). The SRTs of the two repetitions were averaged to give the final SRT. As the list number (40) is less than the run number (60), some lists were used twice for the same listener. This will not have affected the result because this corpus is designed to be semantically unpredictable and suitable to be used twice with the same listener. The order of the test conditions and lists was randomized across listeners. To reduce listener fatigue, the 60 runs (each taking 4–6 min) were divided into two sessions that were at least 12 h apart, and participants took a 15-min break after every 8 consecutive runs. Training was conducted before each session.

2.3. SRTs from objective metric

The BSTI were calculated according to IEC 60268-16 [27] and based on the "better ear" rule (i.e., selecting the larger STI), as described by Liang [26]. The background noise was obtained by generating a pink noise signal and then filtering and adjusting the spectrum according to GB/T 7347 [37], the standard spectrum of Chinese speech. A monaural speech samples was generated from the speech material according to Hu [36], by 30-fold superimposing all synthesized sentences in CMNmatrix. We then obtained binaural speech signals from the speaker at different positions and for various listener head orientations by convolving the monaural speech sample with the corresponding BRIRs obtained in Section 2.1. Because sound waves decay rapidly in an automobile, the impact of the temporal characteristics of the transfer function (i.e., reverberation) on the BSTI should be weak, and so its effect on BSTI can be ignored in automobile applications under different conditions [5,6,8,26]. Consequently, the BSTI calculation was simplified to consider only SNR effects [8,24,26].

According to the Chinese STI-SI model, the subjective listening SI scores were estimated indirectly after the objective metric, i.e., the BSTI, had been determined. Under various conditions and Chinese STI-SI models, the SI scores were obtained under different SNRs successively, then we found the SNRs that produced a SI score closest to 50%, that is the SRT from objective metric, with resolution of 0.1 dB.

3. Results and discussion

3.1. Subjective SRT result in listening room

Fig. 2 shows the subjective SRT result for different speaker locations in the listening room, together with SRT distribution for each participant and the 95% confidence intervals (CIs). As shown in Fig. 2, the subjective SRT clearly varies with the head orientation of the listener in the listening room condition, where the effect of binaural listening plays a dominant role in SNR and SI. Note that, since a higher SRT value represents a poorer SI performance, all graphs in present study adopt the inverse coordinates for SRT.

With speaker in the FP-seat, the subjective SRT reaches a minimum of -14.7 dB when head orientation $\varphi = -30^{\circ}$ (i.e., azimuth of 60°), and reaches a maximum of -7.9 dB when $\varphi = 60^{\circ}$ (i.e., azimuth of 150°) (Fig. 2). With speaker in the RB-seat, the subjective SRT in the listening room increased with head orientation (azimuth) due to the effect of binaural listening, with a maximum of -6.9 dB when $\varphi = 60^{\circ}$ (i.e., azimuth of 193°) and a minimum of -15.8 dB when $\varphi = -60^{\circ}$ (i.e., azimuth of 73°) (Fig. 2). With speaker in the LB-seat, the subjective SRT reaches a maximum of -5.8 dB when $\varphi = 0^{\circ}$ (i.e., azimuth of 180°), and a minimum of -13.1 dB when $\varphi = -60^{\circ}$ (i.e., azimuth of 120°) (Fig. 2). These phenomena are consistent with the results of Liang and Yu [24] as well as Liang [26], namely the better SI (i.e., lower SRT value) appears at the speaker azimuth of the front lateral direction (i.e., azimuth of 60° to 80°), whereas lower SI (i.e., higher SRT value) is recorded when the virtual speaker is located in anterior or posterior positions. This phenomenon can be due to stronger reflection from the head and ipsilateral ear of the listener with a lateral speaker and the head shadow effect with an anterior or posterior speaker [24], especially in the near-field, which is similar to the spatial characteristics of the near-field head-related transfer function [43] and the perceptual difference in binaural loudness with different source directions [44].

Additionally, a two-way repeated-measures analysis of variance (rm-ANOVA) showed that the speaker location, head orientation and the interaction between both factors all had a significant effect on subjective SRT in listening room [$F(2,18) = 637.9, F(4,36) = 598.6$ and $F(8,72) = 196.8$, with $p < 0.0001$]. *Post hoc* pairwise comparisons applying Bonferroni corrections showed insignificant differences of the subjective SRT between the FP-seat and LB-seat when $\varphi = -60^{\circ}$, and between the FP-seat and RB-seat when $\varphi = -30^{\circ}$ ($p > 0.1$). The comparisons also showed that there were insignificant differences of the subjective SRT between head orientation of -30° or -60° and 0° when the speaker is in the FP-seat ($p > 0.06$); there was insignificant difference of the subjective SRT between head orientation of -30° and -60° when the speaker is in the RB-seat ($p = 0.5$); there were significant differences of the subjective SRT among each head orientation when the speaker is in the LB-seat ($p < 0.0001$). We can also find corresponding clues in Fig. 2 for these results.

In fact, the position of the passenger seats in an automobile is fixed, with different azimuth angles and distances among them. Thus, SI involving different speaker locations may benefit from the effect of binaural listening to a varying extent depending on the azimuth of the speaker relative to the listener's head. Note that, as determined by the effect of binaural listening under various speaker locations in the listening room, the worst SI performance (i.e., highest SRT value) occurs when the speaker is in the LB-seat, except when $\varphi = 60^{\circ}$ which achieves a best performance (i.e., lowest SRT value) (Fig. 2). The subjective SRT when the speaker is in the FP-seat is lower than those when the speaker is in the RB-seat at head orientations except when $\varphi = -60^{\circ}$ (Fig. 2). In the listening room, the average and standard deviation (SD) of the subjective SRT among different head orientations for the speaker in the FP-seat, RB-seat, and LB-seat were -12.3 ± 2.6 dB, -12.1 ± 3.8 dB, and -9.6 ± 2.9 dB, respectively. This indicates that, on average, the dominant effect of binaural listening means a listener in the D-seat achieves optimal SI when the speaker is in the FP-seat, and the worst SI occurs when the speaker is in the LB-seat. It is worth mentioning that, caused by the distance attenuation factor (the $1/r$ raw, i.e., the sound pressure generated by a point source with constant power is inversely proportional to the distance between the sound source and the receiver in a free field), the SNRs in listener's ears with speaker in RB-seat and LB-seat are additionally reduced by 2.77 dB and 0.37 dB respectively, compared with that in FP-seat. If the distance attenuation factor is excluded, the azimuth-dependent effect of binaural listening means a listener in the D-seat location achieves optimal SI when the speaker is in the RB-seat when the reflections are weak or almost non-existent.

