Comparisons between simulated and in-situ measured speech intelligibility based on (binaural) room impulse responses

Peisheng Zhu[a], Fangshuo Mo[b]*, Jian Kang[c]* , Guofeng Zhu[a]

[a] *School of Architecture and Fine Art, Dalian University of Technology, Dalian 116023, China*
[b] *Institute of Acoustics, Tongji University, Shanghai 200092, China*
[c] *School of Architecture, University of Sheffield, Sheffield S10 2TN, UK*

* Corresponding author

Abstract: This study systematically compares acoustic simulation and in-situ measurement in terms of speech transmission index (STI), speech intelligibility scores and relationship curves when considering (binaural) room impulse response and four general room conditions, namely, an office, a laboratory, a multimedia lecture hall and a semi-anechoic chamber. The results reveal that STI can be predicted accurately by acoustic simulation (using room acoustics software ODEON) when there is a good agreement between the virtual models and the real rooms and that different reverberation time (RT) and signal-to-noise ratio (SNR) may exert less significant influence on the simulated STI. However, subjective intelligibility may be overestimated when using acoustic simulation due to the head-related transfer function (HRTF) filter used, and the score bias may be minimal and difficult to detect in everyday situations. There is no obvious score tendency caused by different RT, though with the decrease in the SNR, score bias may increase. Overall, considering that the accurate acoustic modelling of rooms is often problematic, it is difficult to obtain accurate speech intelligibility prediction results using a simulation technique, especially when the room has not yet been built.

Key words:Speech intelligibility; simulation; in-situ measurement; (binaural) room impulse responses

2015 Applied Acoustics
Date Received: 17 January 2014 **Date Accepted**: 10 April 2015
Publish online: 24 April 2015

1. Introduction

Speech intelligibility is an important metric and can be used to evaluate the sound transmission quality of auditoriums. The assessment of speech intelligibility mainly includes subjective evaluation and objective evaluation [1, 2]. However, performing such measurements in real rooms has limitations, such as schedule conflicts, and it is difficult to perform a speech intelligibility test simultaneously with a large number of subjects at a single receiver position. In recent years, the rapid progress in the acoustic simulation technique offers a potential solution to these limitations and provides an unlimited capacity to reproduce the same listening environments while also making it possible for speech intelligibility to be assessed in a room before it is built [3, 4, 5, 6, 7, 8]. However, before it can be used with confidence, the acoustic simulation technique must be validated in comparison with in-situ measurement in real rooms.

Subjective intelligibility tests were performed in virtual and real classrooms, and the results were compared by Yang and Hodgson [3] using the CATT-Acoustics prediction and auralization system. The results showed that auralized subjective intelligibility tests were found to be reliable if the classroom was neither very absorptive nor noisy. However, in their study, the comparison of the objective evaluation metric speech transmission index (STI) was not involved. Subjective intelligibility tests were also performed in virtual and real classrooms, and the results were compared by Hodgson et al. [4] using the CATT-Acoustics and ODEON prediction and auralization system. The results suggested that auralization is not accurate in the case of high noise or low reverberation. The comparison of the objective evaluation metric STI, however, was still not involved in this study. Peng et al. [5, 6, 7, 8] made many meaningful attempts on using acoustic simulation technique to assess the speech intelligibility of Chinese. The results showed that the relationship between the subjective intelligibility scores and STI can be better reflected based on acoustic simulation, which is an effective method for the evaluation of speech intelligibility. However, their conclusions obtained are mainly based on simulation, and in-depth comparison and validation with the in-situ measurement are still needed. Overall, there is still a lack of study on the systematic comparison and validation of simulation technique for the evaluation of speech intelligibility.

The aim of this study is therefore to systematically compare the simulated speech intelligibility scores, STI and the curve thus produced with that of the in-situ measured. Finally, the influence factors for simulation bias, if any, should be considered carefully.

2. Methods

This section starts with selecting four general rooms and then, based on the room impulse response (IR) and binaural room impulse response (BRIR) and the STI and Chinese speech intelligibility scores of 12 receiver positions, a total of 48 listening environments in the four general rooms were obtained by the two types of methods: in-situ measurement and acoustic simulation. A general flowchart of this study is given in Figure 1.

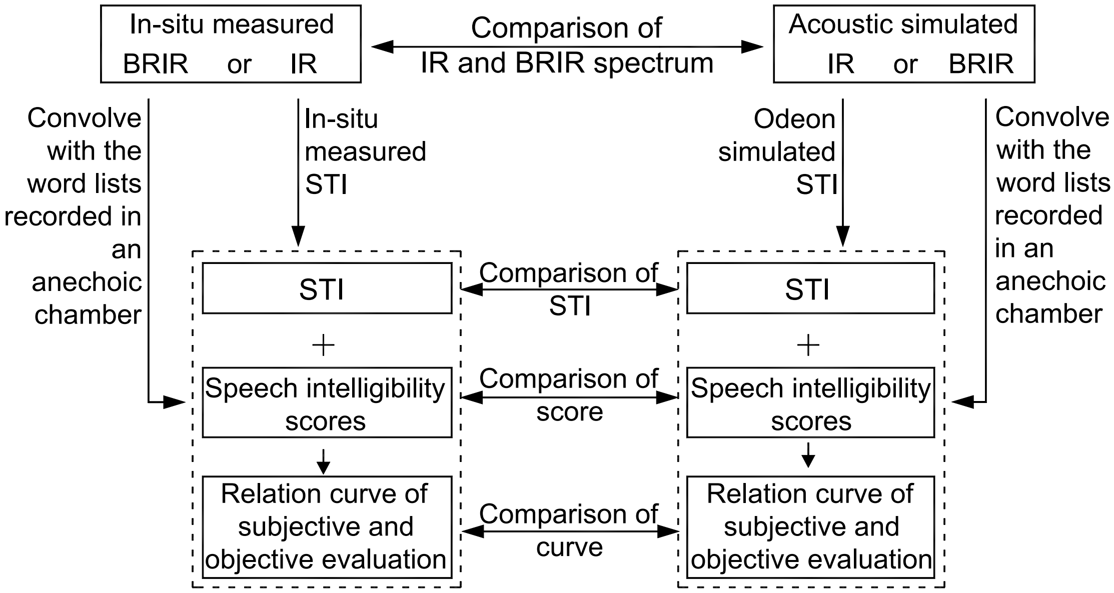
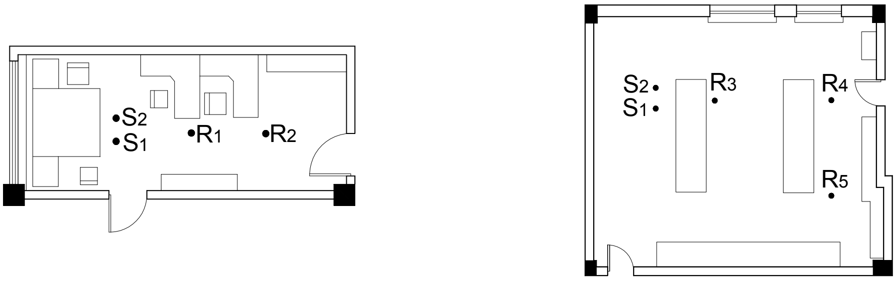


Fig. 1. Flowchart of the general experimental procedure.

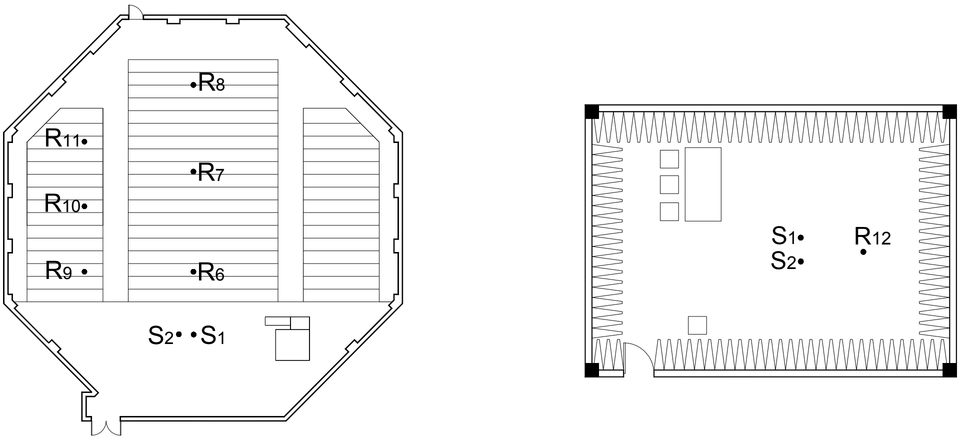
2.1 Experimental arrangement

Four general rooms were selected as the test rooms in this study, including an office, a laboratory, a multimedia lecture hall and a semi-anechoic chamber (with one desk and four chairs inside), of which, the office, laboratory and semi-anechoic chamber are rectangular, and the multimedia lecture hall is octagon. There are two receiver positions arranged in the office, three receiver positions in the laboratory, six receiver positions in the multimedia lecture hall and one receiver position in the semi-anechoic chamber. The layout of the receiver positions and the sound sources are shown in Figure 2.

