

Research Paper

Study on Chinese Speech Intelligibility Under Different Low-Frequency Characteristics of Reverberation Time Using a Hybrid Method

Wuqiong HUANG^{(1),(2)*}, Jianxin PENG^{(1)*}, Tinghui XIE⁽³⁾

⁽¹⁾ *School of Physics and Optoelectronics, South China University of Technology
Guangzhou, China*

⁽²⁾ *State Key Laboratory of Subtropical Building Science, South China University of Technology
Guangzhou, China*

⁽³⁾ *School of Architecture and Art, Shijiazhuang Tiedao University
Shijiazhuang, China*

*Corresponding Author e-mail: hwq2020@scut.edu.cn (Wuqiong Huang); phjxpeng@163.com (Jianxin Peng)

(received March 15, 2022; accepted November 19, 2022)

Reverberation time (RT) is an important indicator of room acoustics, however, most studies focus on the mid-high frequency RT, and less on the low-frequency RT. In this paper, a hybrid approach based on geometric and wave methods was proposed to build a more accurate and wide frequency-band room acoustic impulse response. This hybrid method utilized the finite-difference time-domain (FDTD) method modeling at low frequencies and the Odeon simulation at mid-high frequencies, which was investigated in a university classroom. The influence of the low-frequency RT on speech intelligibility was explored. For the low-frequency part, different impedance boundary conditions were employed and the effectiveness of the hybrid method has also been verified. From the results of objective acoustical parameters and subjective listening experiments, the smaller the low-frequency RT was, the higher the Chinese speech intelligibility score was. The syllables, consonants, vowels, and the syllable order also had significant effects on the intelligibility score.

Keywords: low frequency; speech intelligibility; classroom; finite-difference time-domain method.



Copyright © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/>) which permits use, distribution, and reproduction in any medium, provided that the article is properly cited. In any case of remix, adapt, or build upon the material, the modified material must be licensed under identical terms.

1. Introduction

Concerning the reverberation time (RT) in a room, most studies only pay attention to the mid-high frequency RT, less care about the low-frequency part of RT in a room. RT is the primary index of the acoustical design for all kinds of halls, but the requirement of RT in the low-frequency range is still controversial. Around the 1960s, Beranek defined the RT ratio of low frequency (125–250 Hz) and medium frequency (500–1000 Hz) as the bass ratio (BR), and put forward to the ideal value of 1.1–1.5 for BR (BERANEK, 1962), namely the bass-rise reverberation characteristic. This bass-rise characteristic has been seen as desirable or at least tolerated in auditoria, especially in the USA (BARRON, 2010). However, a flatten RT curve has been more favorable in Europe recently, even BERANEK (2010)

questioned himself after numerous and elaborate investigations. After measuring many performing venues with good sound quality, FUCHS and STEINKE (2015) found that these buildings had a relatively flat frequency curve of RTs, hence they suggested that the BR close to 1 was more favorable to the low-frequency sound. In addition, ADELMAN-LARSEN (2015) emphasized the necessity to improve clarity by controlling the low-frequency RT in his analysis of large-scale venues.

The low-frequency sound in an auditorium can increase the feeling of warmth in the hall (BERANEK 1996), however intelligibility is more important than warmth in a speech hall, such as a classroom. In China, most of the Chinese acoustic standards are still based on the characteristic of the bass-rise RT (GB/T, 2005; GB, 2010; JGJ/T, 2012). Moreover, there are few sound standards for frequency below 500 Hz. As we

know, Chinese is a tonal language, which is different from the intonation language of western countries. For Chinese, vowels are longer than consonants, as the low-frequency vowels are easy to mask the mid-high frequency consonants, which ultimately affect the Chinese speech intelligibility in a room. Some studies have demonstrated acoustic problems by applying low-frequency sound-absorbing structures and sound-absorbing materials into actual buildings. By adding low-frequency sound absorption structures in the classroom, ZHA and LYU (2020) reduced the low-frequency noise and flattened the low-frequency RT characteristics, in which way they obtained the satisfactory sound environment for a classroom. PENG *et al.* (2020) compared the objective parameters and subjective speech perception in two classrooms with similar RTs, but one classroom had a better speech perception than the other. They suggested that this discrepancy might be due to the difference in a low-frequency RT or a background noise level. In the follow-up study, XU *et al.* (2021) used the Odeon software to calculate RTs of these two classrooms and then carried out the Chinese speech intelligibility listening test by a headphone, which eliminated the influence of background noise and confirmed that reducing the low-frequency RTs was helpful to improve the Chinese speech intelligibility in the classroom.