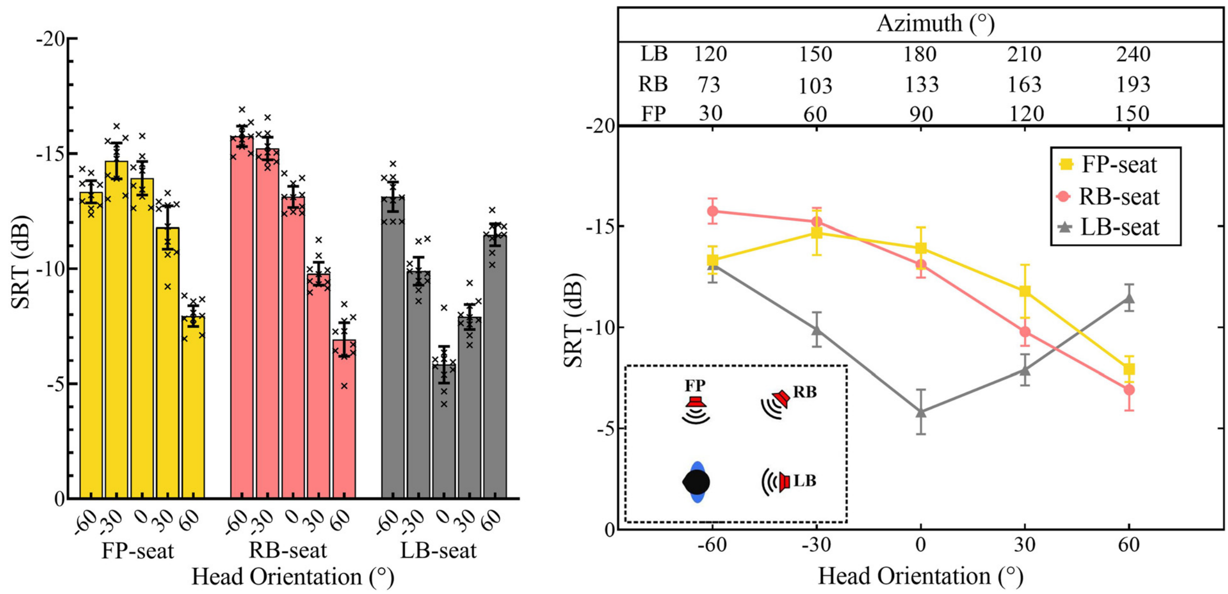

Fig. 2. The SRT results under listening room condition at various head orientations and speaker locations (FP-seat, RB-seat and LB-seat). The left graph shows the SRT distribution for each participant (×) as well as the mean (bar) and the 95% CIs (the bottom and top edges) across participants. The right graph shows the means and standard deviations of the SRTs across participants. The upper label on the x axis gives the azimuths of speaker relative to listener's head when the speaker is in FP-seat, RB-seat, and LB-seat, corresponding to various head orientations. Inset shows a schematic diagram of the location relationship between speaker and listener, and the dotted line indicates that there is almost no reflections in listening room.

3.2. Subjective SRT result in automobile

Fig. 3 shows the subjective SRTs result in automobile for various listener head orientations (or azimuths) when the speaker is in different locations, together with SRT distribution for each participant and the 95% CI.

For speaker in FP-seat, the minimum subjective SRT (−15 dB), corresponding to the best SI, appears when head orientation $\varphi = -30^{\circ}$, and increasing or decreasing the head orientation causes the subjective SRT to increase by more than 3 dB (Fig. 3). This demonstrates that, when the speaker is in the FP-seat, it is detrimental to SI for listeners in the D-seat to turn their heads left to observe their surrounding environment or to communicate with passengers. This result is somewhat different from the STI result in Liang [26], where there are no obvious differences among the STIs at head orientaions below 0°.

For speaker in RB-seat, the subjective SRT as a function of head orientation varied from −10.5 dB (when $\varphi = -60^{\circ}$) to −5.9 dB (when $\varphi = 60^{\circ}$) in the automobile condition (Fig. 3). This reveals that with speaker in RB-seat, the driver turning their head toward the listener (i.e., to the right or decreasing head orientation) appears to enhance SI compared with facing toward the front of the automobile, with an enhancement in SI equivalent to an effective SNR of more than 4.5 dB.

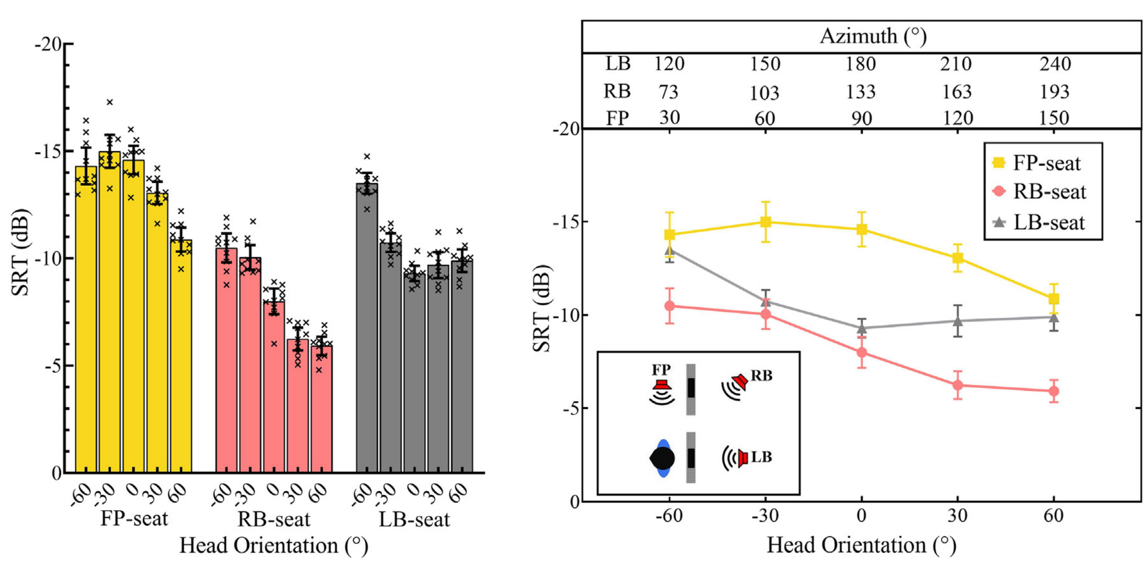

Fig. 3. The SRT results under the automotive condition at various head orientations and speaker locations (FP-seat, RB-seat and LB-seat). The left graph shows the SRT distribution for each participant (×) as well as the mean (bar) and the 95% CIs (the bottom and top edges) across participants. The right graph shows the means and standard deviations of the SRTs across participants. The upper label on the x axis gives the azimuths of speaker relative to listener's head when the speaker is in FP-seat, RB-seat, and LB-seat, corresponding to various head orientations. Inset shows a schematic diagram of the location relationship between speaker and listener.