To obtain a wide range of the STI, an interference noise source (monitor loudspeaker GENELEC 8020B) was placed at a distance of 0.5 m beside the signal source. A dodecahedral sound source was not used as an interference noise source in this experiment because, for the dodecahedral sound source, there was no main radiation and the directivity changed with orientations, the equalisation and calibration was difficult [9], and room acoustics software could hardly simulate a real dodecahedral sound source. Accordingly, these factors may exert significant influence on the comparison results. In an anechoic chamber, the sound pressure level (SPL) on the front axis at 1 m of the signal source (artificial mouth GRAS 44AA) was set at 60 dBA [10]. The noise source reproduced a males spectra shaped [10] pink noise, and the SPL was adjusted simultaneously to make the positions 1 m away from the two sound sources correspond to four distinct relative background noise levels (RBNLs): 5 dB, 0 dB, -10 dB, and -20 dB. The SPL on the front axis at 1 m of the monitor loudspeaker 8020B was set as 65 dBA, 60 dBA, 50 dBA, and 40 dBA, respectively. The RBNL equals the signal-to-noise ratio (SNR) in a noiseless anechoic chamber; however, due to the influence of different reflections and sound source directivity patterns, and possible environmental noise, the RBNL does not equal the actual SNR at the R_1-R_{12} receiver positions. The signal source and the noise source, preset in an anechoic chamber, were placed in the corresponding sound source positions in the test rooms and at each receiver position, the STI, IR and BRIR as well as the operational speech level and background noise levels were each measured in turn. Both of the sound source systems were equalised using their inverse filter systems calculated from the impulse responses measured on the front axis of the sources in an anechoic chamber [9].



Office Laboratory



Multimedia lecture hall Semi-anechoic chamber

Fig. 2. The layout of the receiver positions and the sound sources in the four rooms. S_1 is the signal source and S_2 is the noise source. R_1–R_{12} are the receiver positions. The height of the receiver positions is 1.2 m, and the height of the loudspeaker is 1.5 m.

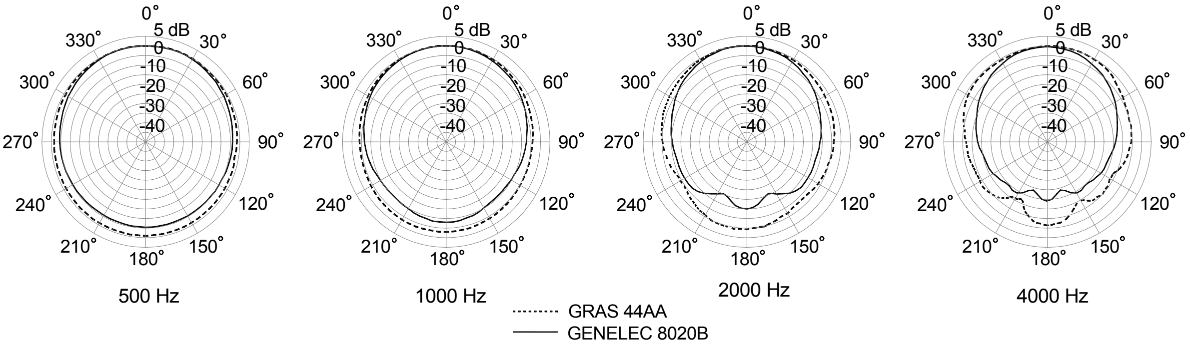
2.2 Virtual room modelling

Four room models were erected corresponding to the four real rooms using the ODEON version 12.0 [11] room acoustics software. During the simulation, the virtual signal source and the virtual interference noise source, namely, virtual-44AA and virtual-8020B, respectively, were erected in ODEON using the horizontal and vertical directivity patterns of the artificial mouth GRAS 44AA and the monitor loudspeaker 8020B. In the ‘Directivity Polar Plot Editor’ menu, both the virtual-44AA and the virtual-8020B were marked with ‘Natural’, the horizontal and vertical directivity patterns of each octave band were established, and the ‘+EQ’ of each octave band was adjusted to ensure that, in the ‘Point Source Editor’ menu, the SPL of the virtual-44AA on the front axis at 1 m (which should be 20 dB higher than the SPL on the front axis at 10 m) was set to the same octave band (from 125 to 8000 Hz) SPL as that measured on the front axis at 1 m in an anechoic chamber when reproducing a composite signal of seven half-octave band carriers without modulation with a SPL of 60 dBA using the artificial mouth 44AA. The SPL of the virtual-8020B on the front axis at 1 m was set to the same octave band (from 125 to 8000 Hz) SPL as that measured on the front axis at 1 m in an anechoic chamber when reproducing a male spectra shaped [10] pink noise with a SPL of 60 dBA using the monitor loudspeaker 8020B. In addition, in the ‘Point Source Editor’ menu, the ‘+EQ’ of each octave band was set to 0 dB for both virtual-44AA and virtual-8020B, the ‘Overall gain’ was set to 0 dB for the virtual-44AA, and 5 dB, 0 dB, -10 dB, and -20 dB, respectively, for the virtual-8020B. The octave band SPL for the artificial mouth GRAS 44AA and the monitor loudspeaker 8020B measured on the front axis at 1 m in an anechoic chamber are presented in Table 1. In Figure 3, the horizontal and vertical directivity patterns of the monitor loudspeaker 8020B and the artificial mouth GRAS 44AA at 500, 1000, 2000, and 4000 Hz are shown based on the data provided by the manufacturer of the monitor loudspeaker 8020B, while the data for the artificial mouth 44AA were obtained through measurements taken in an anechoic chamber for this study. The corresponding receiver and sound source positions for the four virtual models were set in accordance with Figure 2.

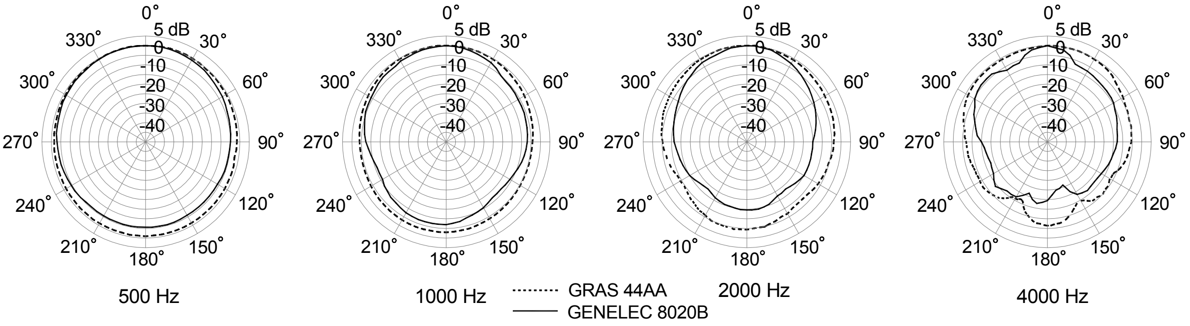
The acoustic parameters were obtained based on hybrid methods combining ray tracing and image source modelling in ODEON [11]. The acoustic properties, such as the absorption and scattering coefficients of the materials in the four virtual rooms, were adjusted carefully to ensure good agreement between the simulated and measured acoustic parameters for each receiver position. The simulated and measured early decay time (EDT), reverberation time (RT), clarity (C_{80}) and SPL for each octave band (from 125 to 8000 Hz) at the R_1 through R_{12} receiver positions are presented in Appendix A, where the virtual-44AA and artificial mouth GRAS 44AA were employed. As the simulated and measured acoustic parameter results, with the sources being the virtual 8020B and the GENELEC 8020B, were similar to those of the virtual-44AA and the GRAS 44AA, they are not listed. Appendix A indicates that good agreement was found between the simulated and measured RT and that, compared with the measured values, most prediction differences were generally within approximately 0.10 s. Furthermore, it is evidenced from Appendix A that the largest difference, 0.20 s, appeared at the receiver position R_6 at the central frequency of 1000 Hz and that, with respect to EDT, good agreement was found between the simulated and measured values, while the prediction differences were greater than those of RT, especially at the receiver positions R_6, R_7 and R_8. The largest difference, 0.31 s, appeared at the receiver position R_8 at the central frequency of 500 Hz. With respect to C_{80}, good agreement was found in the office and the laboratory, most prediction differences were generally within approximately 1 dB, the largest difference was rarely more than 2 dB. Furthermore, the prediction differences for the multimedia lecture hall and the semi-anechoic chamber were greater than those for the office and the laboratory, especially when the frequency was low. The greatest difference attained was 6.59 dB in the multimedia lecture hall. This appeared at the receiver position R_8 at the central frequency of 125 Hz where it reached 10.47 dB in the semi-anechoic chamber and appeared at the receiver position R_{12} at the central frequency of 250 Hz. With respect to the SPL, most prediction differences were generally within 2 dB and often within 1 dB. The greatest difference reached was 5.04 dB, which appeared at the receiver position R_8 at the central frequency of 4000 Hz.