The room acoustic simulation is an important part of the architectural design process, which is convenient and cost-saving. However, popular methods of predicting room acoustic characteristics are based on a geometric acoustic model of ray-like sound propagation, considering that these kinds of software are suitable for small wavelengths in comparison to the dimensions of the enclosure and internal objects. At larger wavelengths, the ray-like assumption no longer holds and the phenomena such as diffraction caused by a low-frequency acoustic wave cannot be ignored (SOUTHERN *et al.*, 2013). Solving the low-frequency sound problem the method based on the acoustic wave theory should be used. Moreover, the finite-difference time-domain (FDTD) method can simulate the frequency-dependent boundaries with desired sound absorption characteristics through the digital impedance filter (KOWALCZYK, VAN WALSTIJN, 2008). As the FDTD method has become more mature, this method is showing its superiority in both simulation accuracy and calculation speed (BOTTELDOOREN, 1995; OXNARD, 2018). Even so, the FDTD method is very memory-intensive, especially when modeling large volumes or a wide frequency bandwidth such as the range of hearing. Thus, this work used a combined method of FDTD and a geometric method to get a synthetic wide frequency-band room impulse response (RIR), which includes the 63 Hz octave frequency band.

This study aims to investigate the effect of different RTs in low frequencies on speech intelligibility. It is

better to control RTs in the mid-high frequencies by modeling methods and change only the RTs in low frequencies. Through the combined method four kinds of low-frequency reverberation characteristics were established, and the differences in Chinese speech intelligibility in a classroom before and after the improvement were compared.

2. Experimental method

2.1. Room

In universities, most classrooms are large with a rectangular shape. In the study, a large classroom with volume of $15.82 \times 8.22 \times 4.9 \text{ m}^3$ was selected in the university, which has two windows and two metal doors on the left side, four windows on the right side and a blackboard on the front wall. All the walls and the ceiling are of plastered brick, the floor is covered with ceramic tiles, and the seats and desks are composed of multi-plywood. The RIRs of six receiving positions in the classroom were recorded by using the B&K 4189 microphone, and using the B&K 4296 dodecahedral loudspeaker as an omnidirectional sound source. The sound source (S) with a height of 1.4 m from the ground and all the receiving positions (R1–R6) with the same height of 1.2 m were arranged as shown in Fig. 1. During the measurement, the doors and windows were closed, and the RIRs were measured by using the swept-frequency method in the unoccupied classroom. Whereafter, RIRs were calculated by Dirac4.1 software, and objective acoustic parameters such as RT were obtained. After the calculation, the measured average RT of the six receiving positions was 3 s at 63 Hz, rising to 3.7 s at 125 Hz, and then decreasing gradually. In the Odeon model, the sound absorption coefficients were basically set according to the material of each surface in the actual classroom, while slightly adjusted to make the average RT of each receiving position close to the measured average RT. The sound absorption coefficients of all surfaces in the Odeon model are listed in Table 1.

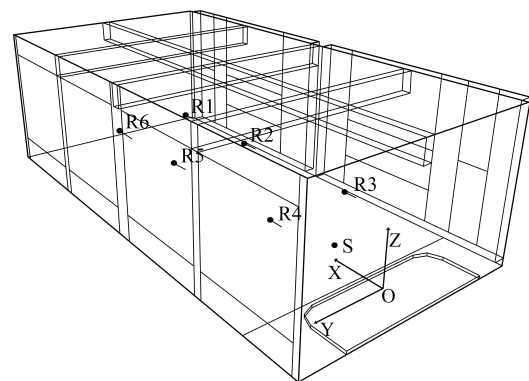


Fig. 1. Location of source and receiving points in the Odeon model.