For speaker in LB-seat, the maximum subjective SRT (about -9.3 dB) in the automobile appears when $\varphi = 0^\circ$, which is determined by the effect of binaural listening [24], and then decreasing or increasing the head orientation reduces the SRT (Fig. 3). There is no doubt that with speaker in LB-seat, when the listener in the D-seat turns their head, either left or right, SI is considerably enhanced. For instance, the listener turning their head to the right by 60° (head orientation of -60°) reduces the SRT by 4.2 dB.

In addition, a two-way rm-ANOVA showed that the speaker location, head orientation and the interaction between both factors all had a significant effect on subjective SRT [$F(2,18) = 812.3, F(4,36) = 184.1$ and $F(8,72) = 38.1$, with $p < 0.0001$]. Note that, compared with in listening room as Section 3.1, the significance level of SRT difference in automobile for different speaker locations increases, while that for different head orientations reduces. *Post hoc* pairwise comparisons applying Bonferroni corrections showed significant differences of the subjective SRT among different speaker locations for each head orientation ($p < 0.05$). The comparisons also showed that there were insignificant differences of the subjective SRT among head orientation of $0^\circ, -30^\circ$ and -60° when the speaker is in the FP-seat ($p > 0.09$); there were insignificant differences of the subjective SRT between head orientation of 30° and 60° and between head orientation of -30° and -60° when the speaker is in the RB-seat ($p > 0.8$); there were insignificant differences of the subjective SRT among head orientation of $0^\circ, 30^\circ$ and 60° when the speaker is in the LB-seat ($p > 0.2$). We can also find corresponding clues in Fig. 3 for these results.

As shown in Fig. 3, in automobile condition, the subjective SRTs when the speaker is in the RB-seat are worse than when the speaker is in the FP-seat or the LB-seat, even though the binaural listening has a positive impact on SI when $\varphi < 0^\circ$. In automobile condition, the means and SDs of subjective SRT for different head orientations when the speaker is in the FP-seat, RB-seat, and LB-seat were -13.6 ± 1.6 dB, -8.1 ± 2.1 dB, and -10.6 ± 1.7 dB, respectively. Note that, the mean subjective SRT when the speaker is in the LB-seat, which is an adverse condition in terms of the effect of binaural listening (Fig. 2), is lower than when the speaker is in the RB-seat in automobile, unlike the situation in listening room condition. In similar to the situation in listening room in Fig. 2, the SI when the speaker is in the FP-seat is better than when the speaker is seated elsewhere.

3.3. Comparison of the subjective SRT result in different measured conditions

As forementioned, for the automotive environment with dense early reflections from the interior boundaries and seat-back occlusions [1,2,4], the subjective SRT results are different from those obtained in the listening room. The combined effect of the special acoustic conditions in the automobile and diffraction around the listener's head (i.e., the effect of binaural listening) influences SI considerably.

Fig. 4 shows the subjective SRT results under two measured conditions when the speaker is in FP-seat, RB-seat and LB-seat respectively. Table 1 shows the statistical SRT results under different measured conditions and speaker locations. As shown in Table 1, the maximum and minimum SRTs in automobile always appear at the same head orientations as those in listening room, even though their values are different. Actually, the trends of SRT variation with head orientation in automobile are similar to the results in the listening room condition if we neglect small differences between them, that is, under various head orientations, the closer the speaker is to front lateral direction of the listener's head (i.e., about azimuth of $60^\circ - 80^\circ$), the smaller the SRT value (i.e., greater SI) will be, as shown in Fig. 4. These results reveal that even though the early reflections from interior boundaries or seat-back occlusions affect SI in automobile considerably, they are still not sufficient to subvert the trend of SI variation under different listener head orientations that is mainly determined by the effect of binaural listening [24–26]. In fact, as described by Arweiler and Buchholz [12], the direct sound energy is always more efficient than the early reflection energy. Thus, even though the early reflection energy can improve SI considerably, it cannot replace the dominant role of direct sound, regardless of whether the direct sound is partially obscured by seat-backs. These results are almost consistent with those of Liang [26], but there are still some differences. The comparison between them will be discussed in Section 3.4.

In comparison with the negligible reflections of the listening room, the effect of the early reflections acts to decrease the SRT values in the automobile. In practice, the extremely small volume of the enclosed automobile means that most reflections arrive within 40 ms and many useful reflections will arrive within 1 ms [1,4]. A delay of within 30 ms indicates that the signals are fully integrated with the direct sound in terms of energy, resulting in an enhancement in the effective SNR, as observed by [9]. Almost all reflections are early reflections, which are useful for improving SI. Early reflections from the boundaries inside the automobile are distributed non-uniformly in the automotive space, so there is no doubt that SI was improved by a varying extent of SRT decrease for various speaker locations and head orientations. Fig. 5 demonstrates the subjective SRT difference between the situation in listening room and automobile, expressed by the subjective SRT in listening room subtracting those in automobile.

Note that, due to the advantages of early reflections, SI when the speaker is in the FP-seat is enhanced in automobiles than in the listening room especially for positive head orientations (i.e., listener turning head to left) (see Fig. 4 and Fig. 5), where the listener is at a disadvantage in terms of the effect of binaural listening, compared with negative head orientations. For instance, the early reflections reduced the subjective SRT by 2.9 dB when head orientation $\varphi = 60^\circ$, while reduced by lower than 1 dB when $\varphi < 0^\circ$ (see Fig. 5). This can be attributed to the fact that, when $\varphi = 60^\circ$, the ears of the listener in the D-seat are more readily exposed to the early reflections in the automobile, because the listener is facing toward the left front window, which acts as the main reflector. Additionally, the early reflections are essential to achieving adequate intelligibility when the direct sound is particularly weak [11], such as when the listener's head is turned away from the speaker, exactly as when the listener's head orientation is 30° or 60° with the speaker in the FP-seat. Table 1 also shows that the mean as well as the maximum and minimum SRTs measured in the automobile with speaker in FP-seat are lower than those in listening room. Two-way rm-ANOVA showed that the measured condition, head orientation and the interaction between both factors all had a significant effect on subjective SRT [$F(1,9) = 145.4, F(4,36) = 244.7$ and $F(4,36) = 10.1$, with $p < 0.0001$]. *Post hoc* pairwise comparisons applying Bonferroni corrections showed insignificant differences of the subjective SRT between the measured conditions of the automobile and listening room when $\varphi = -30^\circ$ and $\varphi = 0^\circ$ ($p > 0.2$).

