The speech intelligibility and applicability of the speech transmission index in large spaces

Hongshan Liu,[a] Hui Ma,[a] Jian Kang,[a,b] and Chao Wang[a,*]

[a] *School of Architecture, Tianjin University, 300072 Tianjin, China*

[b] *UCL Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), Central House, 14 Upper Woburn Place, London WC1H 0NN, UK*

Abstract

This paper aims to explore the influence factors of speech intelligibility and the applicability of the speech transmission index (STI) in large spaces, where the sound energy is unevenly distributed and non-exponentially decays. The subjective speech intelligibility tests were conducted in Mandarin (China mainland) in two large spaces with volumes of 97000 m³ and 246000 m³. Objective indicators such as the Reverberation Time (RT), Early Decay Time (EDT), Definition (D_{50}), and Speech Transmission Index (STI) under different signal-to-noise ratio (SNRs) were also measured in these two spaces. The results showed that both the SNR and room acoustics had significant effect on the speech intelligibility in these two spaces, but the effect of room acoustics on speech intelligibility was also affected by SNR. The speech intelligibility scores significantly increased with the increase in SNR when the SNR was less than 14.4dB. In terms of room acoustics, D_{50} was more relevant to speech intelligibility than RT and EDT in these two large spaces when SNR ranged from -5dB to 15dB. The STI value in large spaces should not be used as in ordinary spaces to evaluate the speech intelligibility. Based on the tests in this paper, the corresponding relation between STI and speech intelligibility in large spaces was modified, and a new rating threshold of STI was also proposed according to the revised relation, which indicated a necessity to modify the rating criteria of using STI to predict speech intelligibility in large spaces.

Keywords: Speech intelligibility; Speech transmission index; Large spaces;

* pdwangchao@tju.edu.cn; Corresponding author.

1. Introduction:

In recent years, large space buildings have been extensively built worldwide, where the sound reinforcement system is of great significance not only for the general public address but also for the evacuation of people in emergency situations. Speech intelligibility is a measure of the sound quality of the sound reinforcement system, which is affected by many factors such as room acoustics and SNR.

The effects of the RT and SNR are important for determining the speech intelligibility, and their correlations have been explored through a series of studies [1-6]. Bradley explored the correlations by conducting English speech intelligibility tests in real classrooms and simulated sound fields [1-4]. He pointed out that both RT and SNR had significant effect on the speech intelligibility, while the effect of the SNR was much more important. Peng established the relation between Mandarin speech intelligibility and objective indicators in ordinary spaces with volume varied from 342 m^3 to 41656 m^3 by using the auralization method, and similar conclusions were obtained [5, 6]. However, the sound field in an ordinary space is different from that in large spaces, where the sound energy is unevenly distributed and non-exponentially decays [7]. And the different forms of sound energy decay curves would significantly influence the effect of human perception of speech [8], so the non-exponentially decay curves might cause significant changes in the relation among RT, SNR and speech intelligibility. Many acoustic measurements were conducted in famous large spaces [9-11], and the speech intelligibility of St Paul's Cathedral in London was discussed by Lewers, but the relation between the speech intelligibility and objective indicators was not obtained [9]. Therefore, the influence factors of the speech intelligibility in a large space must be further explored.

The evaluation methods of the speech intelligibility include subjective tests and objective measurements [12, 13]. Subjective tests are the basic and direct method to obtain accurate speech intelligibility, but it costs too much time and labour. The speech transmission index (STI), which was proposed by Hougast and Steeneken [14-16], is the most widely used objective index in ordinary spaces. In the calculation of the STI, the modulation transfer function is used to reflect the effect of the room acoustics and SNR on the speech intelligibility. Houtgast and Steeneken proposed a series of modified models of the STI [17-20], and the repetition of information in adjacent bands of speech was taken into account. They verified the validity of a modified model to evaluate the speech intelligibility of Western language systems under different environments in ordinary spaces.

The variation of the room acoustics may affect the applicability of the STI [21-23]. Zhu used the auralization method to investigate the applicability of the STI on the evaluation of Mandarin speech intelligibility in different types of spaces with volume varied from 108 m^3 to 1674 m^3, and pointed out that there were significant differences in the relation between STI and speech intelligibility among these spaces [22]. Kang explored the relation between speech intelligibility and STI in the seminar room and long space [23]. It was found that for a given STI, the speech intelligibility scores in the long space were higher than those in the ordinary space. However, in large spaces, the sound field is significantly different from that of the ordinary space. For example, the interval time between late reflections and direct sound in a large space is very long, which makes the late reflections similar to the echoes. This issue may cause an overestimation of speech intelligibility by the STI based on the research in IEC 60268-16,4rd. [13]. Therefore, it is necessary to verify the applicability of the STI in large spaces.

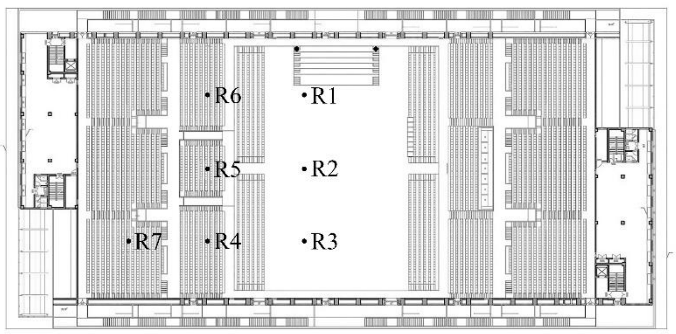
This study aims to explore the influence factors of speech intelligibility through subjective tests in two large spaces. The speech intelligibility scores were obtained under six different SNRs in these spaces, and the objective indicators at the receivers were measured. Based on these data, the applicability of using the STI to predict the speech intelligibility in large spaces was studied.

2. Experiment Method

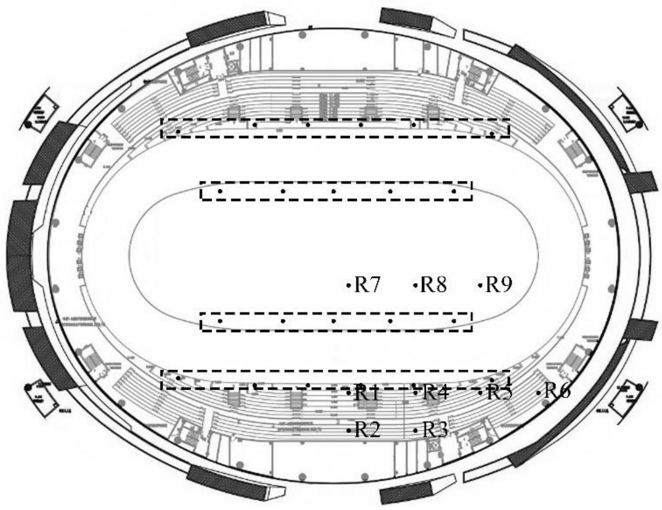
In this study, to explore the applicability of the STI and influence factors of Mandarin (China mainland) speech intelligibility, subjective tests and objective measurements were conducted at the same receivers in two large spaces in Tianjin, China: Tianjin University Stadium (TJU) with the volume of 97000 m^3 and Tianjin Sports Centre (TSC) Velodrome with the volume of 246000 m^3.

2.1 Objective measurement

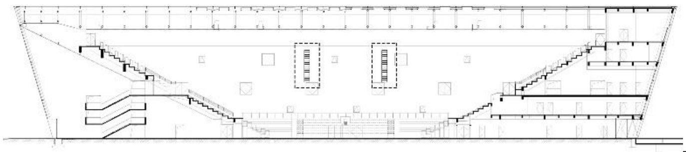
Seven receiver positions were arranged in the stadium and nine receiver positions were arranged in the velodrome. The layout of sound source points and receiving points are shown in Fig. 1. The T_{20}, EDT, D_{50} and decay curves at different receivers were measured using Dirac 6.0 through the impulse response method recommended by ISO-3382 [24]. The impulse response was measured by using the sine-sweep signal, which was generated from Dirac 6.0 and played by sound reinforcement systems in the two spaces. To compensate the effect of the uneven frequency response of the sound reinforcement system on the impulse response and speech intelligibility, the sound reinforcement system was equalized by employing an inverse filter in seven octave bands with centre frequency ranging from 125Hz to 8000Hz in advance [25]. The measurement instruments included a PC computer, a Yamaha UR242 soundcard, a DBX 1231 equalizer and a B&K2270 hand-held analyser. During the measurement, no audible echo appeared in either of the two large spaces.