Table 1. Sound absorption coefficients of the classrooms in the Odeon models.

Frequency [Hz]	500	1000	2000	4000
Floor	0.03	0.04	0.04	0.05
Glass windows	0.03	0.03	0.02	0.02
Seats	0.32	0.29	0.35	0.32
Front wall	0.03	0.04	0.04	0.07
Door	0.01	0.01	0.02	0.02
Ceiling	0.02	0.02	0.02	0.02
The walls of Model A&B	0.02	0.02	0.02	0.03
The walls of Model C&D	0.56	0.56	0.56	0.56

2.2. FDTD acoustic model

The formulation of the FDTD scheme used in this study is the numerical solution of the wave equation that governs sound wave propagation in an ideal isotropic medium:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p, \quad (1)$$

where p denotes sound pressure and c denotes the sound speed which is set to 344 m/s, t is time in seconds and $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the 3D Laplacian operator. FDTD schemes for numerical simulation of the wave equation are derived by approximating time and space derivatives with finite difference operators according to KOWALCZYK and VAN WALSTIJN (2008). Assuming an equal distance between grid points in all directions, the 3D discretized wave equations take the form of Eq. (2):

$$\begin{aligned} p_{l,m,i}^{n+1} = & \kappa^2 (p_{l+1,m,i}^n + p_{l-1,m,i}^n + p_{l,m+1,i}^n \\ & + p_{l,m-1,i}^n + p_{l,m,i+1}^n + p_{l,m,i-1}^n) \\ & + 2(1 - 3\kappa^2) p_{l,m,i}^n - p_{l,m,i}^{n-1}, \end{aligned} \quad (2)$$

where $\kappa = cT/X$ denotes the Courant number, T is the time step, X is the grid spacing, l , m , and i denote the spatial indexes in x , y , and z directions, and n is the time index. To ensure numerical stability in simulations, the stability condition should be satisfied, that is $\kappa \leq \frac{1}{\sqrt{3}}$. Besides, the grid spacing should not be too long, generally less than one tenth of the wavelength. Combined with the stability condition, in this study the grid spacing was set to 0.06 m and the time step was set to 100.7 μ s, and the derivative of a Gaussian function was chosen as the excitation source. Under these conditions, the spectral characteristic of the excitation source is non-flat in the whole frequency band, which will distort the listening material. To eliminate the non-flat effects of the excitation source, the RIRs calculated by the FDTD method were corrected by an inverse-filtering technique (SAKAMOTO *et al.*, 2008).

In general, the reflected wave has a phase and an amplitude that differ from those of the incident wave, and such changes diverge with frequency. Assuming that a digital filter also has a frequency response in amplitude and phase, so the frequency-dependent boundary can be incorporated in a FDTD model with a digital filter. Since the infinite impulse response filter (IIR filter) and the specific impedance of the boundary have a similar form, this study expressed the boundary in terms of IIR filter. KOWALCZYK and VAN WALSTIJN (2008) presented the FDTD formulation of the digital impedance filter (DIF) in a rectilinear grid, and the update formula for a boundary node could be expressed as:

$$\begin{aligned} p_{l,m,i}^{n+1} = & \left[2\kappa^2 (p_{l-1,m,i}^n + p_{l,m-1,i}^n + p_{l,m,i-1}^n) \right. \\ & + \kappa^2 \left(\frac{g_x^n}{b_x} + \frac{g_y^n}{b_y} + \frac{g_z^n}{b_z} \right) + 2(1 - 3\kappa^2) p_{l,m,i}^n \\ & + \left. \left(\frac{\kappa a_x}{b_x} + \frac{\kappa a_y}{b_y} + \frac{\kappa a_z}{b_z} - 1 \right) p_{l,m,i}^{n-1} \right] \\ & \times \left(1 + \frac{\kappa a_x}{b_x} + \frac{\kappa a_y}{b_y} + \frac{\kappa a_z}{b_z} \right)^{-1}, \end{aligned} \quad (3)$$