In addition to the beneficial early reflections, adverse conditions such as seat-back occlusions also occur in automobiles, which affect the sound propagation from the speaker to the listener's ears. In practice, speech communication between front and rear passengers is a well-known problem [8,45]. When the speaker is in the rear passenger seats (e.g., RB-seat and LB-seat), the sound waves propagating from the speaker to the listener's ears are partially blocked by the seat-backs, especially for the ear closer to the speaker (i.e., ipsilateral ear), which benefits more from direct sounds [14,26]. The SNR of the ipsilateral ear appears to be the

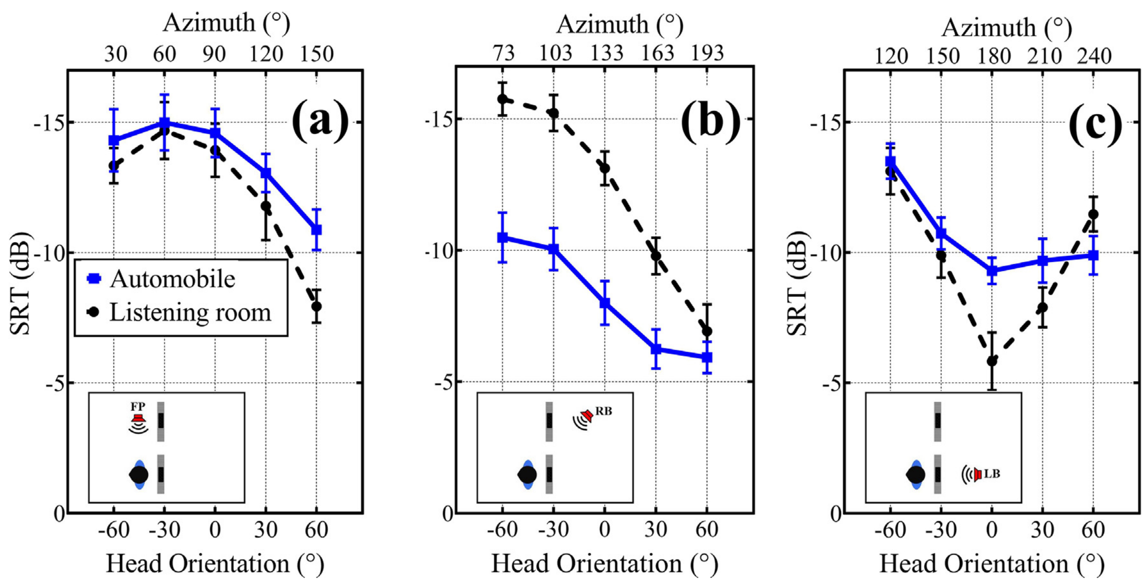

Fig. 4. Subjective SRT results at various HOs under different measured condition when the speaker is in (a) FP-seat, (b) RB-seat, and (c) LB-seat) respectively. Insets show schematic diagrams of the location relationship between speaker and listener.

Table 1
Statistical SRT results under different measured conditions and speaker locations.

Speaker location	Condition	Maximum		Minimum		Range/dB	Mean/dB	SD/dB
		SRT/dB	φ /°	SRT/dB	φ /°			
FP-seat	Listening room	−7.9	60	−14.7	−30	6.8	−12.3	2.6
	Automobile	−10.9	60	−15	−30	4.1	−13.6	1.6
RB-seat	Listening room	−6.9	60	−15.8	−60	8.9	−12.1	3.8
	Automobile	−5.9	60	−10.5	−60	4.6	−8.1	2.1
LB-seat	Listening room	−5.8	0	−13.1	−60	7.3	−9.6	2.9
	Automobile	−9.3	0	−13.5	−60	4.2	−10.6	1.7

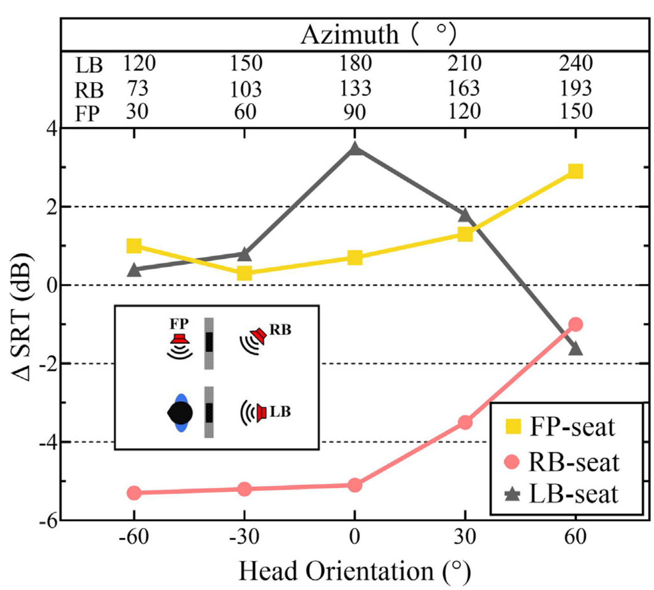

Fig. 5. The SRT difference between the situation in listening room and automobile for various head orientations and speaker locations, expressed by the subjective SRT in listening room subtracting those in automobile. The upper label on the x axis gives the azimuths of speaker relative to listener's head when the speaker is in FP-seat, RB-seat, and LB-seat, corresponding to various head orientations respectively. Inset shows a schematic diagram of the location relationship between speaker and listener.

main influence in determining the in-automobile SI performance when the binaural difference in the speech sound pressure level is significant [15,26]. Thus, seat-back occlusions reduce the effective SNRs for SI.

As can be seen in Fig. 4, caused by the seat-back occlusions, the subjective SRTs in automobile are higher than those in listening room when the speaker is in the RB-seat. The difference between the two measured conditions (up to 5 dB when $\varphi = -60°$) decreases with increasing head orientation when the speaker is in the RB-seat, as shown in Fig. 5. This may be because the listener's ears move toward the closer reflective boundaries as they turn their head away from the speaker (therein increasing the head orientation), thus exposing their ears to more early reflections from these boundaries, which compensates for the adverse occlusions. Similarly, Liang [26] found that the reflections are helpful to the SNR in the 8 kHz octave band, with an enhancement of up to 3 dB when the listener turns their head to the left (at positive head orientations) with the speaker in the RB-seat. This also confirms the idea that the early reflections are more helpful in enhancing SI under adverse conditions when direct sounds cannot easily reach the ears [11]. Two-way rm-ANOVA showed that the measured condition, head orientation and the interaction between both factors all had a significant effect on subjective SRT [$F(1,9) = 2365, F(4,36) = 537.7$ and $F(4,36) = 46.8$, with $p < 0.0001$]. *Post hoc* pairwise comparisons applying Bonferroni corrections showed significant differences of the subjective SRT between the measured conditions of the automobile and listening room for each head orientation ($p < 0.004$). Table 1 also shows that the mean as well as the maximum and minimum SRTs measured in the automobile with speaker in RB-seat are higher than those in listening room.