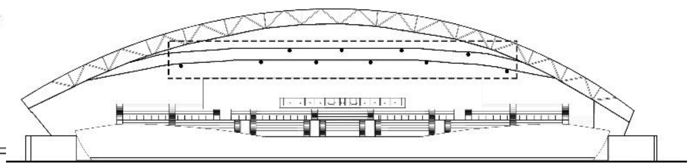
a. Plan of Tianjin University Stadium



b. Plan of Tianjin Sports Centre Velodrome



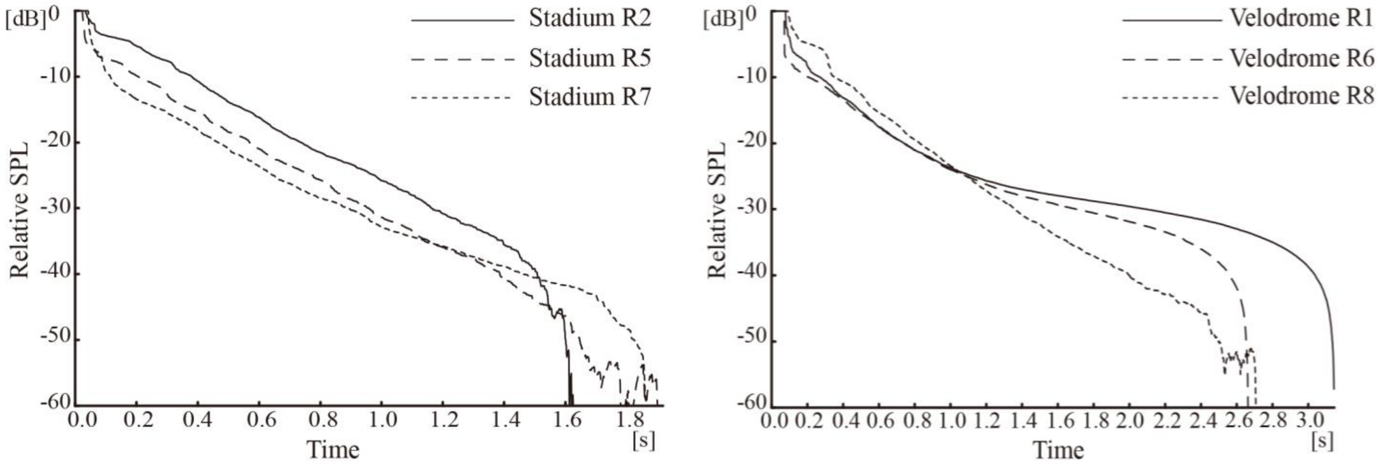
c. Section of Tianjin University Stadium



d. Section of Tianjin Sports Centre Velodrome

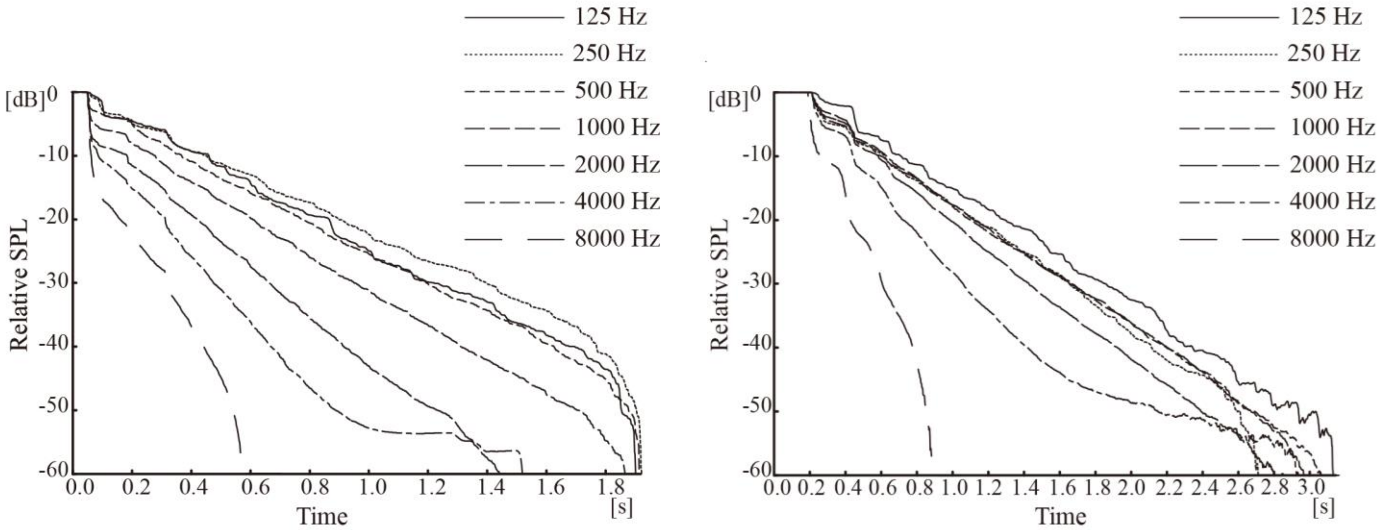
Fig. 1. Layout of the receiver positions and sound sources in the two rooms. In the large spaces, the signal and noise were played by the sound reinforcement system. The loudspeakers of the stadium and velodrome were arranged in a line array and a distributed arrangement, respectively (the sound reinforcement system in the large space is inside the dotted line).

T_{20}, EDT and D_{50} in six octave bands with centre frequency ranging from 125 Hz to 4000 Hz are shown in Appendix A and the average T_{20} and EDT in two octave bands with centre frequency of 500 Hz and 1000 Hz, the average D_{50} in four octave bands with centre frequency ranging from 500 Hz and 4000 Hz are also shown in Appendix A. These particular octave bands were chosen according to the work by Peng [6], which would make the comparison between the two studies straightforward and robust. The full band decay curves of different receivers in large spaces are shown in Fig.2, and the decay curves in seven octave bands with centre frequency ranging from 125 Hz to 8000 Hz are shown in Fig.3.



a. Tianjin University Stadium b. Tianjin Sports Centre Velodrome

Fig. 2. Full band decay curves in two large spaces.



a. Tianjin University Stadium R5 b. Tianjin Sports Centre Velodrome R8

Fig. 3. Decay curves in seven octave bands with centre frequency ranging from 125Hz to 8000Hz.

2.2 Subjects

The experimental subjects were selected from graduate students of Tianjin University, aged between 22 and 30 years old. They were able to listen and speak Mandarin normally, and all subjects passed the pure tone hearing test. Prior to the formal experiment, all subjects were trained and pre-tested to ensure that the speech intelligibility score of each subject was at least 95% under normal conditions. Finally, 10 subjects participated in the experiments, including 5 males and 5 females. During the experiment, 10 subjects took turns participating in the experiment at each receiver to ensure that each subject completed six experiments with different SNRs at each receiver. A total of 2320 individual speech intelligibility scores were obtained in the two large spaces.

2.3 Speech intelligibility test

The speech materials of the subjective speech intelligibility test usually included two types: the Diagnostic Rhyme word list [26] and the Phonemically Balanced word list [12]. Since the Diagnostic Rhyme test is a closed-set test, which achieves a relatively high speech intelligibility score when the speech intelligibility is high, the speech intelligibility tests were carried out using the Mandarin PB word list. Every Mandarin PB word list contained 75 syllables and was divided into 25 groups by random combination; thus, each group had three syllables and was embedded in a carrier phrase. All word lists were recorded at a rate of 4.0 words per second spoken by one male and one female in the anechoic chamber; meanwhile, an 8-second silence was added between each group using Audition.

The sound pressure level of the speech signal at receiver 1 was set at 70 dB by adjusting the volume of mixer, and the SPL of the remaining points was measured by the Norsonic 140 sound level meter. For a wide range of speech intelligibility, noise and speech signals were simultaneously played through the sound reinforcement system. The experiment was set at five SNRs, -5 dB, 0 dB, 5 dB, 10 dB and 15 dB in Tianjin University Stadium. To obtain high speech intelligibility, no artificial noise interference condition was added to the Velodrome. The SNRs of the remaining receivers were also measured by Norsonic 140, as shown in Appendix B. The impulse response and the SNRs in seven octave bands were also used to calculate the full STI, which was recommended by IEC 60268-16.

Since the speech signals and noise might have different temporal and spectral characteristics, which makes it difficult to compare the SPL of speech signals and noise, the following measures were taken. First, the pink noise was modulated according to the standard spectrum of Mandarin speech to ensure identical SNRs for all octave bands. Secondly, the SPL of the discontinuous speech signals was calibrated by eliminating the effect of the silent period, which was achieved by applying a threshold to the overall RMS value of speech signals [13].