Table 1. The octave band (from 125 to 8000 Hz) SPL for the artificial mouth GRAS 44AA and the monitor loudspeaker 8020B measured on the front axis at 1 m in an anechoic chamber, as used for the calibrations of the virtual-44AA and virtual-8020B.

Frequency band (Hz)	125	250	500	1000	2000	4000	8000	L_A
SPL for the artificial mouth GRAS 44AA (dB)	63.11	62.62	59.22	53.19	47.00	41.61	35.65	60
SPL for the monitor loudspeaker 8020B (dB)	63.07	62.75	59.19	52.79	47.80	41.61	35.58	60



(a) Horizontal directivity patterns for GRAS 44AA and GENELEC 8020B



(b) Vertical directivity patterns for GRAS 44AA and GENELEC 8020B

Fig. 3. The horizontal and vertical directivity patterns of the artificial mouth GRAS 44AA and monitor loudspeaker 8020B.

2.3 STI simulation and in-situ measurement

The STI can be derived from ODEON using the virtual-44AA signal source in the virtual room to a given receiver position [11]. The STI parameter takes into account the background noise level, which can be adjusted by the user in the 'Room Setup' menu in ODEON [11]. For the STI parameter to be valid, it is important to adjust the background noise accordingly. In this study, the background noise levels calculated by ODEON using the virtual-8020B noise source were inputted into the ROOM SETUP menu, and the STI was then derived from ODEON.

There are two in-situ STI measurement methods recommended by IEC 60268-16 [10], namely, the direct method, which is based on signal modulation, and the indirect method, which is based on impulse response. With the exception of certain commercial STIPA meters that adopt the direct measurement method [12], the STI measurement platforms available, such as Dirac [13], Aurora [14], etc. [15], all adopt the indirect measurement method. In this study, Aurora was used as the STI measurement platform, and the test signal for the measurement of operating speech level and the SPL calibration of signal source 44AA in an anechoic chamber, the noise signal for the measurement of background noise level and the SPL calibration of interference noise source 8020B in an anechoic chamber were also compiled. The test signal was a compound signal of seven half octave-band carriers without frequency modulation but including the male spectrum as described in IEC [10], and the noise signal was a males spectra shaped [10] pink noise. The operating speech level was measured at the receiver position in the four rooms by reproducing the test signal via the signal source 44AA, and the background noise level was measured at the receiver position in the four rooms by reproducing the interference noise signal via the noise source 8020B at four different SPLs. Using the measured operating speech level and background noise level, the correction SNR was obtained and the STI was then calculated.

The measurement system includes signal source GRAS 44AA, noise source GENELEC 8020B, audio interface B&K ZE-0948, microphone B&K 4189 (power supply is B&K 1704), sound recording software Audition (v3.0) and binaural microphone B&K 4101-A (for the measurement of BRIRs). To reduce the inaccuracy caused by the measurement system, loop calibration was performed to ensure that it was a linear time-invariant (LTI) system without harmonic distortions.

2.4 Subjective intelligibility test for simulation and in-situ measurement

2.4.1 Test materials

Chinese was used as the test speech in this study. As previous study [16] showed that the curve of Chinese speech intelligibility scores corresponding to STI was approximately similar to that of English, and in this paper the speech intelligibility tests were exactly the same for the two methods and only the comparison results were considered, the results are also suitable for Western languages. Subjective Chinese speech intelligibility tests are usually conducted in one of two ways: a Diagnostic Rhyme Test (DRT), specified by GB/T 13504-2008 [17] and SJ2467-84 [18], or a PB word dictation test, specified by GB/T 15508-1995 [19]. DRT is a type of closed set test, for which higher scores are more easily obtained [20]. Therefore, the discrimination of this test is lower at high intelligibility levels, which does not typically reflect the relationship between subjective and objective evaluation. Therefore, Chinese PB word lists specified by GB/T 15508-1995 [19] were used as the test material in this study. Each Chinese PB word list comprises 25 three-syllable rows, and the three syllables in each row were randomly arranged, thus a total of 75 syllables were used and embedded in a carrier phrase. All the test word lists spoken by a male speaker were recorded at a rate of 4 words per second in an anechoic chamber, and were used as the training PB word lists according to GB/T 15508-1995 [19], to guarantee all the subjects were familiarised with the PB word lists.

The subjective intelligibility test was conducted by reproduction through headphones. For the in-situ measurement method, the speech material was the signal convolving the in-situ measured BRIR of each receiver position with the PB word list recorded in an anechoic chamber. Additionally, four types of background noise from each receiver position were recorded in the real room and mixed. The BRIRs were measured at the receiver position using a binaural microphone B&K 4101-A with the ear canals blocked placed on the author’s head and pointing to the signal source. Specifically, the signal source artificial mouth GRAS 44AA was used. For the acoustic simulation method, the BRIRs were obtained by combining the simulated IRs with that of a virtual listener, as defined by head-related transfer functions (HRTFs). The virtual listener was modelled using the HRTF of Subject_021 (the default HRTF in ODEON, which is the Kemar dummy head with blocked ear canals) while facing the speaker [11]. The auralisation procedure involved the creation of the BRIR, separately for the virtual-44AA signal source and the virtual-8020B noise source. The BRIR was then convolved from the virtual-44AA with the word lists recorded in an anechoic chamber and the BRIR from the virtual-8020B with the noise signal (a males spectra shaped [10] pink noise) for each listening environment, respectively. The convolved wav files were then mixed, ensuring the correct relative levels, to give the speech intelligibility test signals [11]. Finally, the amplitudes of the simulated speech material lists were adjusted, after applying a headphone equalisation filter, to ensure the same speech level as the corresponding in-situ measurement speech material lists. The headphone used in the test is the Sennheiser HD-600, the corresponding power amplifier is the Rane-HC4s, and the audio interface is the

B&K ZE-0948. HD-600 is an open-air headphone and is, therefore, inclined to be affected by ambient noise. Accordingly, the test was conducted in a semi-anechoic chamber.

2.4.2 Test procedures

Sixteen testing juries with four subjects per jury (sixty-four subjects) participated in the subjective test. The sixteen juries were divided into two groups. One group of eight juries listened to the speech material lists for the R_1 through R_6 receiver positions using the two methods, while the other group of eight juries listened to the speech material lists for the R_7 through R_{12} receiver positions using the two methods. For one group of eight juries, every two juries composed a test pair. Thus, there were four test pairs obtained, which corresponded to the arranged four RBNLs. Each jury in a test pair listened to twelve speech material lists, six lists for the simulation method and six lists for the in-situ measurement method. Accordingly, twelve different word lists recorded in an anechoic chamber were used and divided into two groups, one for each of the two methods. These twelve word lists were also used by the other group of eight juries. Two juries in a test pair listened to the speech material lists with the same receiver position and same RBNL condition, except for the six word lists that were interchanged for the two methods. Finally, each sound environment for the two methods received eight scores from the paired juries. Thus, with four scores from each jury using the two different word lists, the eight scores were then averaged. As the repeated use of word lists may significantly influence the testing results [21], the twelve word lists listened that each jury listened to were different in this study. To reduce the inaccuracy caused by listening sequence and memory, the twelve speech material lists listened to by one jury were carefully arranged to avoid, as much as possible, the same adjacent RBNL and method. The speech material lists and the listening sequence for the R_1 through R_6 receiver positions for each of the two methods listened to by one of the groups of eight juries are listed in Appendix B.

All subjects were junior students at a university between 20-22 years old, and the hearing threshold level (HL) of each subject was in normal range. The pre-experiment indicated that the scores from different subjects may vary greatly even under identical conditions, especially when the listening environment is poor. Because the hearing threshold level of each subject was examined prior to the test, the disparity may be related to the personal quality of the subjects. To reduce this inaccuracy, systematic training was provided to all of the subjects prior to the test to ensure that they understood the entire test process and could master the key points.

Many studies have debated the ability of headphones to produce the same listening effect as that of actual listening on site [22, 23]. Therefore, to guarantee the reliability of the reproduction through headphones, the following items were carefully considered in the test: whether the spectrum of the reproduction speech signals were the same as the actual recording spectrum on site, whether the SPL of the reproduction was exactly the same as the SPL on site, whether the pre-set RBNLs were truly reflected, as well as the equalisation processing of the reproduction system, etc. In this study, the conditions can be realised via the signal processing program compiled based on binaural technology [24]. The comparison results between the signal recorded by an artificial head at the receiver position R_5 (-10 dB RBNL) when the word list was reproduced through the artificial mouth GRAS 44AA and the signal recorded by the artificial head when the corresponding test speech material list was reproduced through the headphones are provided in Table 2. It shows that the reproduced spectrum and speech level were similar to that on site.