In particular, when the speaker is in the LB-seat, the in-automobile SIs are better than those in listening room benefiting from the early reflections from boundaries inside the automobile (see Fig. 4), even if sounds traveling from the speaker to the ears of the listener in the D-seat are blocked by the seat-backs. This is contrary to the case of the speaker in the RB-seat. In fact, when the speaker is in the LB-seat behind the listener, the early reflections from very close boundaries (e.g., the left front window) compensate for the adverse occlusions in the SNRs, as reported by Liang [26] and Viktorovich [46]. Liang [26] also found that the early reflections from the left front window, which is very close to the listener, play a dominant role relative to the overall early reflections, and this is helpful in enhancing SI when the speaker is in the LB-seat and the listener is in the D-seat. Under such adverse conditions in terms of the effect of binaural listening and seat-back occlusions, in which the direct sounds are blocked by head, auricles and seat-backs, the effect of the early reflections seems to become indispensable [11]. The existence of the early reflections enhances SI inside the automobile, with the subjective SRT decreasing by up to 3.47 dB when $\varphi = 0^\circ$ (see Fig. 5). Table 1 also shows that the mean as well as the maximum and minimum SRTs measured in the automobile with speaker in LB-seat are lower than those in listening room. It seems that increasing the reflective boundaries inside the automobile as a means of increasing the early reflections is conducive to interpersonal communication inside the automobile, especially when front-row listeners are listening to the rear-row speakers. Two-way rm-ANOVA showed that the measured condition, head orientation and the interaction between both factors all had a significant effect on subjective SRT [$F(1,9) = 45.8, F(4,36) = 407.1$ and $F(4,36) = 68.6$, with $p < 0.0001$]. *Post hoc* pairwise comparisons applying Bonferroni corrections showed insignificant differences of the subjective SRT between the measured conditions of the automobile and listening room when $\varphi = -60^\circ$ ($p = 0.48$).

In general, the SDs of SRT among the different head orientations are lower than in the listening room condition. This indicates that the SI variation with head orientation is smaller in the automobile than in the listening room (see Table 1), because the early reflections compensate the SNR when the listener is in an adverse conditions in term of the effect of binaural listening [24,26].

3.4. Comparison of the SRT from subjective experiment and objective metric

In previous work [26], we calculated the BSTI and then indirectly calculated the SI under various head orientations based on the BSTI and the Chinese STI-SI model from Chinese phonetically balanced (PB) work test of Zhu [28]. However, the STI-SI model is affected by the speech test materials, languages, presentation way for test signals, subjects and listening conditions etc [28,31]. Table 2 lists several STI-SI models for different types of speech test materials, languages, presentation ways and listening conditions, and their curves are shown in Fig. 6. These models consist of the Chinese rhyme test curve derived by Peng [30], Chinese PB words test curves by Peng [31] and Zhu [28], English rhyme test curve by Bradley [32] and English PB words test curve by Anderson [33].

As shown in Fig. 5, the differences between the curves for the PB words test and rhyme test are considerable, and there is relatively small difference among the curves for PB words test with different language and listening conditions. Obviously, the influence of speech test materials on STI-SI model is extremely important. However, to the best of our knowledge, the results of the CMNmatrix test [36] that differs from PB words test and rhyme test have never previously been compared with any of the objective metrics, nor was the STI-SI model established. Also, the existing curves are based on measurements and tests in traditional rooms with much larger volume than automobile (see Table 2), which may not be applicable to the much smaller size and complex sound field of an automotive space. For these reasons, there will be a mismatch between the SIs based on the Chinese STI-SI models and that based on subjective experiment.

As a comparison, the STI-SI models listed in Table 2 were used in this study to obtained the SRTs from objective metric, as the description in Section 2.3. Fig. 7 plots the SRT results based on different STI-SI models at various head orientations and measured conditions when the speaker is in FP-seat, RB-seat and LB-seat respectively. As can be observed in Fig. 7, the SRT results under different STI-SI models are shown as the overall fluctuation of the curve, that is, the overall law for the SRT under different conditions and head orientations remains unchanged, but the specific values of the SRT are affected by the STI-SI models. As shown in Fig. 7, the SRT based on different STI-SI models only partially expresses the variation rule in SRT obtained from subjective experiment as in Fig. 4. For instance, with speaker in LB-seat, the SRTs in automobile based on STI-SI models of Peng [30] and Bradley [32] slightly increase with increasing head orientation when $\varphi > 0^\circ$ (see Fig. 7), while there is insignificant difference of the subjective SRT among head orientation of $0^\circ, 30^\circ$ and 60°, even a slight decline with increasing head orientation, as shown in Fig. 4 and described in Section 3.2. Or when the speaker is in RB-seat, the SRT in automobile based on the STI-SI model of Zhu [28] when $\varphi = 60^\circ$ is lower than that when $\varphi = 30^\circ$ (see Fig. 7), unlike the results based on other STI-SI models or the subjective SRT result. This is probably because the advantages of binaural hearing in SI are disregarded under "better ear" rule.

However, there are distinct differences among the SRT results from different STI-SI models in term of the absolute value, as shown in Fig. 7. This is mainly caused by the choice of the STI-SI model. To intuitively explain the performance of different STI-SI models, Fig. 8 shows the relationship between the SRT result from subjective experiment ($SRT_{subjective}$) and based on different STI-SI models ($SRT_{objective}$) based on linear regression. The linear regression models for the SRT result from subjective experiment and based on different STI-SI models are shown in Table 3. Note that, the correlation coefficient of the STI-SI models of Peng [30] and Bradley [32] for rhyme test are higer than others (i.e., 0.89 and

Table 2
The list of the STI-SI models for different types of speech test materials, languages, presentation ways and listening conditions.