where a , b , and g have been clarified in (KOWALCZYK, VAN WALSTIJN, 2008) and are not explained here. In this FDTD model, the DIFs combined Bessel high-pass filters and Bessel low-pass filters were used to obtain the sound absorption coefficient of each interface. From the parameters a and b of the filter, the corresponding specific acoustic impedance (ξ) can be obtained. According to the relationship among the specific acoustic impedance, the reflection coefficient and the absorption coefficient (α), the absorption coefficient can be expressed as:

$$\alpha = \frac{4 \operatorname{Re}(\xi)}{|\xi|^2 + 2 \operatorname{Re}(\xi) + 1}. \quad (4)$$

2.3. Verification

The RIRs obtained by the FDTD method were processed by low-pass filter with upper limit frequency of 355 Hz (355 Hz is corresponding to the upper cut-off frequency of 250 Hz octave band), while the RIRs obtained by Odeon software were processed by the high-pass filter with the lower cut-off frequency of 355 Hz. Then, these two kinds of simulation results were combined to obtain synthetic RIRs over the entire audible spectrum.

To verify the accuracy of the DIF boundary, a classroom model named Model A, was established, whose absorption coefficients of all boundaries were adjusted so that the average RTs of the six receiving positions were close to the measurements of the classroom.

Figure 2 exhibits that the simulated average RTs of six receiving positions are basically consistent with the measured results, which is within plus and minus 5% of the measurement. The comparison result has verified the effectiveness of this combined simulation method.

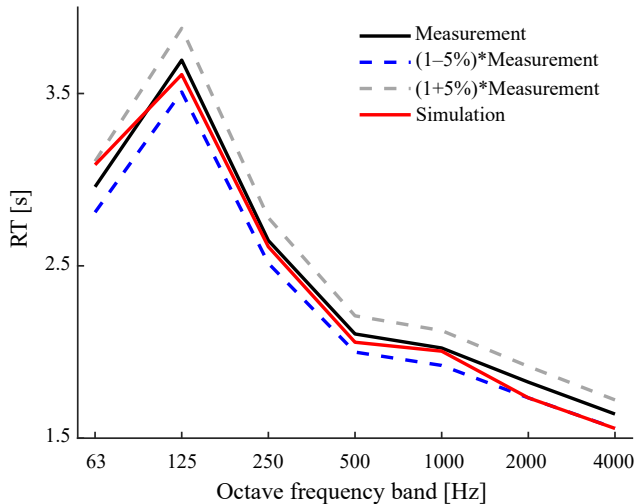


Fig. 2. Comparison of RTs between the measurement and simulation.

Based on Model A, Model B with relatively flat RT at low frequencies was established. According to the specification (GB, 2010), the RT in the classroom larger than 300 m^3 should be lower than 0.8 s. Therefore, both of Model C and Model D with an average RT of 0.8 s at mid-high frequency were established. In addition, Model C and Model D have a rising and a flat RT at low frequencies, respectively. The RTs of these four models are shown in Fig. 3 and the sound absorption coefficients of these FDTD models are listed in Table 2. The absorption coefficients of FDTD models were obtained through Eq. (4). Most of the sound absorption processing were for plastered brick walls and

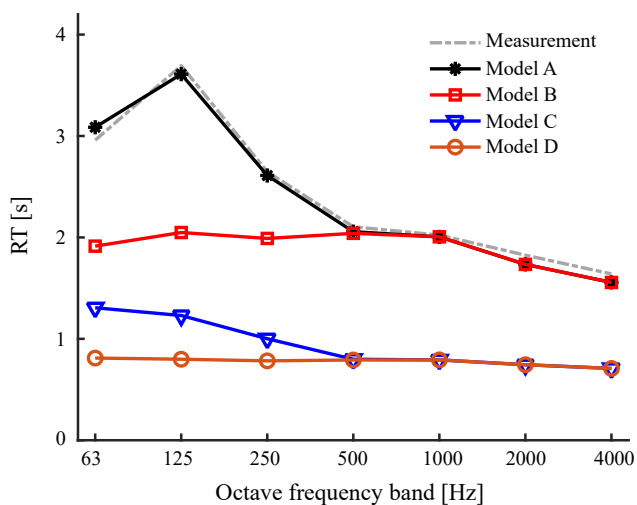


Fig. 3. RTs of measurement and four models.