The SPL calibration of word lists recorded in an anechoic chamber for convolution is also a factor that can easily cause inaccuracy. Generally, speech signals are not continuous and contain numerous pauses; the calibration method for its SPL is based on the removal of the silent parts of the speech signals, i.e. the pauses between the words [25], which were used to calibrate the word lists in this study. Naturally, the SPL of all other test signals, such as the noise signal and test signal for the direct method, were accurately controlled at the beginning of signal generation by filtering, to ensure the complete equivalence of the SPL between the test signals and speech signals.

Table 2. The comparison results between the signal recorded by an artificial head at the receiver position R_5(-10 dB RBNL) when the word list was reproduced through the artificial mouth GRAS 44AA (Sig. 1) and the signal recorded by the artificial head when the corresponding test speech material list was reproduced through the headphones (Sig. 2). The SPL in the table is the overall RMS values of the recorded signal, including both speech and noise.

Frequency band (Hz)		125	250	500	1000	2000	4000	8000	L_A
Sig. 1 (dB)	Left ear	50.93	55.37	53.90	52.78	47.06	52.02	46.66	58.36
	Right ear	50.47	55.06	55.10	53.39	47.94	53.02	48.06	59.23
Sig. 2 (dB)	Left ear	50.34	54.85	53.56	52.08	46.97	51.99	44.29	57.88
	Right ear	49.97	54.64	55.14	53.34	47.90	53.02	46.03	59.03
Difference between Sig. 2 and Sig. 1 (dB)	Left ear	-0.59	-0.52	-0.34	-0.70	-0.09	-0.03	-2.37	-0.47
	Right ear	-0.50	-0.42	0.04	-0.05	-0.04	0.00	-2.03	-0.20

3. Results and Discussion

We first compare the STIs for each sound environment of the two methods and then compares the speech intelligibility scores for each sound environment of the two methods. The curve between the Chinese speech articulation scores and the STIs for each of the two methods are subsequently established and compared. The factors that may have influenced the simulation inaccuracy are then discussed. The STI values and speech intelligibility scores for each of the two methods, as well as their differences, in the 48 sound environments of the four rooms are presented in Appendix C.

3.1 Comparison of STI

The differences of simulated and in-situ measured STIs are presented in Figure 4. Based on Figure 4, it is noted that only three STI differences, -0.032, 0.037 and 0.054, exceed a just noticeable difference (JND) [26]. Furthermore, they appear when the RBNL is -20 dB at the receiver position R_6 and when RBNLs are -10 dB and -20 dB at the receiver position R_8. The mean value of the 48 STI differences is -0.00163. To further evaluate these inaccuracies, the paired-sample T-test statistical method is used to confirm whether there exists significant difference between the simulated and measured STIs. The p-value is found to be 0.546, indicating that there is no significant variation. These results reveal that the STI can be predicted accurately when there is a good agreement between the virtual models and the real rooms.

The mean STI differences of the two methods with different RBNLs for the four rooms are presented in Figure 5. It is noted that all mean STI differences are below one JND. The maximum STI difference variation range for different RTs is 0.022, which is less than one JND. This difference appears between the semi-anechoic chamber and the laboratory when the RBNL is -10 dB, indicating that the influence exerted by different RTs is not highly significant. The maximum STI difference variation range for different RBNLs is 0.014 in the semi-anechoic chamber, which, again, is less than one JND, indicating that the influence exerted by different SNRs is not highly significant.

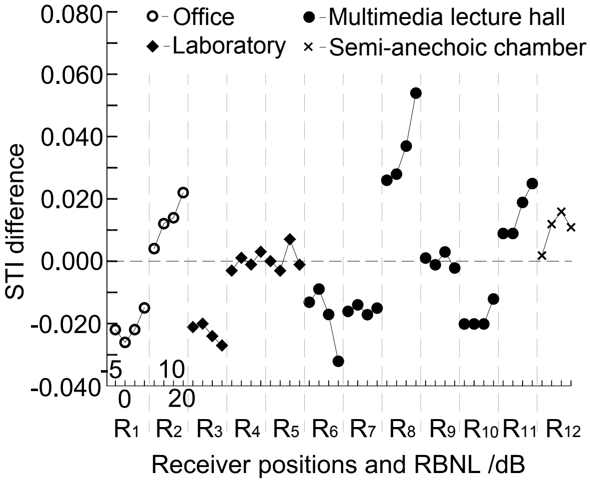


Fig. 4 The difference between the simulated and in-situ measured STIs.

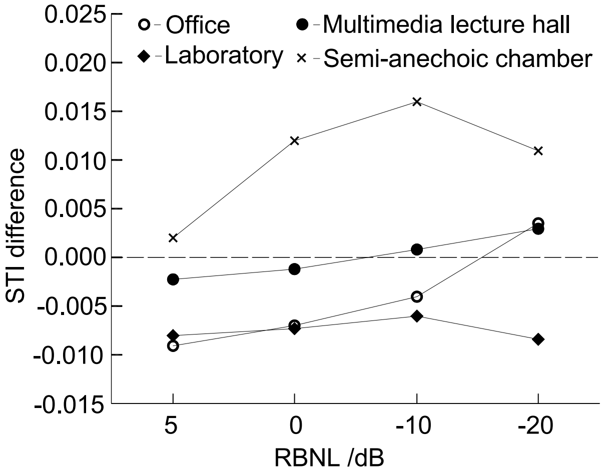


Fig. 5 The mean difference between the simulated and in-situ measured STIs under different RBNLs for each of the four rooms.

3.2 Comparison of speech intelligibility scores

The differences between the speech intelligibility scores of the simulation and in-situ measurement methods are provided in Figure 6. It is evidenced that most of the score differences are not large and there is no obvious tendency caused by the different RTs for the four rooms. The mean score difference of the 48 measurement conditions is approximately 1%, and the largest score difference, 8.6%, appears when the RBNL was 0 dB at the receiver position R_9. To evaluate these inaccuracies, the paired-sample T-test statistical method is used to confirm whether there exists a significant difference between the simulated and the measured scores. The p-value of 0.011 indicates that the bias of the simulated scores is significant and, therefore, cannot be ignored.

The mean score differences of the twelve receiver positions under different RBNLs are presented in Figure 7. Figure 7 reveals that with the decrease of the SNR, score bias may increase, a finding that is similar to Hodgson’s conclusion [4] that auralisation is not accurate in the case of high noise.

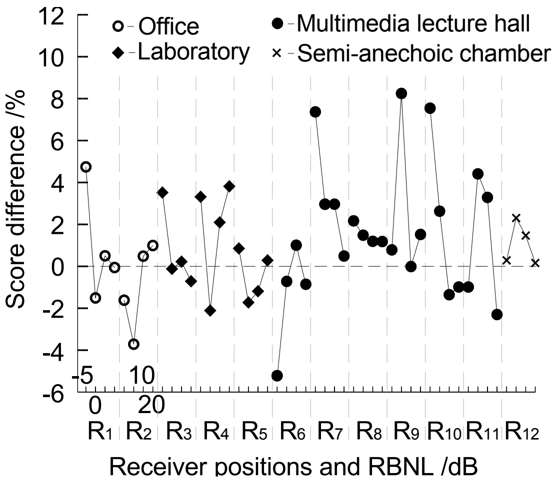


Fig. 6 The difference between simulated and in-situ measured speech intelligibility scores.

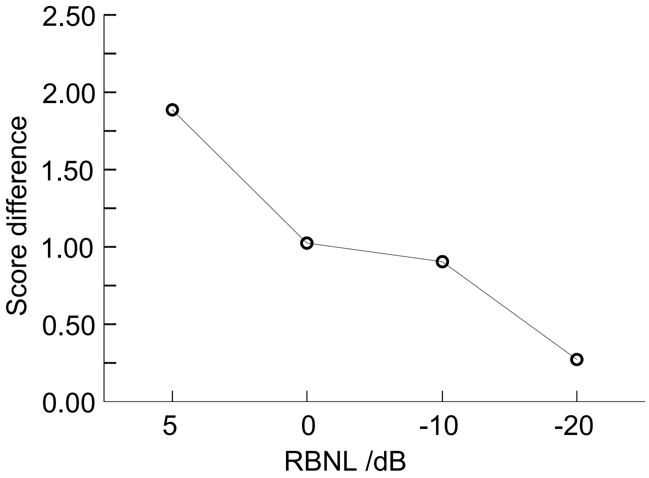


Fig. 7 The mean score differences of the twelve receiver positions under different RBNLs.