3. Results

3.1 Effect of the SNR

The SNR and room acoustics are the most important influence factors of speech intelligibility. The subjective speech intelligibility scores were obtained in two large spaces with different volumes under different SNRs from

-5 dB to 30 dB. The relation between speech intelligibility scores and SNR for two large spaces and their best fit third-order polynomial is shown in Fig. 4.

There is a high correlation between speech intelligibility and SNR in both large spaces. The determination coefficient R square of the two curves was 0.95 and 0.94, respectively. The analysis of variance shows that the SNR significantly affects the speech intelligibility in large spaces (F=254.1, P=0.000). The LSD multiple comparison shows that the speech intelligibility significantly varied with the increase in SNR when the SNR was lower than 14.4 dB and hardly changed when the SNR was over 14.4 dB.

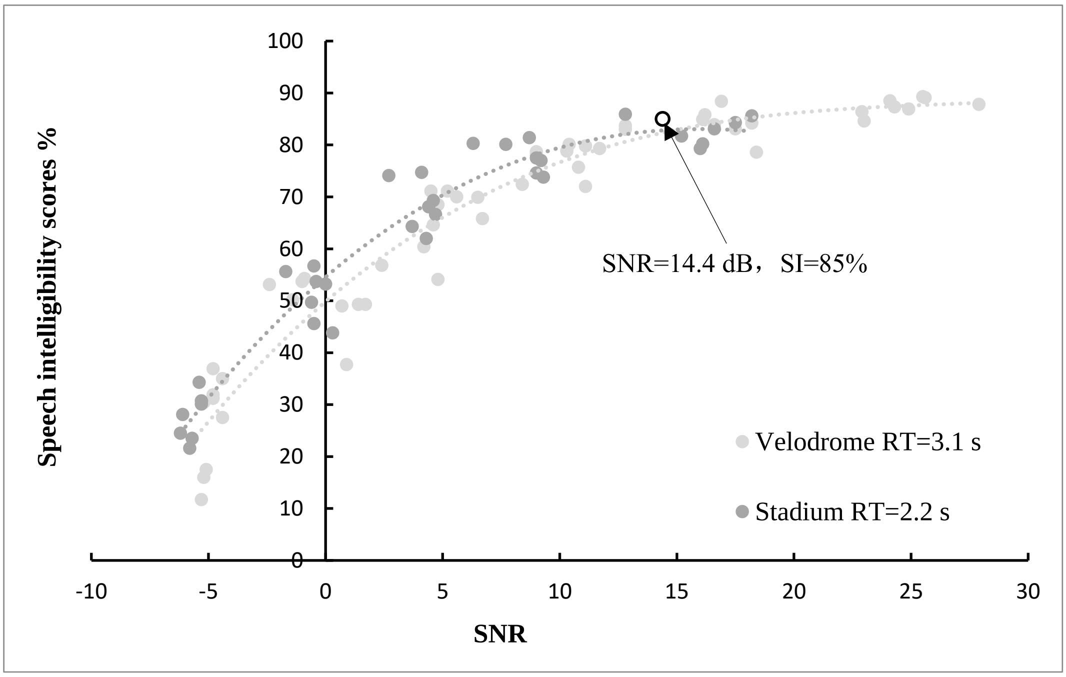


Fig. 4. Relation between speech intelligibility scores and the SNR of two large spaces (the black circle is the turning point in the growth rate of speech intelligibility scores).

When the SNR was relatively low, the difference in speech intelligibility among various receivers was more significant. The spatial dispersion of the speech intelligibility under different SNRs is shown in Table 1. The results in Fig. 4 and Table 1 show that for a relatively low SNR, the difference in speech intelligibility between each receiver was more significant, and the largest difference (approximately 25.2%) appeared when the SNR was near -5 dB. In addition, when the SNR was greater than 10 dB, the change in receiver position had little effect on the speech intelligibility scores in two large spaces, and the speech intelligibility scores were always larger than 72.0%.

Table 1. Standard deviation of speech intelligibility at each receiver position under different SNRs (SNRs in the

Table were approximate values).

SNR	-5 dB	0 dB	5 dB	10 dB	15 dB	25 dB
Stadium	4.22	4.59	4.98	2.94	2.26	--
Velodrome	8.63	5.25	5.51	3.00	2.00	1.51

The speech intelligibility scores of the large spaces and Peng's ordinary spaces under different SNRs are compared in Fig. 5, and the RT of both spaces was approximately 2.2 s. The curve obtained in the large spaces was similar to that established by Peng [6] in ordinary spaces when the SNR was near -5 dB, 0 dB and 15 dB, while the largest difference in speech intelligibility score, which was approximately 6%, appeared when the SNR was near 5 dB.

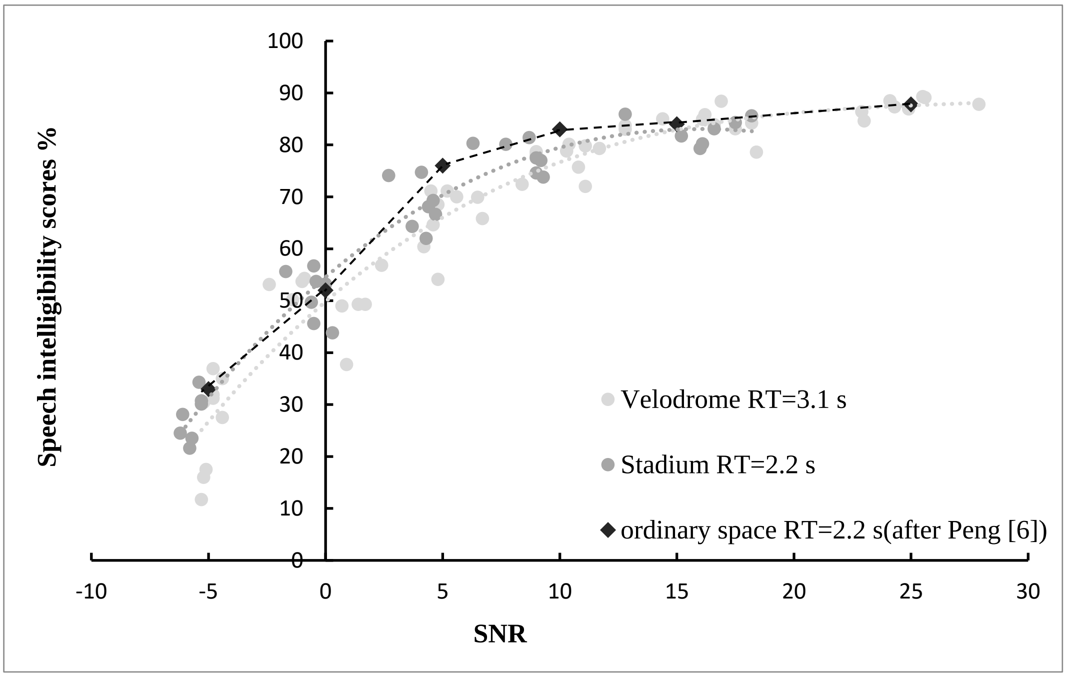
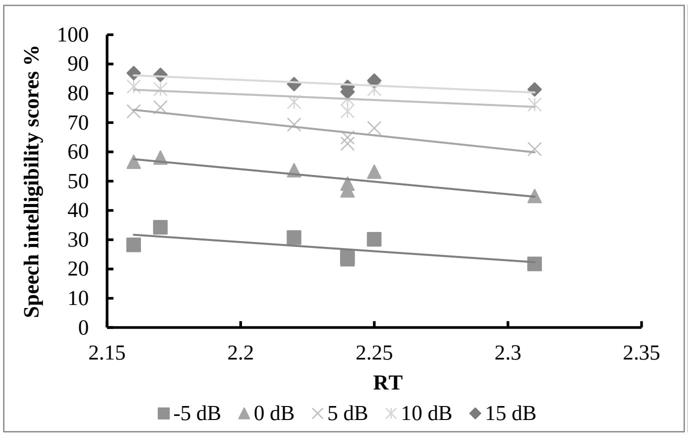


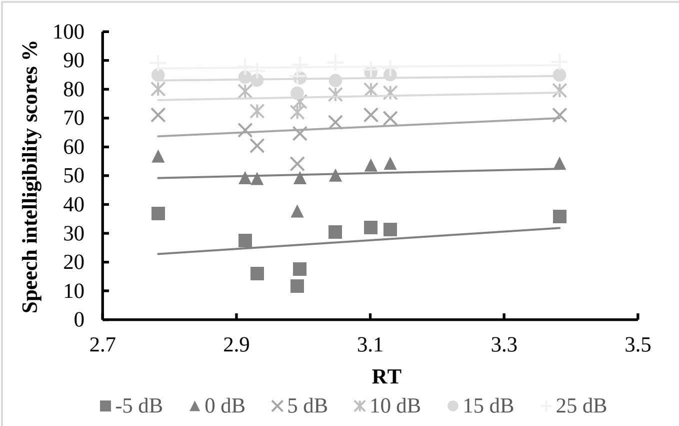
Fig. 5. Comparison of the relation curve between speech intelligibility scores and the SNR in different spaces.