Table 2. Sound absorption coefficients of the classrooms in the FDTD models. The rests models stand for models with the same absorption coefficient.

Boundary	Model	Frequency [Hz]		
		63	125	250
Floor	A	0.07	0.1	0.15
	D	0.14	0.13	0.14
	Rests	0.1	0.12	0.15
Glass windows	A	0.11	0.07	0.15
	Rests	0.12	0.06	0.2
Seats	D	0.28	0.28	0.34
	Rests	0.24	0.25	0.31
Front wall	A	0.1	0.13	0.2
	Rests	0.12	0.15	0.2
Wall	A	0.26	0.09	0.2
	B	0.25	0.25	0.25
	C	0.51	0.51	0.51
	D	0.88	0.81	0.73
Ceiling	A	0.1	0.07	0.15
	B	0.33	0.33	0.33
	C	0.4	0.4	0.4
	D	0.79	0.72	0.69

ceiling, while other boundaries were just slightly adjusted. From Fig. 3, Model A and Model B have the same RT at 500–4000 Hz, while Model C and Model D have the same RT at 500–4000 Hz. Model A and Model C have a rising RT at 63–500 Hz, while Model B and Model D have a flat RT at 63–500 Hz. It can be seen that this hybrid method can deal with the RT quantitatively, which is beneficial to analyze the effects of low-frequency RT on speech intelligibility.

2.4. Speech intelligibility test

As the simulation results of Model A were consistent with the measured results, this study only conducted listening tests on the simulated models. In a quiet room where the background noise level was less than 30 dBA, subjective evaluation of the speech intelligibility was conducted using recordings of each receiving position. These signals had been processed based on the Mandarin Chinese speech intelligibility test word list as specified by GB 15508-1995 (GB, 1995). Each receiving position used two different lists, which were recorded by a man and a woman reading, respectively. Each list has 75 syllables, which are randomly divided into 25 rows of three syllables with no coherent meaning. The lists keep the balance of difficulty and phonemic characteristic, where each consonant, vowel, or tone appears with the same frequency in each list. Each row is preceded by a carrier phrase, for example “the tenth row is ā, ér, jǐng”, where the carrier phrase

gives a hint of row number and “ā, ér, jìng” stands for the three syllables. The words were recorded in an anechoic chamber, and spoken by a male or a female speaker at the rate of about 4 syllables per second. There is a pause of about 5 seconds between each row for the listener to write down the syllables. The testing word lists were convolved with simulated binaural RIRs by Cooledit Pro software.

Fourteen graduate students aged between 22–30 participated in this speech intelligibility test, who were trained and familiar with Chinese phonetic alphabet. All participants are native speakers of Mandarin Chinese and had absolute thresholds of less than 15 dB HL at octave frequencies between 125 Hz and 8000 Hz. For each test condition, up to 4 participants could participate in the test at the same time using the HP-S4 headphone amplifier, and each participant wore the same type of the Sennheiser HD580 headphone at a speech sound pressure level of 60–65 dBA. The listening material was played through Cool Edit Pro software and controlled by the tester. Finally, the results were scored by testers against the correct answer. As the tonal error rate was very low and some subjects accidentally mis-marked tones, the tonal results will not be discussed in the study. Only the consonant and the vowel in each syllable are correct, the syllable is counted as a correct syllable. The correct rate of syllables in each list is calculated by percentage, then the average score of all participants is the Chinese speech intelligibility score of each receiving position. The speech intelligibility score of consonants depends on the correct rate of consonants, no matter whether its vowel part is correct or not. Similarly, the speech intelligibility score of vowels does not consider whether the consonants are right or not.