3.3 Comparison of simulated and in-situ measured curves

Data for the in-situ measurement method and the simulation method, based on 48 subjective and objective evaluations and their best-fitting third-order polynomial curves, are provided in Figure 8. It is evidenced from Figure 8 that the simulated curve is higher than the in-situ measured curve, especially in the case of high noise, although the score bias is quite small. The largest score difference appears whenthe STI is in the vicinity of 0.25, thus reaching approximately 3%. Considering the great majority of the standard deviations for the eight scores for each sound environment in Appendix C exceed 3%, the score bias may be difficult to detect in everyday situations. With respect to the in-situ measurement method, the curve covers speech intelligibility scores that range from 39.2% to 99.3% and STIs that range from 0.250 to 0.979. The coefficient of determination, R^2, of the curve is 0.9049, and the standard deviation between the speech intelligibility scores and the curve is 5.38%. With respect to the simulation method, the curve covers speech intelligibility scores ranging from 42.5% to 99.0% and STIs ranging from 0.250 to 0.990. The coefficient of determination, R^2, of the curve is 0.9048, and the standard deviation between the speech intelligibility scores and the curve is 5.17%. The relationship between speech intelligibility scores and STIs for the in-situ measurement method ($y_{\text{in-situ}}$) and the simulation method ($y_{\text{simu.}}$) are given in Eq. (1) and Eq. (2), respectively:

$$y_{\text{In-situ}} = 156.69x^3 - 449.53x^2 + 432.73x - 42.04 \tag{1}$$

$$y_{\text{Simu.}} = 119.33x^3 - 372.29x^2 + 379.37x - 29.072 \tag{2}$$

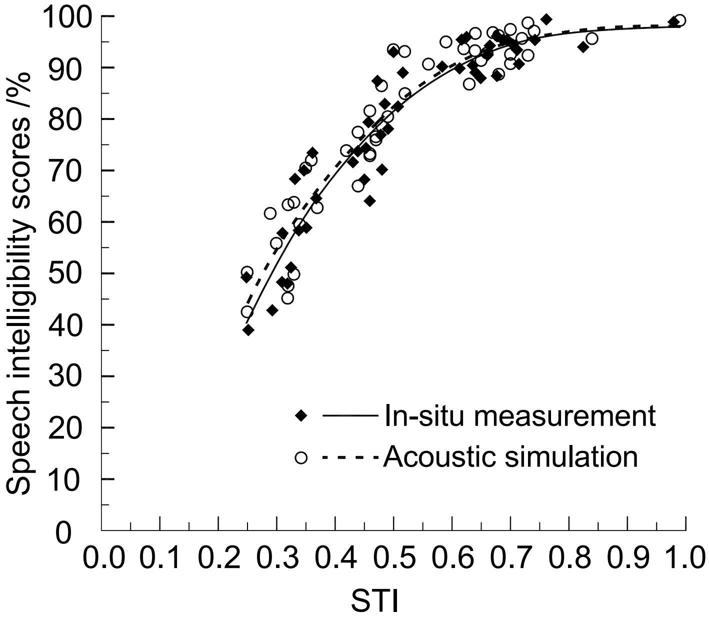


Fig. 8. Relationship between speech intelligibility scores and STI for the simulation and in-situ measurement methods.

3.4 Discussion

In the process of simulation, though the speech related acoustic parameters [27] such as EDT, RT, C_{80} and SPL have been carefully controlled, biases still exist between the simulated and in-situ measured speech intelligibility scores, and accordingly, further investigation is required to determine the cause.

The spectrum difference between the simulated and measured IRs and BRIRs following the normalisation of the octave-band levels to 60 dBA are presented in Tables 3 and 4. During the simulation of the BRIRs and IRs in ODEON, the HRTF of Subject_021 (M=0, without headphone equalisation) and the default filter (A(stop): 40 dB; A(pass): 0.1 dB; band overlap: 5%) were used.

It is evidenced from Table 3 that the spectrum difference between the simulated and measured IRs was less significant and that there was no obvious tendency. A finding that is similar to the SPL result in Appendix A. From Table 4, it is noted that the spectrum difference between the simulated and measured BRIRs was significant in the low frequency range and that, with the decrease of frequency from 500 Hz, the difference tended to increased. Specifically, the mean difference between the simulated and measured BRIRs of the twelve receiver positions increased by up to -20.62 dB when the frequency decreased to 125 Hz. This BRIR difference tendency is likely the reason that the simulated scores are slightly higher than the in-situ measured scores. Because the simulated speech material lists were adjusted to have the same speech levels as the corresponding in-situ measurement speech material lists, the energy in the high frequency range of simulated speech material lists should be higher than that of the in-situ measured speech material lists, a factor that should positively influence the understanding of language [28]. Though the spectrum differences between the simulated and the measured BRIR were considerable, the score bias was less significant in this experiment. This is perhaps because both the BRIRs from the virtual-44AA and those from the virtual-8020B had the same spectrum tendency, and thus, as the SNR was not influenced, it remained similar to that of the in-situ measurement. In addition, according to the spectrum difference between simulated and measured BRIRs seen in Table 4, the low and mid-frequency portions of the simulated speech material lists were lower. This can be deemed to reflect the sensitivity difference in the low and mid-frequency range between the binaural microphone used by ODEON and that used in the in-situ measurement. If both binaural microphones were calibrated for a certain frequency, for example 1000 Hz, the speech intelligibility scores using simulated BRIRs would certainly be lower than those using in-situ measured BRIRs due to the sound level decrease in the low and mid-frequency bands. Thus, the amplitudes of the simulated speech material lists were adjusted to have the same speech level as the corresponding in-situ measurement speech material lists in this paper. Moreover, if the background noise levels are not considered by ODEON or considered in the actual measurement, only the speech material lists are simulated by ODEON and adjusted to have the same speech level as that of the corresponding in-situ measurement. Accordingly, the score bias may increase.

To further investigate the reason for the simulated BRIR spectrum bias, the comparison results between the BRIR spectrum for Subject_021 with normal incidence and clockwise 80° from the front axis incidence direct sound measured by CIPIC Interface Laboratory (http://interface.cipic.ucdavis.edu/index.htm) and the BRIR for Subject_021 with normal incidence and clockwise 80° from the front axis incidence direct sound simulated by ODEON in a free field (using default filter, M=0, without headphone equalisation) are presented in Figures 9 and 10. Accordingly, it is noted that compared to the original BRIR measured by CIPIC Interface Laboratory (http://interface.cipic.ucdavis.edu/index.htm), the spectrum of simulated BRIR decreased significantly when the frequency was below 630 Hz (with normal incidence direct sound) and 650 Hz (with clockwise 80° from the front axis incidence direct sound). This decrease, however, may be the result of the HRTF filter used by ODEON.

Disregarding the influence of the HRTF filter used by ODEON, the different heads used in the BRIR simulation and BRIR in-situ measurement may still have contributed to the large spectrum difference seen in Table 4. In this paper the simulated BRIRs were obtained by combining the simulated IRs with the HRTF of Subject_021, and the in-situ BRIRs were measured at the receiver position using a binaural B&K 4101-A microphone placed on the author’s head. In addition, the HRTF filter–induced BRIR spectrum decrease also shows some differences by direction. Moreover, studies [29, 30, 31] also showed that the predictive accuracy of room acoustic parameters at low frequencies is lower than that at middle and high frequencies. However, these influences may not be so large, because the spectrum difference between simulated and measured BRIRs in Table 4 has exactly the same tendency as the HRTF filter–induced BRIR spectrum decrease in Figures 9 and 10.

Many factors could exert significant influences on the comparison results, and some measures have been adopted in this study to obtain correct and reliable results: a monitor loudspeaker was used as the interference noise source instead of the dodecahedral sound source, as the room acoustics software could hardly simulate a real dodecahedral sound source; real sound sources were calibrated in an anechoic chamber, and the data were used to calibrate the virtual sound sources in the model. The virtual interference noise source and the virtual signal source were erected directly using the horizontal and vertical directivity patterns of the monitor loudspeaker 8020B and the artificial mouth GRAS 44AA and were marked with ‘Natural’, and the ‘+EQ’ of each octave band was set to 0 dB in the ‘Point Source Editor’ menu, to ensure the accurate equivalence of virtual and real sound sources and avoid the frequency response of the sound source being included twice in the auralisation. Both of the sound source systems were equalised using their inverse filter systems calculated from the impulse responses measured on the front axis of the sources in an anechoic chamber. The noise signal for the measurement of background noise level was compiled and used in the auralisation, to ensure accurate equivalence of simulated and measured background noise levels; to reduce inaccuracy due to listening sequence and memory, subject speech intelligibility tests were carefully designed. The SPLs for the word lists were calibrated based on the removal of the silent parts of the speech signals, and the SPLs of all other test signals were accurately controlled at the beginning of signal generation by filtering, to ensure the complete equivalence of the SPL between test signals and speech signals.

Table 3. Spectrum difference (in dB) between simulated and measured IR at the R_1–R_{12} receiver positions in the four test rooms. The sources are the virtual-44AA and artificial mouth GRAS 44AA after the octave band levels have been normalised to an A-weighted level of 60 dB.