3.2 Effect of the room-acoustic parameters

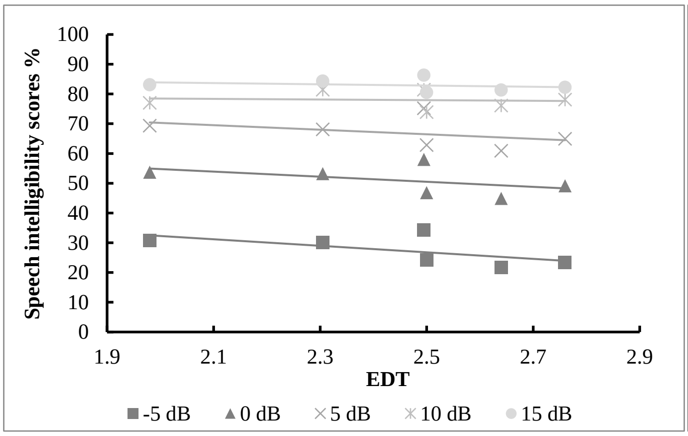
The room acoustics is an important factor that affects the speech intelligibility. It can be seen from the test results that there were significant differences in sound field between ordinary spaces and large spaces. On the one hand, the difference in acoustic characteristics of each receiver in ordinary spaces was small, but it was more significant in large spaces as shown in Appendix A. On the other hand, the decay curves in large spaces as shown in Fig.2 showed significant double-curvature, which was different from the exponential decay curves in ordinary spaces [27]. Moreover, the early decay rate at low-and-medium frequency sound which was the main components of speech had obvious fluctuation. The difference in decay curves between large spaces and ordinary spaces might lead to significant differences in human perception, even if the two types of space had the same room-acoustic parameters [8]. Therefore, the relation between speech intelligibility and acoustic characteristics of each receiver in large spaces should be discussed. Since the sound decay in a large space is very complex, three different parameters were chosen to represent the relation between acoustic characteristics and speech intelligibility: RT, which is the attenuation of sound energy in the room; EDT, which is the early decay of sound energy in the room; and D_{50}, which is the ratio of the early sound energy to the total sound energy.



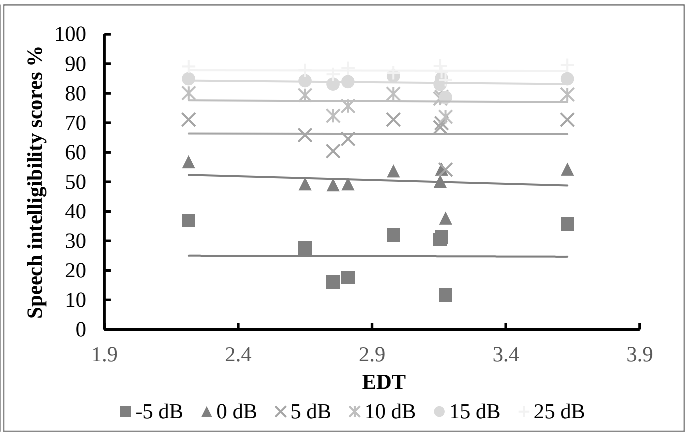
a. Tianjin University Stadium



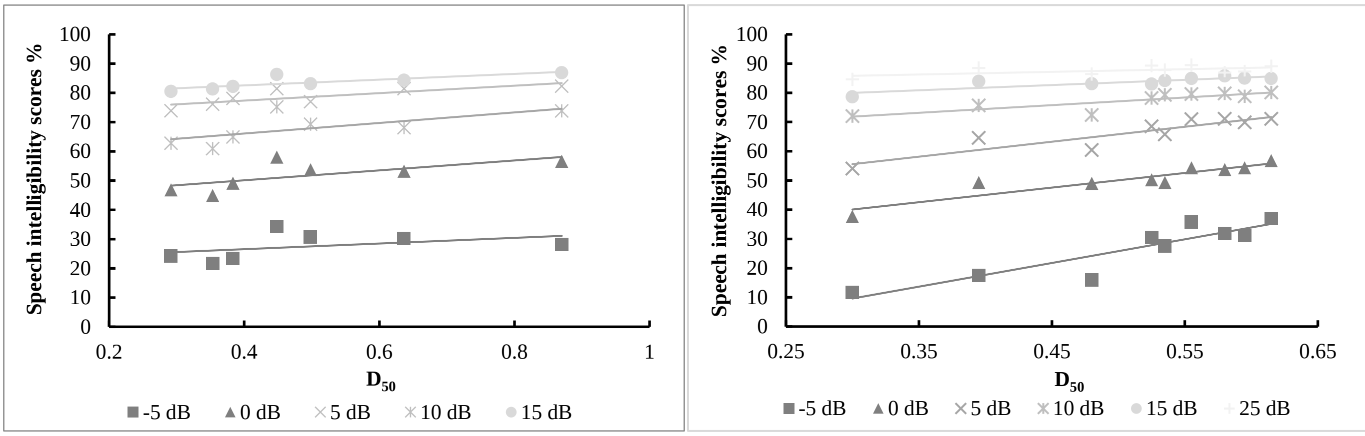
b. Tianjin Sports Centre Velodrome



c. Tianjin University Stadium



d. Tianjin Sports Centre Velodrome



e. Tianjin University Stadium f. Tianjin Sports Centre Velodrome

Fig. 6. Relation between speech intelligibility scores and the RT, EDT and D_{50} of two large spaces.

The relation between the speech intelligibility scores and the RT, EDT and D_{50} in two large spaces is shown in Fig. 6. For the TJU stadium, RT (500Hz-1000Hz) was 2.16-2.31 s, EDT (500Hz-1000Hz) was 1.98-2.76 s, and D_{50} (500Hz-4000Hz) was 0.29-0.87. For the TSC velodrome, RT (500Hz-1000Hz) was 2.78-3.38 s, EDT (500Hz-1000Hz) was 2.22-3.63 s, and D_{50} (500Hz-4000Hz) was 0.3-0.62. The speech intelligibility in the stadium with a smaller volume decreased with the increase in RT or EDT, which was also found in ordinary spaces [5, 27]. However, in the velodrome, for a relatively low SNR, this phenomenon was observed only when RT was less than 3 s; when RT was more than 3 s, the speech intelligibility increased with the increase in RT. In both spaces, the speech intelligibility scores increased with the increase in D_{50} under different SNRs.

The results show that the speech intelligibility had a high correlation with the SNR in large spaces. However, the room acoustics parameters such as the RT, EDT, and D_{50} had different effects on the speech intelligibility under different SNRs. Therefore, based on the results of subjective tests and objective measurements, different prediction equations considering the SNR and other objective indicators (RT, EDT and D_{50}) were obtained by using the regression method, as shown in Table 2.

Table 2. Different prediction equations considering the SNR and other objective indicators.

	Objective indicators	**Expression**	**Determination coefficient R^2**
Equation 1	SNR	$SI=3.67SNR-0.09SNR^2+50.42$	0.927
Equation 2	SNR and RT	$SI=3.67SNR-0.09SNR^2-2.72RT+57.75$	0.932
Equation 3	SNR and EDT	$SI=3.68SNR-0.09SNR^2-2.98EDT+58.11$	0.937
Equation 4	SNR and D_{50}	$SI=3.69SNR-0.09SNR^2+25.20D_{50}+36.68$	0.955

As shown in Table 2, Equations 2 and 3 had higher determination coefficients than Equation 1, which indicated that the addition of RT or EDT improved the accuracy of the prediction equation with only SNR. Equation 4 had the highest determination coefficients among these four equations, which indicated that the combination of SNR and D_{50} might be the more effective method to predict the speech intelligibility. Due to the directivity of public address systems and the non-diffuse sound field in large spaces, the direct energy has a higher proportion at most receivers, which makes the energy-decay curve non-linear. Hence, D_{50} is more advantageous in the prediction of speech intelligibility than RT and EDT.