Model	Language	Material	Relationship	Presentation way	Listening condition	
					Volume/m^3	T_{30} /s
Peng, 2008 [30]	Chinese	Rhyme test	3-order polynomial	Headphone	–	0.39–5.85
Peng, 2011 [31]	Chinese	PB words	3-order polynomial	Headphone	343–41656	0.6–5
Zhu, 2014 [28]	Chinese	PB words	3-order polynomial	Headphone	108–1674	0.63–1.6
Bradley, 1986 [32]	English	Rhyme test	3-order polynomial	Loudspeaker	362–2000	0.8–3.8
Anderson, 1987 [33]	English	PB words	3-order polynomial	Headphone	–	–
Present work	Chinese	CMNmatrix	–	Headphone	16.2	0.06

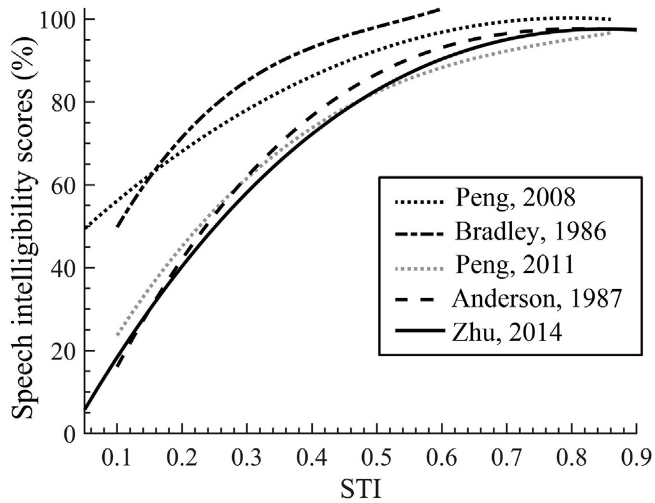

Fig. 6. The curve of the STI-SI models for different types of speech test materials, languages, presentation ways and listening conditions.

0.94), and their standard errors (i.e., −0.93 and 2.76) are also lower than others. Comparison of the results in Fig. 4 and Fig. 7 also show that the SRT results based on the STI-SI model for rhyme test in Peng [30] and Bradley [32] are closer to that subjective SRT result than others for PB words test in term of the absolute value, especially the former that is almost in line with the subjective SRT curves in Fig. 4. While the SRTs based on the STI-SI models for PB words test cannot be comparable with the subjective result in terms of absolute value, with standard errors exceed 5 dB (see Table 3). This is probably because the CMNmatrix test [36] adopted in present work is more similar to the rhyme test, especially Chinese rhyme test, rather than PB words test.

Note that, the change rules of the SRT in listening room (see Fig. 7) seem to be better expressed compared with those from subjective experiment (see Fig. 4), rather than that in automobile. In order to quantify the performance, the Spearman correlation analysis between the BSTIs and subjective SRTs under different measured conditions was conducted. Result shows that the correlation coefficients for listening room and automobile are −0.957 and −0.985 respectively. It seems that the correlation between the BSTI and SRT in listening room or automobile is strong

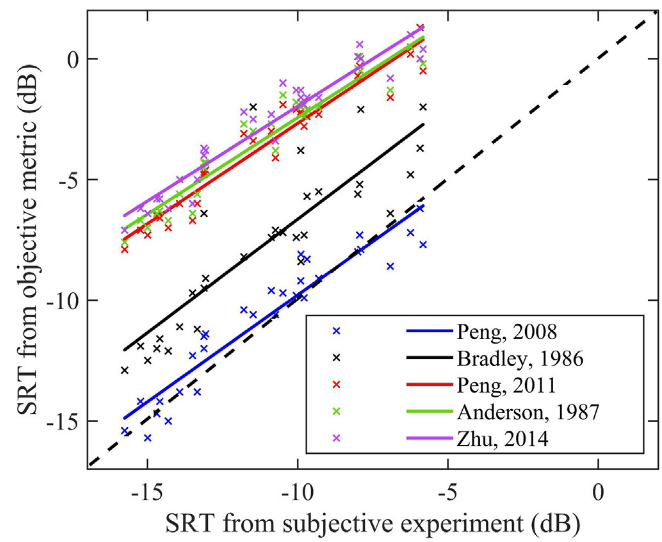

Fig. 8. Relationship between the SRT result from subjective experiment and based on different STI-SI models. × represents the SRT for each conditions, consisting of two measured conditions, three speaker locations and five head orientations. The solid lines represent the linear regression models for the SRT result from subjective experiment and based on different STI-SI models.

Table 3
The linear regression models for the SRT result from subjective experiment and based on different STI-SI models.

Model	Regression equation
Peng, 2008	$SRT_{objective} = 0.89 \times SRT_{subjective} - 0.93$
Peng, 2011	$SRT_{objective} = 0.83 \times SRT_{subjective} + 5.64$
Zhu, 2014	$SRT_{objective} = 0.79 \times SRT_{subjective} + 5.9$
Bradley,1986	$SRT_{objective} = 0.94 \times SRT_{subjective} + 2.76$
Anderson, 1987	$SRT_{objective} = 0.8 \times SRT_{subjective} + 5.56$

and there is no obvious difference between two measured conditions, which further reveals that the difference between subjective and objective results is mainly caused by the inapplicability of STI-SI model derived from the traditional room in the automotive environment. Actually, this difference is a common problem, with objective SI metrics often being mismatched with subjective methods in automotive environments [15,29,47] due to the limitations

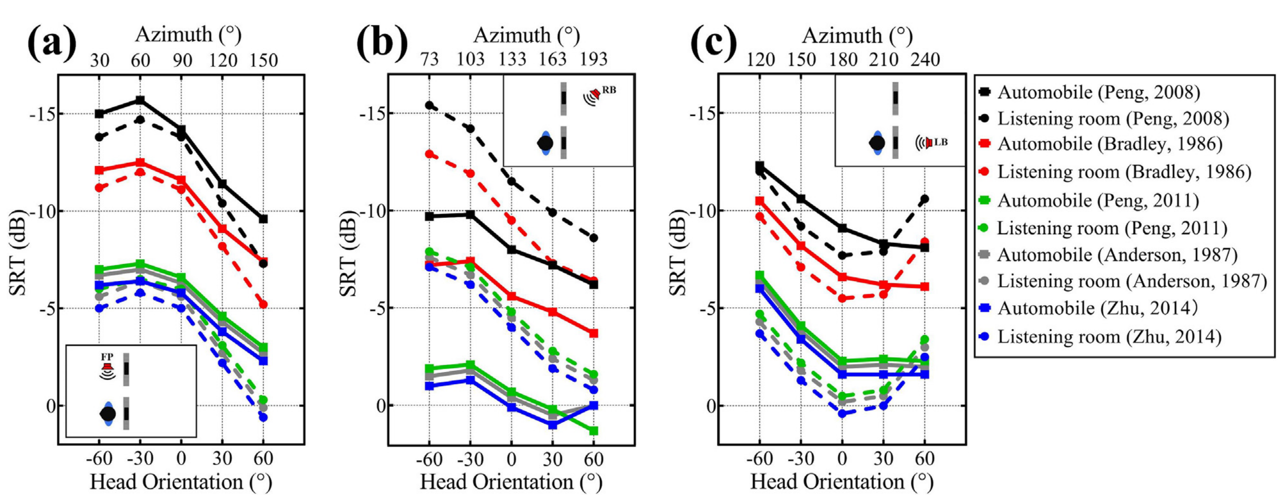

Fig. 7. SRT results based on different STI-SI models in listening room and automobile at various head orientations (corresponding to various azimuths of speaker relative to listener's head), when the speaker is in (a) the FP-seat, (c) RB-seat and (e) LB-seat respectively. The upper label on the x axis in each subgraph gives the azimuths of speaker relative to listener's head, corresponding to various head orientations. Insets show schematic diagrams of location relationship between speaker and listener.

of traditional metrics in automotive applications, including the band importance function and the speech or background spectrum etc. In general, the relationship between the SI scores and STI maybe particularly problematic in automobiles compared with traditional indoor environments, which remains further discussions.