Room Type	Receiver position	Frequency bands (Hz)						
		125	250	500	1000	2000	4000	8000
Office	R_1	-0.45	-0.86	-0.30	-1.08	-0.43	-0.31	0.67
	R_2	-2.38	-2.38	-0.28	-1.84	-0.79	-0.53	1.22
Lab	R_3	-1.17	0.78	-0.40	-0.55	-0.74	-0.33	0.72
	R_4	0.15	-1.69	-0.73	-0.70	-1.55	-0.52	2.20
	R_5	0.32	-1.40	-1.45	-0.77	-1.44	-0.51	2.18
Multimedia lecture hall	R_6	2.28	-0.86	-0.97	0.81	0.69	-0.01	-0.27
	R_7	-0.46	2.34	-1.04	-0.29	-1.08	0.37	0.36
	R_8	-3.34	-3.04	-3.12	-2.42	-1.78	0.59	1.17
	R_9	-0.34	1.78	-3.06	-2.80	-0.11	-1.15	1.22
	R_{10}	-0.98	1.40	-2.97	-1.80	-0.76	0.40	0.54
	R_{11}	-3.16	0.54	-3.55	-1.75	-1.38	-0.10	1.15
Semi-anechoic chamber	R_{12}	5.42	-1.01	0.67	-0.29	0.21	0.12	-0.12
Mean difference of twelve receiver positions		-0.34	-0.37	-1.43	-1.12	-0.76	-0.16	0.92

Table 4. Spectrum difference (in dB) between simulated and measured BRIR at R_1-R12 receiver positions in the four test rooms. The sources are the virtual-44AA and artificial mouth GRAS44AA after the octave band levels have been normalised to an A-weighted level of 60 dB. The data in the Table are the average of the two ear canals.

Room Type	Receiver position	Frequency bands (Hz)						
		125	250	500	1000	2000	4000	8000
Office	R_1	-21.05	-7.41	-3.91	-1.76	-2.63	0.95	0.27
	R_2	-22.71	-10.62	-4.61	-3.97	-3.26	1.39	0.22
Lab	R_3	-19.42	-8.77	-2.49	-1.95	-2.05	1.59	-0.86
	R_4	-18.10	-11.19	-4.52	-2.55	-3.07	1.16	0.62
	R_5	-20.73	-11.43	-4.37	-2.40	-3.33	0.86	1.28
Multimedia lecture hall	R_6	-16.39	-7.92	-3.11	-0.84	-1.69	1.45	-1.26
	R_7	-21.68	-7.82	-5.08	-1.46	-3.08	1.69	-0.24
	R_8	-21.52	-10.76	-5.69	-3.28	-5.16	1.81	0.87
	R_9	-20.07	-8.25	-7.84	-2.68	-4.08	0.62	0.90
	R_{10}	-22.44	-9.26	-5.79	-2.55	-2.76	1.04	0.28
	R_{11}	-24.01	-10.96	-6.87	-3.92	-4.94	0.96	1.51
Semi-anechoic chamber	R_{12}	-19.28	-10.10	-2.14	-1.78	-1.25	1.90	-1.79
Mean difference of twelve receiver positions		-20.62	-9.54	-4.70	-2.43	-3.11	1.28	0.15

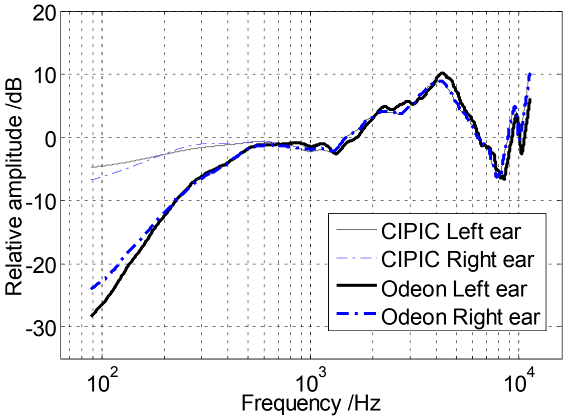


Fig. 9 BRIR spectrums for the Subject_021 with normal incidence direct sound measured by CIPIC Interface Laboratory and simulated by ODEON in a free field.

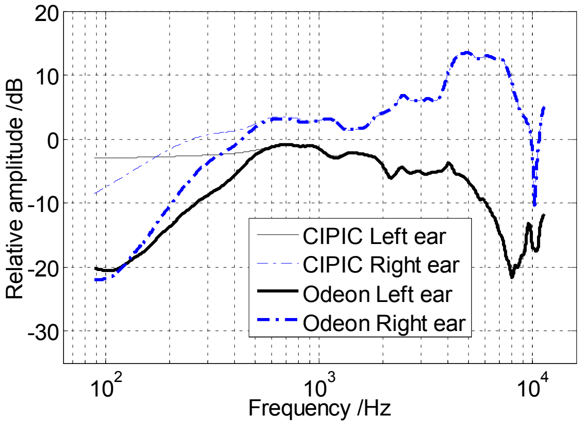


Fig. 10 BRIR spectrums for the Subject_021 with clockwise 80° from the front axis incidence direct sound measured by CIPIC Interface Laboratory and simulated by ODEON in a free field.

4. Conclusions

In this paper, the simulation technique for evaluating speech intelligibility under general room conditions was systematically compared. The results reveal that the STI can be predicted accurately by acoustic simulation when there is a good agreement between the virtual models and the real rooms and that different RT and SNR may exert less significant influence on the simulated STI. However, subjective intelligibility may be overestimated by acoustic simulation due to the HRTF filter used by ODEON, especially in the case of high noise. While this is consistent with the comparison results of the simulated and in-situ measured curves, the score bias may be minimal and difficult to detect in everyday situations. There is no obvious score tendency caused by different RT, though with a decrease in the SNR, the score bias may increase. Overall, considering that the accurate acoustic modelling of rooms is often difficult, it is difficult to obtain accurate speech intelligibility prediction results using a simulation technique, especially when the room has not yet been built.

Acknowledgments

This work was supported by National Natural Science Foundation of China, PR China (Grant No. 51278078, 11274243, 51378139 and 51308087) and the National Key Scientific Instrument and Equipment Development Project of China, PR China (Grant No. 2012YQ15021306).

References:

[1] French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. J Acoust Soc Am 1947;19:90–119.
[2] Anderson BW, Kalb JT. English verification of the STI method for estimating speech intelligibility of a communications channel. J Acoust Soc Am 1987;81:1982–5.
[3] Yang WY, Hodgson M. Validation of the auralization technique: comparative speech-intelligibility tests in real and virtual classrooms. Acta Acust United Acust 2007;93:991–9.
[4] Hodgson M, York N, Yang WY, et al. Comparison of predicted, measured and auralized sound fields with respect to speech intelligibility in classrooms using CATT-Acoustic and ODEON. Acta Acust United Acust 2008;94:883–90.
[5] Peng JX. Feasibility of subjective speech intelligibility assessment based on auralization. Appl Acoust 2005;66:591–601.
[6] Peng JX. Relationship between Chinese speech intelligibility and speech transmission index using diotic listening. Speech Commun 2007;49:933–6.
[7] Peng JX, Bei CX, Sun HT. Relationship between Chinese speech intelligibility and speech transmission index in rooms based on auralization. Speech Commun 2011;53:986–90.
[8] Peng JX. Relationship between Chinese speech intelligibility and speech transmission index in rooms using dichotic listening. Chin Sci Bull 2008;53:2748–52.
[9] Zhu PS, Mo FS, Kang J. Influence of sound source characteristics in determining objective speech intelligibility metrics. Appl Acoust 2015;89:188–98.
[10] IEC 60268-16,4rd ed. Sound system equipment-Part 16: Objective rating of speech intelligibility by speech transmission index. IEC; 2011.
[11] Christensen CL, Koutsouris G. Industrial Auditorium and Combined Editions User Manual. Lyngby (Denmark): ODEON A/S; 2013.
[12] Mapp P. Is STIPa a robust measure of speech intelligibility performance. AES 118th Convention, Barcelona, Spain, 2005.
[13] User Manual, V5.0, DIRAC Room Acoustics Software, Lyngby (Denmark): Brüel & Kjær Sound & Vibration Measurement A/S; 2010.
[14] Farina A, User Manual of Aurora43, Parma (Italy): University of Parma A/S; 2012.
[15] Bowden EE, Wang LM. Verifying two commercial software implementations of impulse-response-based speech intelligibility measurements. Appl Acoust 2007;68:717–28.
[16] Zhu PS, Mo FS, 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.
[17] GB/T 13504-2008. Diagnostic Rhyme Test (DRT) method of Chinese articulation. Standard of PR China, 2008.
[18] SJ2467-84. Test methods for Chinese articulation in communication equipment. Standard of PR China, 1984.
[19] GB/T 15508-1995. Acoustics-Speech articulation testing method. Standard of PR China, 1995.
[20] Meng ZH, Dai L. Test study on relationship between clarity of mandarin monosyllable and STIPA caused by reverberation. Audio Eng 2009;2:4–8 [in Chinese].
[21] Zhang JL. Chinese human-machine communication system, first ed., Shanghai scientific & Technical Publishers, Shanghai, 2002, p. 500–74. [in Chinese].
[22] Rao D, Wu SX. Auralization differences between individualized and nonindividualized binaural room impulse responses. J South China Univ Technol (Natural Science Edition) 2008;8:123–7 [in Chinese].
[23] Rao D. Analysis of non-individualized compensation for headphone transfer function. Audio Eng 2007;3:41–3 [in Chinese].
[24] Henrik M. Fundamentals of binaural technology. Appl Acoust 1992;36:171–218.
[25] Steeneken HJM, Houtgast T. Comparison of some methods for measuring speech levels. Report IZF 1986-20. TNO-HFRI Human Factors Research Institute, Soesterberg, The Netherlands, 1986.
[26] Bradley JS, Reich R, Norcross SG. A just noticeable difference in C50 for speech. Appl Acoust 1999;58:99–108.
[27] Tang SK. Speech related acoustical parameters in classrooms and their relationships. Appl Acoust