The stepwise method was conducted to test the significance of the effects of SNR, RT, EDT and D_{50} on the prediction equations of speech intelligibility. The results in Table 3 showed that SNR contributed most to explain the change in speech intelligibility, and EDT and D_{50} could also effectively explain the change in speech intelligibility, therefore, the addition of EDT and D_{50} to the equation could improve the prediction accuracy. However, RT had no significant effect on improving the prediction accuracy, so the equation that combined SNR and RT was excluded.

Table 3. Significance effect of variables in regression models.

	Variable	**Standardized coefficient**	**T**	**Sig.**
Equation 1	SNR	1.535	26.262	.000
	SNR2	-.738	-12.628	.000
Equation 2	SNR	1.546	33.147	.000
	SNR2	-.747	-16.018	.000
	RT	Removed		
Equation 3	SNR	1.541	27.816	.000
	SNR2	-.732	-13.206	.000
	EDT	-.091	-3.300	.001
Equation 4	SNR	1.546	33.147	.000
	SNR2	-.747	-16.018	.000
	D_{50}	.163	7.088	.000

3.3 The relation between speech intelligibility and STI

The relation between speech intelligibility and STI in two large spaces and the best fitting third-order polynomial are shown in Fig.7. The relation between STI and speech intelligibility was established in both ordinary and long spaces. The comparisons of large space curves with Kang's curve in long spaces [23] and Zhu's curve in ordinary spaces [22] are also shown in Fig.7. All curves were obtained using the Mandarin PB-words test.

As shown in Fig.7, for a relatively low STI, the curves in the three types of spaces show a significant difference, while the difference decreases with the increase in STI. For a given STI value, the long space always had the highest scores, while the large space always had the lowest scores. These differences may be due to different shapes and volumes of these spaces. For example, there might be a large number of late reflections that appeared at long intervals due to the long propagation distance in large spaces. These late reflections did not form "audible echo", but they had the similar acoustic characteristics as the echo which might lead to the overestimation of the speech intelligibility when using the STI [13].

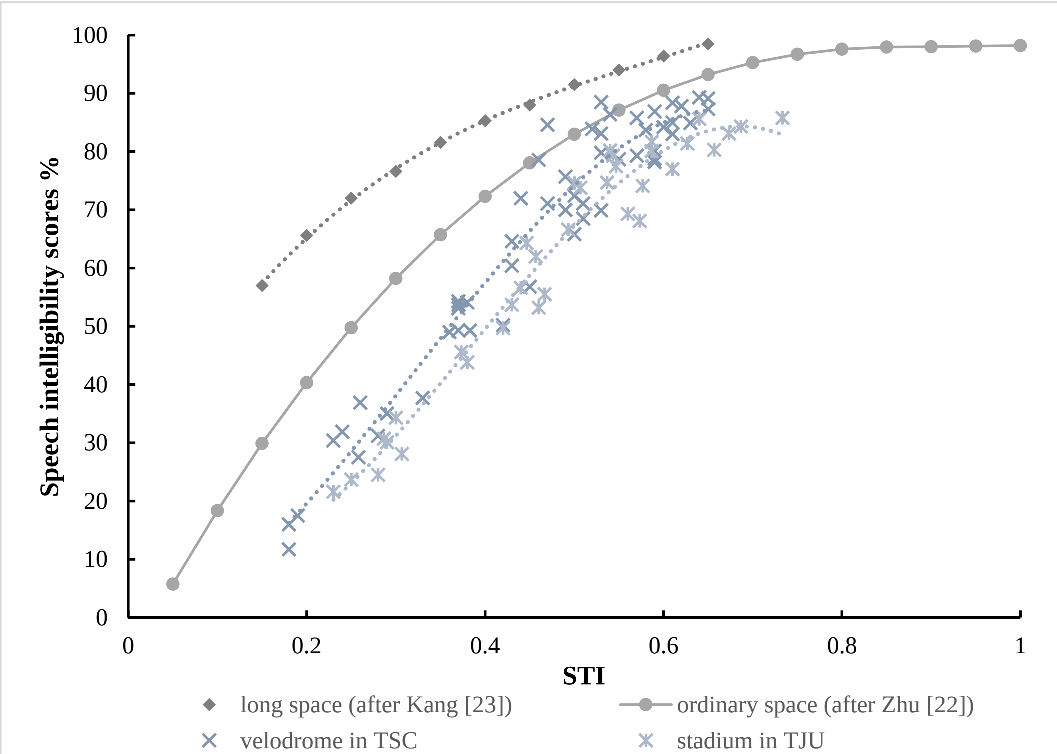
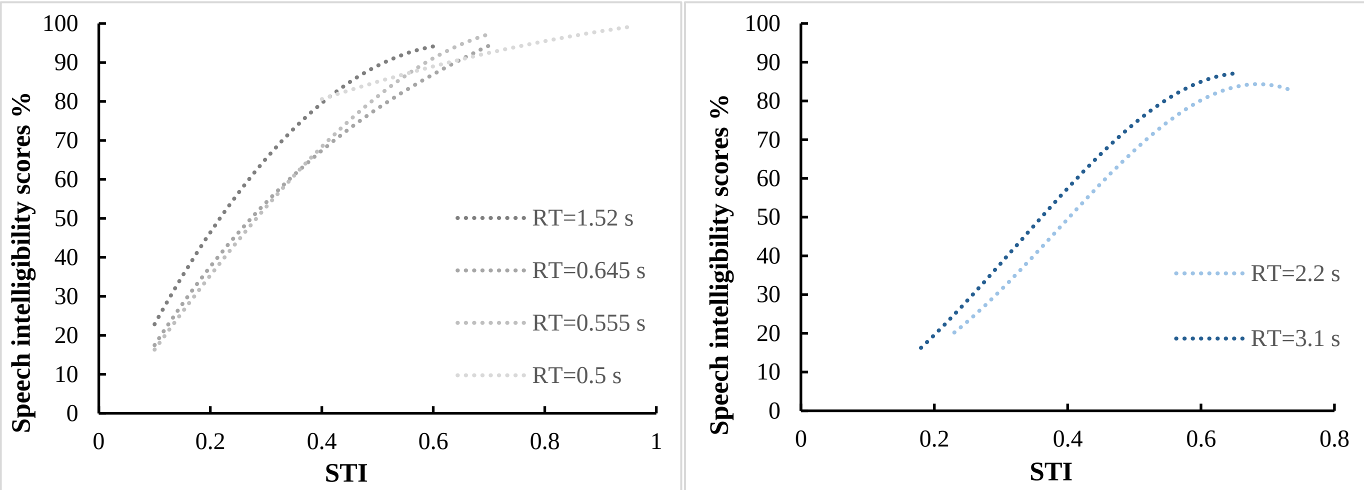


Fig. 7. Relation between speech intelligibility scores and STI in different spaces.

Although the relation between speech intelligibility and STI in large spaces was different from that in ordinary spaces, a high correlation between them was also observed in these two large spaces. The determination coefficient R square of the curves were 0.94 and 0.96, and the standard deviations were 5.16% and 4.06%, for the two large spaces. Thus, the STI, which was proven to be effective for ordinary spaces [22], was also useful for large spaces after the relation was modified. The relations between speech intelligibility and STI, as shown in Fig. 5, in two large spaces are:

$$SI_{(stadium)}=13.765-118.4STI+795.14STI^2-688.9STI^3 \quad (5)$$

$$SI_{(velodrome)}=-1.781+20.719STI+542.35STI^2-559.87STI^3 \quad (6)$$



a. ordinary spaces (after Zhu [22]) b. large spaces

Fig. 8. Relation between speech intelligibility scores and STI of different RT conditions.

The effect of RT on the relation between speech intelligibility and STI in ordinary spaces and large spaces are shown in Fig. 8a and 8b, respectively. The Fig.8a was from Zhu's discussion on the difference between the relation curves of speech intelligibility and STI in different ordinary spaces. And the Fig.8b was from Equations 5 and 6. The relation curves obtained under different RT conditions in two types of spaces indicated the same general trend: for a given STI value, a room with longer RT has higher speech intelligibility scores. The reason for these differences may be that the STI overestimated the effect of the room acoustics on the speech intelligibility [3].

For example, despite a 0.9 s difference existed in RT of the two large spaces, the two curves in Fig.8b were approximately similar, and the largest score difference (approximately 2.3%) appeared when the SNR was about 5 dB as shown in Table 4. However, the difference in STI between the two spaces under identical SNRs ranged from 0.06 to 0.11 as shown in Table 5. And according to Fig.5, the corresponding speech intelligibility differed by approximately 10%. The above results showed that the difference in room acoustics between these two large spaces had not caused significant difference in speech intelligibility, but caused a significant change in STI.