4. Conclusions

Under the complicated acoustic field conditions in an automobiles, we investigated how the head orientation of the listener in the driver's seat affects speech intelligibility (SI) for speaker positions in the front passenger seat, right-back seat, and left-back seat, by measuring the speech reception thresholds (SRTs) of ten participants via headphones. It has been revealed that:

(1) The SRT in automobile as well as listening room varies with listener head orientation considerably. The SRT variations caused by head orientations are up to 5 dB in automobile, lower than that of 9 dB in listening room that is only determined by the effect of binaural listening.
(2) In comparison with the result in listening room, early reflections improve SI in automobile by an SRT decrease of up to 3.5 dB, while seat-back occlusions reduce SI by an SRT increase of up to 5.3 dB. The early reflections play a more significant role when the listener is in an adverse position. Thus, the SI variation with head orientation is smaller in the automobile than in the listening room.
(3) The SRT in automobile decreases as the speaker moves closer to the front lateral direction relative to the listener's head, which is basically in line with that in the listening room. It seems that even though the early reflections from interior boundaries or seat-back occlusions affect SI considerably, they are still not sufficient to subvert the trend of SI variation under different listener head orientations. Overall, a higher SI in automobile can be obtained when the listener in the driver seat turns their head inward, i.e., towards the right or with negative head orientations.
(4) The objective method based on speech transmission index (STI) only roughly expresses the variation in SI under various head orientations and measured conditions. The difference between the subjective results and the objective results is mainly caused by the inapplicability of the STI-SI models derived from the traditional room in the automotive environment. It is defective to evaluate the SI in automobile only through binaural speech transmission index and STI-SI model derived from the traditional room. In future work, we will conduct adequate subjective and objective experiments to explore the relationship between the binaural STI and SI under the special and complex sound field conditions in automobiles.

The present work is conducive to systematically comprehend the combined effects of various factors on SI under the special and complex sound field conditions in automobiles, and can provides evidences for future evaluations of SI in automobiles.

CRediT authorship contribution statement

Linda Liang: Conceptualization, Validation, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Software, Conceptualization. **Guangzheng Yu:** Conceptualization, Validation, Methodology, Resources, Data curation, Writing – review & editing, Supervision, Funding acquisition, Project administration. **Huali Zhou:** Methodology, Software, Visualization. **Qinglin Meng:** Methodology, Funding acquisition.

Declaration of Competing 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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China through Grant No. 12074129 and the Guangdong Basic and Applied Basic Research Foundation through Grant No. 2020A1515010386. We are grateful to all the participants for their patience and cooperation during this study.

References

[1] Granier E, Dalenback B, Svensson P, Kleiner B. Experimental auralization of car audio installations. J Audio Eng Soc 1996;44:835–49. URL: http://www.aes.org/e-lib/browse.cfm?elib=7882.
[2] Kleiner M, Tichy J. Acoustics of Small Rooms. Boca Raton, FL: CRC Applied Mathematics Research Press; 2014.
[3] Parizet E. Speech intelligibility mappings in a car compartment. Int J Vehicle Design 1993;14:132–44. https://doi.org/10.1504/IJVD.1993.061830.
[4] Rumsey F. Automotive audio: They know where you sit. J Audio Eng Soc 2016;64(9):705–8. URL: http://www.aes.org/e-lib/browse.cfm?elib=18378.
[5] Samardzic N, Novak C. In-vehicle speech intelligibility for different driving conditions using the speech transmission index. Noise Control Eng J 2011;59:397–407. https://doi.org/10.3397/1.3598380.
[6] Samardzic N, Novak C. In-vehicle application of common speech intelligibility metrics. Int J Veh Noise Vib 2011;7:328–46. https://doi.org/10.1504/IJVNV.2011.043193.
[7] Herrera C, Donoso-Garcia P, Medeiros E. Intelligibility in low-cost automotive audio systems. J Audio Eng Soc 2016;64(5):320–31. URL: http://www.aes.org/e-lib/browse.cfm?elib=18137.
[8] Ebbitt GL, Remtema TM. Automotive speech intelligibility measurements. Sound Vib 2017;51(6):6–11. https://doi.org/10.4271/2015-01-2337.
[9] Lochner JPA, Burger JF. The influence of reflections on auditorium acoustics. J Sound Vib 1964;1(4):426–54. https://doi.org/10.1016/0022-460X(64)90057-4.
[10] Bradley J, Reich R, Norcross S. On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. J Acoust Soc Am 1999;106 (4):1820–8. https://doi.org/10.1121/1.427932.
[11] Bradley J, Sato H, Picard M. On the importance of early reflections for speech in rooms. J Acoust Soc Am 2003;113(6):3233–44. https://doi.org/10.1121/1.1570439.
[12] Arweiler I, Buchholz JM. The influence of spectral characteristics of early reflections on speech intelligibility. J Acoust Soc Am 2011;130(2):996–1005. https://doi.org/10.1121/1.3609258.
[13] Warzybok A, Rennies J, Brand T, Doclo S, Kollmeier B. Effects of spatial and temporal integration of a single early reflection on speech intelligibility. J Acoust Soc Am 2013;133(1):269–82. https://doi.org/10.1121/1.4768880.
[14] Visintainer A, VanBuskirk J. Effects of sound absorption on speech intelligibility in an automotive environment. In Proceedings of the Noise and Vibration Conference, Traverse City; 1997.
[15] Samardzic N, Novak C, Gaspar R. The evaluation of speech intelligibility in a simulated driving environment using the hearing in noise test. Int J Veh Noise Vib 2012;8(4):318–36. https://doi.org/10.1504/IJVNV.2012.051538.
[16] Samardzic N. The applicability of the objective speech intelligibility metrics for vehicle interior speech intelligibility evaluation, considering different listening configurations and background noise spectra. SAE Int J Passeng Cars-Mech Syst 2014;69:434–8. https://doi.org/10.4271/2014-01-9126.
[17] Samardzic N, Novak C. A novel method for in-vehicle speech intelligibility evaluation and statistical variability analysis. Int J Veh Noise Vib 2015;11 (2):114–32. https://doi.org/10.1504/IJVNV.2015.070017.
[18] Wang D, Tan G, Chen S, Jiang J, Su L. Research on speech intelligibility of sound field inside passenger car. Adv Mater Res 2012;403:5214–9. https://doi.org/10.4028/www.scientific.net/AMR.403-408.5214.
[19] Bilzi P, Bozzoli F, Farina A. Influence of artificial mouth's directivity in determining speech transmission index. In: Proceedings of the 119th Audio Engineering Society Convention. New York, USA; 2005. URL: http://www.aes.org/e-lib/browse.cfm?elib=13345.
[20] Yu L, Yu G, Meng Q. Effect of human speaker's head rotation on speech transmission index in vehicle sound environment. In: Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Honolulu, Hawaii; 2018.
[21] Westermann A, Buchholz JM. The effect of spatial separation in distance on the intelligibility of speech in rooms. J Acoust Soc Am 2015;137(2):757–67. https://doi.org/10.1121/1.4906581.
[22] Padilla L, Felipe O. Binaural speech intelligibility as a function of speaker angle and interaural cross-correlation of disturbing noise and reverberation. J Acoust Soc Am 2010;128(4):2319. https://doi.org/10.1121/1.3508175.