2008;12:1318–31.
[28] Shen GL. Discussion on speech sound recording. Motion Picture Video Technol 1997;5:3–11 [in Chinese].
[29] Howarth MJ, Lam YW. An assessment of the accuracy of a hybrid room acoustics model with surface diffusion facility. Appl Acoust 2000;60:237–51.
[30] Zeng XY, Chen KA, Sun JC. Development of a hybrid computer model for simulating the complicated virtual sound field in enclosures. Appl Acoust 2002;63:481–91.
[31] Zeng XY. An improved broad-spectrum room acoustics model including diffuse reflections. Appl Acoust 2005;66:1309–19.

Appendix A Simulated and measured acoustic parameters at the R1-R12 receiver positions in the four test rooms with the source being the virtual-44AA and the artificial mouth GRAS 44AA.

Room Type	Capacity (m^3)	Receiver position	Acoustic parameters		Frequency bands (Hz)						
					125	250	500	1000	2000	4000	8000
Office	108	R_1	T_{30} (s)	simulated	0.60	0.67	0.60	0.66	0.71	0.65	0.54
				measured	0.55	0.70	0.63	0.65	0.68	0.68	0.57
			EDT (s)	simulated	0.61	0.68	0.56	0.62	0.68	0.64	0.52
				measured	0.72	0.68	0.55	0.56	0.58	0.65	0.51
			C_{80} (dB)	simulated	7.50	6.40	8.50	7.50	7.00	7.70	10.20
				measured	6.29	6.33	8.94	8.64	8.35	8.37	12.23
			SPL (dB)	simulated	59.30	65.70	59.00	54.40	49.30	45.10	39.30
				measured	57.60	65.79	61.31	55.78	51.02	46.03	39.85
		R_2	T_{30} (s)	simulated	0.59	0.66	0.61	0.65	0.69	0.70	0.55
				measured	0.64	0.69	0.60	0.65	0.69	0.69	0.58
			EDT (s)	simulated	0.61	0.67	0.59	0.62	0.66	0.69	0.51
				measured	0.62	0.69	0.70	0.67	0.70	0.67	0.55
			C_{80} (dB)	simulated	7.40	6.70	7.90	7.50	7.30	7.10	10.10
				measured	8.06	7.11	6.32	7.17	6.57	6.74	9.83
			SPL (dB)	simulated	58.60	65.20	58.70	54.00	48.80	45.30	38.90
				measured	55.19	64.58	58.23	54.15	49.50	44.08	37.57
Laboratory	238	R_3	T_{30} (s)	simulated	1.45	1.57	1.52	1.61	1.56	1.41	0.98
				measured	1.47	1.52	1.64	1.64	1.57	1.45	1.01
			EDT (s)	simulated	1.47	1.61	1.55	1.63	1.54	1.37	0.84
				measured	1.11	1.66	1.59	1.65	1.67	1.45	0.94
			C_{80} (dB)	simulated	2.60	1.80	2.40	2.30	3.90	4.90	9.10
				measured	2.92	3.33	2.10	3.82	3.88	5.70	10.60
			SPL (dB)	simulated	61.20	67.50	61.60	56.60	50.60	46.30	40.70
				measured	60.76	67.83	59.76	56.83	50.88	46.65	40.49
		R_4	T_{30} (s)	simulated	1.36	1.53	1.60	1.67	1.53	1.43	1.00
				measured	1.50	1.46	1.54	1.63	1.59	1.46	1.04
			EDT (s)	simulated	1.24	1.56	1.63	1.69	1.53	1.42	0.96
				measured	1.31	1.36	1.61	1.69	1.47	1.43	0.99
			C_{80} (dB)	simulated	1.40	-0.10	-0.20	-0.30	1.00	1.80	5.00
				measured	-0.41	1.24	0.21	0.22	0.74	1.83	5.15
			SPL (dB)	simulated	57.10	64.50	58.90	54.10	47.50	43.40	36.60
				measured	54.25	64.86	59.54	53.38	48.76	43.06	34.92
		R_5	T_{30} (s)	simulated	1.80	1.49	1.56	1.64	1.50	1.43	0.99
				measured	1.88	1.45	1.56	1.62	1.59	1.43	1.03
			EDT (s)	simulated	1.81	1.50	1.57	1.65	1.51	1.43	0.96

				measured	1.69	1.52	1.60	1.74	1.58	1.43	0.94
			C_{80} (dB)	simulated	-1.20	-0.10	-0.30	-0.50	0.60	1.30	4.40
				measured	0.70	-0.12	1.47	-1.37	0.60	1.20	4.64
			SPL (dB)	simulated	57.50	63.80	58.30	53.60	46.90	43.30	36.10
				measured	57.54	63.47	58.63	53.10	48.83	42.98	34.33
Multimedia lecture hall	1674	R_6	T_{30} (s)	simulated	1.06	0.91	0.85	0.86	0.84	0.72	0.56
				measured	1.12	0.81	0.79	0.66	0.69	0.62	0.53
			EDT (s)	simulated	1.00	0.72	0.52	0.50	0.52	0.51	0.31
				measured	0.95	0.75	0.74	0.64	0.54	0.50	0.26
			C_{80} (dB)	simulated	4.50	7.00	9.30	9.70	10.30	11.20	14.80
				measured	5.28	8.21	8.75	10.33	11.68	14.78	18.91
			SPL (dB)	simulated	51.40	56.60	49.50	44.00	39.50	35.80	30.60
				measured	49.02	57.32	49.31	43.24	39.09	35.98	31.61
		R_7	T_{30} (s)	simulated	0.95	0.94	0.80	0.80	0.81	0.73	0.58
				measured	0.94	0.90	0.75	0.75	0.81	0.72	0.56
			EDT (s)	simulated	0.83	0.72	0.50	0.47	0.56	0.52	0.43
				measured	0.63	1.02	0.71	0.57	0.49	0.49	0.58
			C_{80} (dB)	simulated	5.30	6.40	9.40	10.10	8.80	9.30	11.90
				measured	3.08	3.32	6.86	9.56	10.44	10.74	15.25
			SPL (dB)	simulated	49.70	55.00	47.30	41.90	37.00	33.20	27.40
				measured	50.90	52.53	46.40	41.10	36.22	30.57	25.40
		R_8	T_{30} (s)	simulated	0.96	0.90	0.88	0.84	0.82	0.81	0.68
				measured	0.89	0.81	0.79	0.78	0.81	0.86	0.63
			EDT (s)	simulated	0.75	0.55	0.47	0.44	0.49	0.59	0.40
				measured	0.85	0.81	0.78	0.52	0.56	0.54	0.40
			C_{80} (dB)	simulated	6.40	8.60	9.50	10.00	9.20	8.40	10.30
				measured	-0.19	4.53	5.22	9.18	9.55	9.93	13.42
			SPL (dB)	simulated	49.00	53.10	46.50	41.00	35.70	32.50	27.40
				measured	45.72	52.36	44.28	39.24	34.36	27.46	22.87
		R_9	T_{30} (s)	simulated	1.05	0.87	0.82	0.71	0.79	0.70	0.55
				measured	1.07	0.85	0.74	0.75	0.72	0.66	0.51
			EDT (s)	simulated	1.06	0.75	0.54	0.46	0.48	0.42	0.23
				measured	0.81	0.68	0.52	0.44	0.51	0.37	0.32
			C_{80} (dB)	simulated	3.50	6.40	9.20	10.60	10.30	11.80	15.10
				measured	3.12	6.71	9.57	10.15	9.93	13.62	17.17
			SPL (dB)	simulated	49.90	54.90	49.50	41.90	37.80	33.90	28.60
				measured	48.52	51.58	50.43	43.44	36.18	32.57	27.18
		R_{10}	T_{30} (s)	simulated	0.91	0.85	0.78	0.81	0.77	0.71	0.56
				measured	0.86	0.84	0.76	0.73	0.71	0.67	0.54