When using the indirect method to calculate the STI [13], the calculation equation included two independent variables: room acoustics and SNR, which meant that when the effect of the SNR on STI was consistent, only the room acoustics would cause the differences of STI as shown in Table 5. In other words, STI overestimated the impact of room acoustics on speech intelligibility. Therefore, reducing the weight of room acoustics in the STI calculation may improve the prediction of the speech intelligibility. And it can be seen from Fig.6 that the effects of the SNR and room acoustics on speech intelligibility were related, which meant that the weights of the room acoustics might vary in the process of calculating the STI under different SNRs.

Table 4. Difference in speech intelligibility between two large spaces under different SNRs (SNRs in the Table were approximate values).

SNR	-5 dB	0 dB	5 dB	10 dB	15 dB
Stadium	27.6%	51.2%	68.4%	77.8%	82.9%
Velodrome	26.5%	50.4%	66.2%	77.2%	83.6%
Difference	1.1%	0.8%	2.3%	0.6%	0.7%

Table 5. Difference in STI between two spaces at different SNRs and the speech intelligibility corresponding to the STI difference (SNRs in the Table were approximate values).

SNR	-5 dB	0 dB	5 dB	10 dB	15 dB
Stadium	0.28	0.42	0.52	0.58	0.63
Velodrome	0.22	0.36	0.44	0.50	0.52
Difference	0.06	0.06	0.08	0.08	0.11
Speech intelligibility	10.0%	11.4%	12.9%	9.9%	10.4%

4. Discussion

4.1 Near-ideal speech intelligibility in large spaces

In an ordinary space, Bradley recommended that achieving 95% speech intelligibility in a classroom was near-ideal conditions because most children could easily achieve 95% speech intelligibility at high SNRs [2]. However, in a large space, even with a very high SNR, it is difficult to reach a speech intelligibility of 90%; based on the test in this paper, if the speech intelligibility that most of the subjects could achieve at the SNR of 15 dB is considered the near-ideal condition, the recommended speech intelligibility is only 80% in large spaces. When SNR was larger than 14.4 dB, the speech intelligibility no longer significantly increased, so the SNR here was determined to be 15 dB. Based on Chen’s survey, in a stadium, the background noise could exceed 65 dB, and in a railway station, the background noise of passengers was 60-70 dB [28]. Thus, in a large space, the signal SPL of the sound reinforcement system should be maintained at 85-90 dB to ensure that the speech intelligibility is the near-ideal condition. However, it is worth noting that the speech intelligibility was not determined entirely by SNR, it also depended on signal levels. Studebaker [29] and Dubno [30] reported that when SNR remained constant, speech intelligibility decreased with the increase in signal level once the signal level exceeded normal level, and Sato [31] pointed out that the upper limit of acceptable signal level was 85dB for noise levels from 55dB to 70dB.

4.2 STI rating for large spaces

To evaluate the level of speech intelligibility in a room with STI, it is necessary to establish a unified relation curve suitable for large spaces. The Equation 7 was produced using all the data from the two large spaces. The determination coefficient R square of the best-fit curve was 0.93 and the standard deviation was 5.76%.

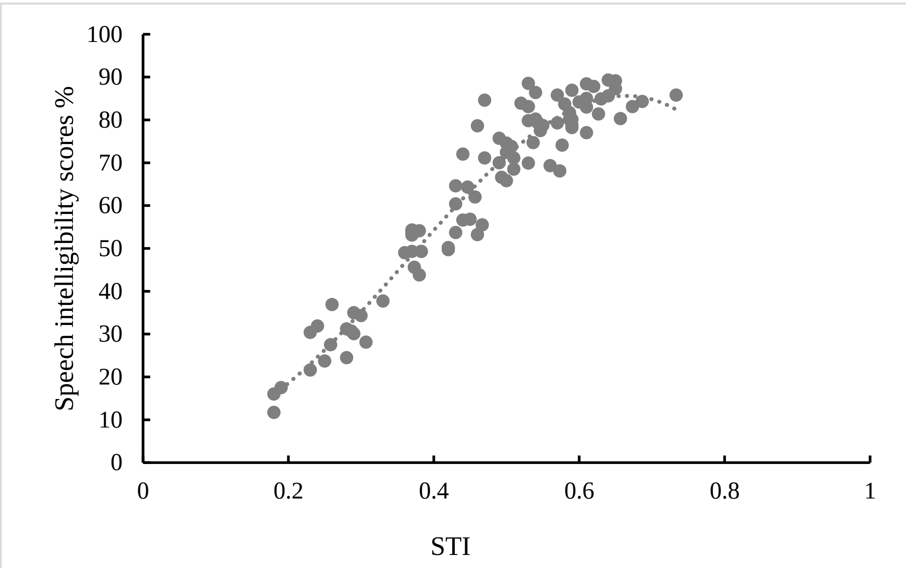


Fig. 9. Relation between speech intelligibility scores and STI in large spaces.

$$SI_{\text{(stadium and velodrome)}}=9.244-74.267STI+742.55STI^2-688.95STI^3 \quad (7)$$

Steeneken divided the STI into five levels based on the relation between STI and speech intelligibility [20]; the dividing point for each level is shown in Table 6: 0.60, 0.45 and 0.30 for the corresponding speech intelligibility of approximately 90.5%, 82.0% and 58.2% in an ordinary space. For a given speech intelligibility, the STI in a large space was significantly larger than that in an ordinary space. Therefore, according to the relation curve of speech intelligibility and STI in the large spaces as shown in Fig.9, the STI values for 90.5%, 82.0% and 58.2% should be 0.68, 0.55 and 0.42, respectively.

Table 6. STI rating in different spaces.

	Bad	Poor	Fair	Good	Excellent

STI in ordinary space	0~0.30	0.30~0.45	0.45~0.60	0.60~0.75	0.75~1
STI in large space	0~0.42	0.42~0.55	0.55~0.68	0.68~--	--

4.3 Further research

In this study, influence factors such as the arrangement and directivity of the sound reinforcement system were not considered. Therefore, the data obtained by the study suggested many possibilities for future work. First, since the tests were based on a limited RT range of 2.4-3.1 s, it will be useful to investigate the relation between speech intelligibility and room acoustics with a larger range, and the phenomena found in this paper must be confirmed in more large spaces. Second, since the results were obtained by playing the Mandarin syllables through a sound reinforcement system, the characteristics of the Mandarin sentence intelligibility and different language systems in large spaces should be further investigated. Third, further studies are needed to determine the boundary between large and ordinary space. Wang proposed the volume threshold of 125000 m^3 for large space according to the validity of the classical sound pressure level prediction model in spaces with different volumes [7]. From the perspective of speech intelligibility, there might also be a maximum tolerance value for STI prediction accuracy and a boundary between large and ordinary spaces. But this requires more research to provide sufficient sample size and credibility.

5. Conclusion

This study explored the influence factors of Mandarin (China mainland) speech intelligibility in large spaces. The results showed that both SNR and room acoustics affect the speech intelligibility, but the effect of room acoustics was affected by SNR. The speech intelligibility rapidly increased with the increase in SNR when the SNR was lower than 14.4 dB, and slightly changed when the SNR was higher than 14.4 dB. A relatively high SNR resulted in a lower spatial dispersion of speech intelligibility scores. The largest difference in the scores, which was approximately 25.2%, appeared when the SNR was approximately -5 dB. In terms of room acoustics, D_{50} was more relevant to the speech intelligibility than RT and EDT. When the SNR was less than 10 dB, the speech intelligibility had no uniform trend with the change in RT and EDT, but it increased with the growth of D_{50}. Moreover, when the SNR was higher than 10 dB, the change in room acoustics did not obviously affect the speech intelligibility.

This study has found that the corresponding relation between speech intelligibility and STI in large space was different from that in ordinary space. Thus, a new relation curve of mandarin speech intelligibility and STI in large spaces was established. The determination coefficient R square of this curve was 0.93 and the standard deviation was 5.76%, which indicated that STI was still suitable for intelligibility prediction after modification. Based on the relation curve, a new rating threshold of STI for large spaces was also proposed.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No.51678401). The author deep appreciated Prof. Zhu Peisheng for his helpful advice and warm support of this study. And thanks to the students who participated in the subjective speech intelligibility test.

Appendix A

Table A. Acoustical characteristics of the two rooms.