[23] Padilla L, Felipe O. Improving speech intelligibility for binaural voice transmission under disturbing noise and reverberation using virtual speaker lateralization. J Appl Res Technol 2015;13(3):351–8. https://doi.org/10.1016/j.jart.2015.07.001.
[24] Liang L, Yu G. Binaural speech transmission index with spatialized virtual speaker in near field: distance and direction dependence. J Acoust Soc Am 2020;148(2):EL202–7. https://doi.org/10.1121/10.0001808.
[25] van Wijngaarden S, Drullman R. Binaural intelligibility prediction based on the speech transmission index. J Acoust Soc Am 2008;123:4514–23. https://doi.org/10.1121/1.2905245.
[26] Liang L, Yu L, Zhao T, Meng Q, Yu G. Speech intelligibility for various head orientations of a listener in an automobile using the speech transmission index. J Acoust Soc Am 2021;149(4):2686–97. https://doi.org/10.1121/10.0004265.
[27] IEC 60268-16. Sound system equipment. part 16: Objective rating of speech intelligibility by speech transmission index, (International Electrotechnical Commission, Geneva, Switzerland, 2011).
[28] Zhu P, Mo F, Kang J. Relationship between chinese speech intelligibility and speech transmission index under reproduced general room conditions. Acta Acust United Acust 2014;100(5):880–7. https://doi.org/10.3813/AAA.918767.
[29] Samardzic N, Moore B. Binaural speech-to-noise loudness ratio at the speech reception threshold in vehicles. Noise Control Eng J 2021;69(2):173–9. https://doi.org/10.3397/1/376917.
[30] Peng J. Relationship between chinese speech intelligibility and speech transmission index in rooms using dichotic listening. Chin Sci Bull 2008;53(18):2748–52. https://doi.org/10.1007/s11434-008-0383-5.
[31] Peng JX. Relationship between chinese speech intelligibility and speech transmission index in rooms based on auralization. Speech Commun 2011;53:986–90. https://doi.org/10.1016/j.specom.2011.05.004.
[32] Bradley J. Predictors of speech intelligibility in rooms. J Acoust Soc Am 1986;80(3):837–45. https://doi.org/10.1121/1.393907.
[33] Anderson B, Kalb J. English verification of the sti method for estimating speech intelligibility of a communications channel. J Acoust Soc Am 1987;81(6):1982–5. https://doi.org/10.1121/1.394764.
[34] Majdak P, Balazs P, Labac B. Multiple exponential sweep method for fast measurement of head related transfer functions. J Audio Eng Soc 2007;55:632–46. URL: http://www.aes.org/e-lib/browse.cfm?elib=14190.
[35] Kondo K, Chiba T, Kitashima Y, Yano N. Intelligibility comparison of japanese speech with competing noise spatialized in real and virtual acoustic environments. Acoust Sci Technol 2010;31(3):231–8. https://doi.org/10.1250/ast.31.231.
[36] Hu H, Xi X, Wong LL, Hochmuth S, Warzybok A, Kollmeier B. Construction and evaluation of the mandarin chinese matrix (cmnmatrix) sentence test for the assessment of speech recognition in noise. Int J Audiol 2018;57(11):838–50. https://doi.org/10.1080/14992027.2018.1483083.
[37] GB/T 7347-1987. The standard spectrum of chinese speech, (National Standard of China, 1987).
[38] Degan ND, Prati C. Acoustic noise analysis and speech enhancement techniques for mobile radio applications. Signal Process 1988;15(1):43–56. https://doi.org/10.1016/0165-1684(88)90027-8.
[39] Kendall G. The decorrelation of audio signals and its impact on spatial imagery. Comput Music J 1995;19(4):71–87. https://doi.org/10.2307/3680992.
[40] Xie B, Shi B, Xiang N. Audio signal decorrelation based on reciprocal-maximal length sequence filters and its applications to spatial sound. In Proceedings of Audio Engineering Society Convention, San Francisco, CA; 2012.
[41] Xiang N, Schroeder M. Reciprocal maximum-length sequence pairs for acoustical dual source measurements. J Acoust Soc Am 2003;113(5):2754–61. https://doi.org/10.1121/1.1561498.
[42] Brand T, Kollmeier B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. J Acoust Soc Am 2002;111(6):2801–10. https://doi.org/10.1121/1.1479152.
[43] Yu G, Wu R, Liu Y, Xie B. Near-field head related transfer-function measurement and database of human subjects. J Acoust Soc Am 2018;143(3):EL194–8. https://doi.org/10.1121/1.5027019.
[44] Lokki T, Pätynen J. Lateral reflections are favorable in concert halls due to binaural loudness. J Acoust Soc Am 2011;130(5):EL345–51. https://doi.org/10.1121/1.3647866.
[45] Bozzoli F, Farina A. Measurement of the speech intelligibility inside cars. In Proceedings of the 113th Audio Engineering Society Convention, Los Angeles, California, USA; 2002. URL: http://www.aes.org/e-lib/browse.cfm?elib=11298.
[46] Viktorovich M. Implementation of a new metric for assessing and optimizing the speech intelligibility inside cars. In: Proceedings of the SAE Noise and Vibration Conference and Exhibition. Michigan: Traverse City; 2005.
[47] Samardzic N, Novak C. The band importance function in the evaluation of the speech intelligibility index at the speech reception threshold within a simulated driving environment. In: Proceedings of SAE Noise and Vibration Conference and Exhibition, Grand Rapids, MI.