			EDT (s)	simulated	0.91	0.77	0.63	0.58	0.55	0.48	0.32
				measured	0.97	0.73	0.71	0.48	0.49	0.41	0.33
			C_{80} (dB)	simulated	4.90	6.30	8.10	9.00	9.40	10.10	13.50
				measured	3.21	5.71	6.16	10.75	10.37	11.88	16.06
			SPL (dB)	simulated	48.70	53.50	45.70	39.70	35.20	31.90	26.50
				measured	48.11	52.14	45.95	40.47	35.16	30.23	25.19
		R$_{11}$	T_{30} (s)	simulated	1.02	0.80	0.76	0.75	0.76	0.69	0.61
				measured	1.04	0.78	0.75	0.70	0.72	0.67	0.57
			EDT (s)	simulated	0.92	0.56	0.50	0.47	0.45	0.42	0.31
				measured	0.89	0.71	0.60	0.53	0.55	0.42	0.27
			C_{80} (dB)	simulated	8.60	8.70	9.70	10.40	10.40	11.00	13.20
				measured	4.00	4.96	6.55	10.20	10.12	11.57	15.50
			SPL (dB)	simulated	47.50	53.10	46.50	40.90	36.00	32.30	27.50
				measured	47.70	51.84	44.91	39.77	34.10	28.77	23.43
Semi-anechoic chamber	550	R$_{12}$	T_{30} (s)	simulated	0.20	0.14	0.10	0.09	0.08	0.08	0.06
				measured	0.16	0.12	0.07	—	—	—	—
			EDT (s)	simulated	0.03	0.04	0.04	0.04	0.03	0.03	0.02
				measured	0.18	0.04	0.04	0.03	0.04	0.06	0.08
			C_{80} (dB)	simulated	30.30	55.60	53.10	46.30	43.90	42.70	43.90
				measured	24.97	45.13	50.69	54.17	41.08	40.70	37.24
			SPL (dB)	simulated	50.00	55.80	50.10	45.30	40.90	37.50	32.70
				measured	46.21	56.83	47.18	46.29	41.49	37.32	32.91

Appendix B The speech material lists and the listening sequence for the R1-R6 receiver positions of the two methods, listened by one group of eight juries. In-situ. is the in-situ measurement method; Simu. is the simulation method; Pair_1 is test pair 1, (a) is one of the two juries in the test pair.

Receiver position	Word list	Method	RBNL (dB)				Listening sequence
			Pair_1 (a)	Pair_2 (a)	Pair_3 (a)	Pair_4 (a)	
R_1	1	In-situ.	5	0	-10	-20	1
R_2	2	In-situ.	0	-10	-20	5	2
R_3	9	Simu.	-10	-20	5	0	3
R_4	4	In-situ.	-20	5	0	-10	4
R_2	8	Simu.	0	-10	-20	5	5
R_6	6	In-situ.	5	-20	-10	0	6
R_5	11	Simu.	-20	-10	0	5	7
R_1	7	Simu.	5	0	-10	-20	8
R_3	3	In-situ.	-10	-20	5	0	9
R_4	10	Simu.	-20	5	0	-10	10
R_6	12	Simu.	5	-20	-10	0	11
R_5	5	In-situ.	-20	-10	0	5	12

Receiver position	Word list	Method	RBNL (dB)				Listening sequence
			Pair_1 (b)	Pair_2 (b)	Pair_3 (b)	Pair_4 (b)	
R_1	7	In-situ.	5	0	-10	-20	1
R_2	8	In-situ.	0	-10	-20	5	2
R_3	3	Simu.	-10	-20	5	0	3
R_4	10	In-situ.	-20	5	0	-10	4
R_2	2	Simu.	0	-10	-20	5	5
R_6	12	In-situ.	5	-20	-10	0	6
R_5	5	Simu.	-20	-10	0	5	7
R_1	1	Simu.	5	0	-10	-20	8
R_3	9	In-situ.	-10	-20	5	0	9
R_4	4	Simu.	-20	5	0	-10	10
R_6	6	Simu.	5	-20	-10	0	11
R_5	11	In-situ.	-20	-10	0	5	12

Appendix C The STI values and speech intelligibility scores for the two methods, and their differences, with 48 sound environments divided among the four rooms (standard deviations for the eight scores of each sound environment in brackets).

Room Type	Receiver position	RBNL (dB)	In-situ measured		Simulated		Difference between simulated and measured	
			STI	Score (%)	STI	Score (%)	STI	Score (%)
Office	R_1	5	0.352	59.0(6.0)	0.33	63.8(6.7)	-0.022	4.8
		0	0.486	83.0(3.3)	0.46	81.5(4.2)	-0.026	-1.5
		-10	0.662	92.5(4.3)	0.64	93.0(2.7)	-0.022	0.5
		-20	0.715	90.8(3.5)	0.70	90.7(3.6)	-0.015	0.0
	R_2	5	0.326	51.3(6.2)	0.33	49.7(6.1)	0.004	-1.6
		0	0.458	79.5(2.6)	0.47	75.8(3.7)	0.012	-3.7
		-10	0.626	96.0(3.3)	0.64	96.5(2.7)	0.014	0.5
		-20	0.678	96.2(1.9)	0.70	97.2(1.9)	0.022	1.0
Laboratory	R_3	5	0.311	58.0(6.4)	0.29	61.5(5.7)	-0.021	3.5
		0	0.440	73.8(4.0)	0.42	73.7(2.5)	-0.020	-0.1
		-10	0.584	90.3(2.4)	0.56	90.5(2.4)	-0.024	0.2
		-20	0.617	95.5(2.1)	0.59	94.8(1.3)	-0.027	-0.7
	R_4	5	0.253	39.2(5.9)	0.25	42.5(5.9)	-0.003	3.3
		0	0.369	64.8(5.0)	0.37	62.7(4.2)	0.001	-2.1
		-10	0.491	78.2(4.2)	0.49	80.3(3.8)	-0.001	2.1
		-20	0.517	89.2(3.5)	0.52	93.0(1.6)	0.003	3.8
	R_5	5	0.250	49.3(5.4)	0.25	50.2(5.0)	0.000	0.9
		0	0.363	73.5(4.2)	0.36	71.8(3.9)	-0.003	-1.7
		-10	0.473	87.5(4.4)	0.48	86.3(3.8)	0.007	-1.2
		-20	0.501	93.0(2.6)	0.50	93.3(2.0)	-0.001	0.3
Multimedia lecture hall	R_6	5	0.333	68.5(4.9)	0.32	63.2(5.2)	-0.013	-5.3
		0	0.479	77.2(4.1)	0.47	76.5(2.4)	-0.009	-0.7
		-10	0.687	95.5(3.3)	0.67	96.5(2.8)	-0.017	1.0
		-20	0.762	99.3(1.0)	0.73	98.5(1.7)	-0.032	-0.8
	R_7	5	0.316	48.5(6.7)	0.30	55.8(5.4)	-0.016	7.3
		0	0.454	74.5(4.8)	0.44	77.5(4.8)	-0.014	3.0
		-10	0.637	90.5(3.6)	0.62	93.5(3.3)	-0.017	3.0
		-20	0.695	95.5(2.8)	0.68	96.0(2.4)	-0.015	0.5
	R_8	5	0.294	43.0(6.2)	0.32	45.2(5.2)	0.026	2.2
		0	0.432	71.7(4.3)	0.46	73.2(3.3)	0.028	1.5
		-10	0.613	90.0(3.1)	0.65	91.2(3.1)	0.037	1.2
		-20	0.666	94.3(2.7)	0.72	95.5(2.2)	0.054	1.2
	R_9	5	0.339	58.5(5.9)	0.34	59.3(5.0)	0.001	0.8
		0	0.481	70.3(4.7)	0.48	78.5(4.1)	-0.001	8.2

		-10	0.677	88.5(3.5)	0.68	88.5(3.9)	0.003	0.0
		-20	0.742	95.5(2.9)	0.74	97.0(2.5)	-0.002	1.5
	R_{10}	5	0.320	48.2(5.2)	0.30	55.7(4.9)	-0.020	7.5
		0	0.460	64.2(5.5)	0.44	66.8(4.1)	-0.020	2.6
		-10	0.650	88.0(3.5)	0.63	86.7(3.5)	-0.020	-1.3
		-20	0.712	93.5(2.7)	0.70	92.5(2.4)	-0.012	-1.0
	R_{11}	5	0.311	48.5(5.7)	0.32	47.5(4.5)	0.009	-1.0
		0	0.451	68.3(4.6)	0.46	72.7(5.4)	0.009	4.4
		-10	0.641	89.2(3.1)	0.66	92.5(3.8)	0.019	3.3
		-20	0.705	94.5(2.5)	0.73	92.2(2.8)	0.025	-2.3
Semi-anechoic chamber	R_{12}	5	0.348	70.0(5.4)	0.35	70.3(5.7)	0.002	0.3
		0	0.508	82.5(3.4)	0.52	84.8(4.0)	0.012	2.3
		-10	0.824	94.0(3.7)	0.84	95.5(3.0)	0.016	1.5
		-20	0.979	98.8(1.5)	0.99	99.0(1.6)	0.011	0.2