Stadium							
T_{20}	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-1000Hz
R1	2.32	2.77	2.49	1.99	1.59	1.27	2.24
R2	2.53	2.74	2.58	2.05	1.56	1.2	2.31
R3	2.46	2.65	2.46	2.01	1.53	1.15	2.24
R4	2.26	2.73	2.43	2.01	1.50	1.04	2.22
R5	2.33	2.76	2.48	2.02	1.51	1.12	2.25
R6	2.39	2.64	2.37	1.99	1.58	1.20	2.17
R7	2.48	2.63	2.33	1.98	1.42	1.04	2.16
EDT	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-1000Hz
R1	2.82	3.00	2.71	2.81	2.62	0.82	2.76
R2	2.30	2.84	2.69	2.59	2.02	1.14	2.64
R3	2.50	2.47	2.64	2.36	1.71	0.88	2.50
R4	2.47	2.29	2.16	1.80	0.61	0.31	1.98
R5	2.65	2.62	2.40	2.21	1.31	0.16	2.31
R6	2.55	2.78	2.58	2.41	1.74	0.30	2.50
R7	2.47	1.93	0.61	0.50	0.48	0.27	0.55
D_{50}	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-4000Hz
R1	0.41	0.27	0.27	0.50	0.78	0.87	0.60
R2	0.26	0.20	0.34	0.36	0.7	0.81	0.55
R3	0.33	0.15	0.32	0.26	0.65	0.80	0.51
R4	0.26	0.50	0.33	0.67	0.86	0.91	0.69
R5	0.46	0.49	0.55	0.73	0.87	0.93	0.77

R6	0.49	0.55	0.45	0.45	0.86	0.92	0.67
R7	0.35	0.64	0.84	0.90	0.92	0.66	0.83

Velodrome							
T_{20}	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-1000Hz
R1	2.91	3.00	3.02	2.91	2.55	1.63	2.96
R2	3.48	3.05	3.18	3.05	2.65	1.86	3.11
R3	2.92	2.97	3.28	3.10	2.71	1.92	3.19
R4	3.07	3.06	3.08	2.78	2.50	1.79	2.93
R5	3.14	3.08	3.13	3.13	2.78	1.98	3.13
R6	2.87	2.94	3.20	3.38	2.92	2.20	3.30
R7	3.48	2.96	3.02	3.00	2.52	1.66	3.01
R8	3.40	3.09	2.93	2.93	2.44	1.69	2.93
R9	3.32	3.02	3.08	2.99	2.63	1.87	3.04
EDT	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-1000Hz
R1	3.37	2.90	2.87	2.43	2.14	1.34	2.65
R2	3.60	3.22	3.03	3.28	3.09	0.53	3.16
R3	2.93	3.23	2.67	3.29	2.88	1.02	2.98
R4	3.02	2.70	2.37	2.06	2.22	0.74	2.22
R5	2.94	2.80	3.02	3.30	2.90	1.47	3.16
R6	3.37	3.45	3.49	3.77	3.50	1.65	3.63
R7	3.72	2.94	2.73	2.89	2.38	2.94	2.81
R8	2.76	3.00	2.78	2.73	2.42	2.28	2.76
R9	2.58	3.05	3.00	3.35	2.81	2.05	3.18
D_{50}	Centre frequency of octave bands						Average
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	500Hz-4000Hz
R1	0.19	0.60	0.41	0.66	0.72	0.82	0.65
R2	0.43	0.22	0.46	0.59	0.66	0.72	0.61
R3	0.24	0.43	0.56	0.60	0.62	0.79	0.64
R4	0.25	0.65	0.51	0.72	0.73	0.87	0.71
R5	0.35	0.55	0.54	0.65	0.76	0.89	0.71
R6	0.46	0.49	0.65	0.46	0.66	0.91	0.67
R7	0.28	0.39	0.42	0.37	0.52	0.68	0.50
R8	0.22	0.54	0.51	0.45	0.51	0.67	0.54
R9	0.28	0.21	0.27	0.33	0.45	0.5	0.39

Appendix B

Table B. SNR of each receiver in two spaces (The value in the upper left corner of each table represents the SNR of the sound input to the sound reinforcement system, and the data in the table represents the actual SNR measured at each receiver).

Stadium								
-5dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	-3.6	-3.7	-8.0	-6.3	1.7	-0.4	-7.0	-5.3
R2	-3.8	-3.7	-8.8	-7.2	1.4	-1.9	-7.9	-5.8
R3	-4.8	-4.8	-8.4	-6.0	1.1	-1.8	-5.7	-6.0
R4	-2.7	-4.4	-8.0	-4.8	0.4	-5.3	-7.7	-5.3
R5	-2.5	-4.6	-7.7	-5.3	-0.9	-3.7	-8.3	-5.3
R6	-2.3	-5.1	-8.1	-5.0	1.1	-2.1	-5.0	-5.4
R7	-3.2	-5.7	-7.1	-6.0	0.1	-1.8	-9.0	-6.1
0dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	2.1	0.8	-2.3	0.1	6.2	3.4	-0.5	0.3
R2	2.0	2.2	-2.6	-0.1	5.9	3.0	-2.4	0.3
R3	1.0	0.5	-3.1	-0.8	5.8	3.3	-1.0	-0.8
R4	3.0	0.2	-2.8	0.1	4.6	0	-2.6	-0.4
R5	3.1	0.3	-2.1	0.5	5.5	1.5	-3.4	0
R6	2.4	-0.1	-3.0	-0.1	5.9	2.6	-0.9	-0.5
R7	1.5	-0.7	-3.0	-1.4	6.4	3.9	-4.2	-1.7
5dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	5.9	6.1	2.4	3.8	10.0	8.2	1.4	5.0
R2	5.1	6.1	1.6	4.4	10.2	7.1	1.1	4.3
R3	4.8	5.5	1.5	3.1	8.7	6.9	1.8	3.7
R4	6.8	5.5	1.8	5.2	11.0	7.0	2.4	4.6
R5	6.8	4.4	2.3	4.9	10.5	7.4	2.0	4.4
R6	7.6	4.5	1.7	4.2	10.6	7.4	2.7	2.1
R7	6.4	4.0	1.3	3.2	10.0	7.7	0.2	2.7
10dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	12.0	11.0	6.6	8.0	14.1	11.4	5.2	9.3
R2	11.3	10.6	6.3	8.3	14.2	10.6	4.3	9.0
R3	11.4	11.5	6.2	8.1	13.3	10.5	6.2	9.0
R4	13.6	10.1	7.2	8.0	13.1	10.0	6.1	9.2
R5	14.9	10.9	6.0	8.1	13.4	10.1	4.4	8.7
R6	11.7	8.1	4.8	7.8	13.2	10.3	6.6	7.7

R7	10.1	8.1	4.7	7.0	13.8	11.7	3.4	6.3
15dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	17.8	18.8	13.9	15.0	19.7	18.4	9.7	15.5
R2	17.9	18.4	13.3	14.6	19.0	17.1	10.2	16.1
R3	17.2	19.0	13.1	14.7	19.5	17.9	10.5	16.0
R4	20.6	17.3	14.5	15.6	20.1	16.5	12.0	16.6
R5	21.4	17.5	13.0	14.4	18.6	16.1	9.2	15.4
R6	18.2	15.5	12.0	14.3	19.6	16.7	10.5	15.1
R7	16.8	15.0	11.2	13.2	20.8	16.3	8.6	12.8

Velodrome								
-5dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	-5.1	-4.6	-4.8	-1.9	-5.0	-4.4	-5.1	-4.4
R2	-4.7	-4.9	-4.7	-6.8	-6.9	-5.5	-4.9	-5.2
R3	-5.4	-5.0	-4.3	-4.5	-4.7	-4.8	-4.8	-4.8
R4	-5.1	-4.6	-5.0	-4.5	-4.7	-4.9	-4.8	-4.8
R5	-4.6	-5.0	-4.3	-4.1	-2.7	-4.9	-5.3	-4.8
R6	-1.0	-4.8	-4.4	-3.8	-2.2	-4.3	-4.8	-4.4
R7	-4.6	-5.3	-4.6	-5.4	-4.7	-5.4	-6.4	-5.1
R8	-5.0	-5.2	-5.6	-5.5	-5.4	-6.1	-6.5	-5.2
R9	-4.8	-5.3	-5.8	-3.3	-3.4	-5.5	-6.0	-5.3
0dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	1.1	1.4	1.3	5.0	1.2	1.3	0.4	1.7
R2	-0.6	-2.1	0.2	0.9	0.9	1.2	1.4	-1.4
R3	1.4	-1.5	0.8	-0.5	-0.2	-0.7	-0.4	-1.0
R4	0.7	3.5	1.4	2.0	1.7	1.4	1.7	2.4
R5	0.2	-0.8	-0.1	-3.3	-0.6	-2.2	-1.6	-0.9
R6	0.4	-3.5	-0.2	-3.2	-1.2	-1.9	1.7	-2.4
R7	1.7	0.8	2.1	0.8	2.3	1.1	0.6	1.4
R8	1.6	0.5	0.1	0.5	-0.2	0.4	0.5	0.7
R9	1.5	0.8	0.6	2.2	1.8	1.7	1.1	0.9
5dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	5.1	6.9	5.7	7.2	4.3	4.7	4.3	6.7
R2	5	4.8	4.7	5.4	4.6	4.5	4.3	4.8
R3	4.4	4.3	5.0	4.8	4.3	4.2	4.2	4.5
R4	4.0	5.7	4.5	5.1	4.9	4.8	4.6	5.2
R5	6.2	7.1	6.3	5.8	5.6	5.4	4.1	6.5

R6	7.2	5.7	5.6	4.8	2.5	5.0	3.4	5.6
R7	5.7	3.9	5.0	4.0	5.1	5.2	4.3	4.6
R8	5.6	4.1	3.0	3.3	4.2	4.2	4.2	4.2
R9	5.0	4.8	4.4	5.9	3.6	3.7	4.7	4.8
10dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	11.3	11.9	10.8	13.2	10.1	10	9.3	11.7
R2	9.7	8.6	9.6	11.1	10.9	10.9	10.2	9.0
R3	9.9	11.4	12.1	9.3	8.9	7.7	7.3	11.1
R4	11.3	10.5	9.9	9.6	10.6	9.3	10.0	10.4
R5	9.9	10.3	10.6	9.8	8.1	9.3	10.8	10.3
R6	9.4	9.4	8.7	7.3	3.3	9.0	10.3	9.0
R7	10.2	10.4	11.4	11.4	10.6	10.5	9.3	10.8
R8	9.3	9.0	7.3	8.9	9.0	9.8	9.3	8.4
R9	11.5	11.3	10.2	11.4	11.4	11.0	10.3	11.1
15dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	17.1	18.8	16.5	17.5	14.9	16.1	14.5	18.2
R2	14.0	12.7	13.1	14.4	11.2	14.6	14.8	12.8
R3	18.7	16.1	16.6	16.0	15.6	15.0	14.6	16.2
R4	16.5	16.7	15.8	16.1	16.4	16.4	16.5	16.1
R5	13.5	14.8	15.0	12.8	4.2	9.9	13.5	14.4
R6	10.9	13.9	12.9	8.7	4.1	12.4	13.8	12.8
R7	17.4	16.2	17.0	15.7	15.6	16.2	15.8	16.6
R8	17.6	18.5	16.8	14.8	15.6	16.0	15.9	17.5
R9	18.0	19.2	17.2	17.2	16.8	17.3	17.1	18.4
25dB	Centre frequency of octave bands							
Receivers	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	L_{Aeq}
R1	20.1	32.2	28.1	26.7	20.9	27.6	24.5	27.9
R2	17.2	27.3	24.6	22.1	20.4	26.9	24.3	25.5
R3	20.5	29.5	27.8	21.9	19.3	25.6	23.7	24.9
R4	17.6	29.1	24.6	22.6	20.4	26.8	24.7	25.6
R5	16.7	27.8	25.7	18.1	11.9	18.6	26	24.3
R6	11.9	20.8	18.1	9.6	4.1	16.2	31.1	16.9
R7	20.8	30.3	26.2	21	21.2	24.2	24.3	24.1
R8	21.6	30.3	23.9	19.3	20.1	22.3	22.8	22.9
R9	20.4	29.2	22.8	18.8	18.6	19.7	18.4	23.0

References

[1] J. S. Bradley, Speech intelligibility studies in classrooms, The Journal of the Acoustical Society of America 80 (3) (1986) 846–854.

[2] J. S. Bradley, H. Sato, The intelligibility of speech in elementary school classrooms, The Journal of the Acoustical Society of America 123 (4) (2008) 2078–2086.
[3] J. S. Bradley, R. Reich, S. Norcross, On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility, The Journal of the Acoustical Society of America 106 (4) (1999) 1820–1828.
[4] S. R. Bistafa, J. S. Bradley, Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics, The Journal of the Acoustical Society of America 107 (2) (2000) 861–875.
[5] P. Jianxin, Chinese speech intelligibility at different speech sound pressure levels and signal-to-noise ratios in simulated classrooms, Applied Acoustics 71 (4) (2010) 386–390.
[6] J. Peng, Chinese syllable and phoneme identification in noise and reverberation, Archives of Acoustics 39 (4) (2014) 483–488.
[7] C. Wang, H. Ma, Y. Wu, J. Kang, Characteristics and prediction of sound level in extra-large spaces, Applied Acoustics 134 (2018) 1–7.
[8] B. Yegnanarayana, B. S. Ramakrishna, J. Anderson, Intelligibility of speech under nonexponential decay conditions, The journal of the Acoustical Society of America 58 (4) (1974) 853–857.
[9] T. Lewers, J. Anderson, Some acoustical properties of st paul's cathedral, london, Journal of Sound and Vibration 92 (2) (1984) 285–297.
[10] A. C. Raes, G. G. Sacerdote, Measurements of the acoustical properties of two roman basilicas, The journal of the Acoustical Society of America 25 (5) (1953) 954–961.
[11] L. Álvarez-Morales, T. Zamarreño, S. Girón, M. Galindo, A methodology for the study of the acoustic environment of catholic cathedrals: Application to the cathedral of malaga, Building and Environment 72 (2014) 102–115.
[12] GB/T 15508-1995. Acoustic-Speech articulation testing method, 1995. (Standard of People's Republic of China, 1995).
[13] IEC 60268-16,4rd ed. Sound System Equipment–Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index. International Electrotechnical Commission; 2011.
[14] T. Houtgast, H. J. Steeneken, The modulation transfer function in room acoustics as a predictor of speech intelligibility, Acta Acustica United with Acustica 28 (1) (1973) 66–73.
[15] H. J. M. Steeneken, T. Houtgast, A physical method for measuring speech-transmission quality, The Journal of the Acoustical Society of America 67 (1) (1980) 318–326.
[16] T. Houtgast, H. J. Steeneken, A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria, The Journal of the Acoustical Society of America 77 (3) (1985) 1069–1077.
[17] H. J. Steeneken, T. Houtgast, Mutual dependence of the octave-band weights in predicting speech intelligibility, Speech communication 28 (2) (1999) 109–123.
[18] H. J. Steeneken, T. Houtgast, Phoneme-group specific octave-band weights in predicting speech intelligibility, Speech Communication 38 (3-4) (2002) 399–411.
[19] S. J. v. Wijngaarden, H. J. Steeneken, Objective prediction of speech intelligibility at high ambient noise levels using the speech transmission index, in: Sixth European Conference on Speech Communication and Technology, 1999.
[20] H. J. Steeneken, T. Houtgast, Validation of the revised stir method, Speech Communication 38 (3-4) (2002) 413–425.
[21] P. Jianxin, Relationship between chinese speech intelligibility and speech transmission index using diotic listening, Speech Communication 49 (12) (2007) 933–936.
[22] P. Zhu, F. Mo, J. Kang, Relationship between chinese speech intelligibility and speech transmission index under reproduced general room conditions, Acta Acustica united with Acustica 100 (5) (2014) 880–887.
[23] J. Kang, Comparison of speech intelligibility between english and chinese, The Journal of the Acoustical Society of America 103 (2) (1998) 1213–1216.
[24] ISO 3382. Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms. International Organisation for Standardisation, 2008.
[25] GB/T 9396-1996. Methods of measurement for main characteristics of loudspeakers. (Standard of People's Republic

of China, 1996).
[26] GB/T 13504-2008. Diagnostic rhyme test (DRT) method of Chinese articulation. (Standard of People’s Republic of China, 2008).
[27] S.K. Tang, Speech related acoustical parameters in classrooms and their relationships, Applied acoustics 69 (12) (2008) 1318–1331.
[28] J. Chen, H. Ma, Analysis of the Acoustic Environment in Large Interior Spaces, South Architecture (3) (2019) 48-53.
[29] G. A. Studebaker, R. L. Sherbecoe, Monosyllabic word recognition at higher-than-normal speech and noise levels , The Journal of the Acoustical Society of America 105 (4) (1998) 2431-2444.
[30] J. R. Dubn, A. R. Horwitz, J. B. Ahlstrom, Recognition of filtered words in noise at higher-than-normal levels: Decreases in scores with and without increases in masking, The Journal of the Acoustical Society of America 118 (2) (2005) 923-933.
[31] H. Sato, M. Morimoto, R. Ota, Acceptable range of speech level in noisy sound fields for young adults and elderly persons, The Journal of the Acoustical Society of America 130 (3) (2011) 1411-1419.