3. Results

Figure 4 shows the Chinese speech intelligibility scores at each receiving position in four models. It can be obviously seen from the results that the Chinese speech intelligibility scores of Model B at each receiving position are higher than the scores of Model A, and the scores of Model D at each receiving position are higher than Model C, which indicates the intelligibility score of flat RT at a low frequency is higher than that of rising RT at a low frequency. Besides, the standard deviation of the Model A and Model B at each receiving position is basically larger than Model C and Model D. The scores of Model C and Model D are much higher than that of Model A and Model B, which indicates that the RT characteristic in the original classroom is insufficient. From the repeated measurement analysis of variance, the model ($F(3, 234) = 400.927$) and receiving position ($F(5, 78) = 49.377$) have significant effects on the Chinese speech intelligibility scores ($p < 0.001$). As there is a significant interaction be-

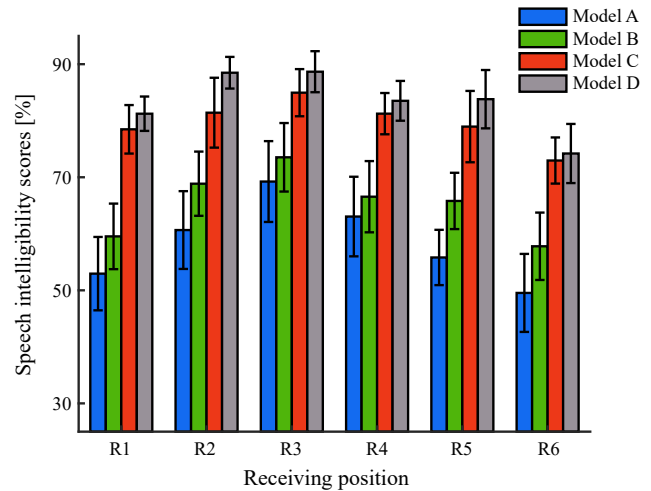


Fig. 4. Chinese speech intelligibility scores in four models.

tween the model and the listening location ($p < 0.001$), it needs paired comparison of the simple effect.

The results of the paired comparison indicates that the speech intelligibility scores of Model C and Model D are significantly higher than that of Model A and Model B ($p < 0.05$). For receiving positions R1, R2, R5, and R6, there is a significant difference between Model A and Model B ($p < 0.05$), whereas there is no significant difference between Model A and Model B for receiving positions R3 and R4 ($p > 0.05$). Meanwhile, between Model C and Model D, there is no significant difference for receiving positions R1, R3, R4, and R6, while there is a significant difference for receiving positions R2 and R5. Therefore, for most locations, flattening the low-frequency characteristics in the classroom with $RT_{(500-1000\text{ Hz})}$ of 2 s can significantly improve the speech intelligibility. However, flattening the low-frequency characteristics in the classroom with $RT_{(500-1000\text{ Hz})}$ of 0.8 s can improve the speech intelligibility but not significantly. Namely, it is more important to improve the entire-frequency RT than improve the low-frequency RT only.

4. Discussion

To further explore the effect of low-frequency RT on speech intelligibility, the following is analyzed in terms of syllables. Mandarin Chinese speech sounds range from very low (about 100–125 Hz) to very high frequencies (above 10 kHz or 12 kHz for some sounds). A Chinese syllable must contain an initial consonant and a vowel, or only contain a vowel. The vowels are low in frequency and high in sound energy, while the initial consonants are much higher in frequency and lower in sound energy (WU, 1964). GELFAND (1998) wrote in his book that low frequencies tended to be effective maskers over a very wide range of frequencies, while higher frequencies were not good maskers of

low frequencies. When the low-frequency RT is much longer than the mid-frequency RT, low-frequency reverberant sounds can be emphasized by room modes and then mask speech sounds in a classroom (WU *et al.*, 2014). For this speech intelligibility test, three syllables are in a row, and each row is preceded by a carrier phrase. Therefore, due to the difference in RTs, syllables in different positions and carrier phrase will affect the test results. The speech intelligibility scores of syllable, the consonant and the vowel (SCV) for tested syllable orders in turn are shown in Fig. 5.

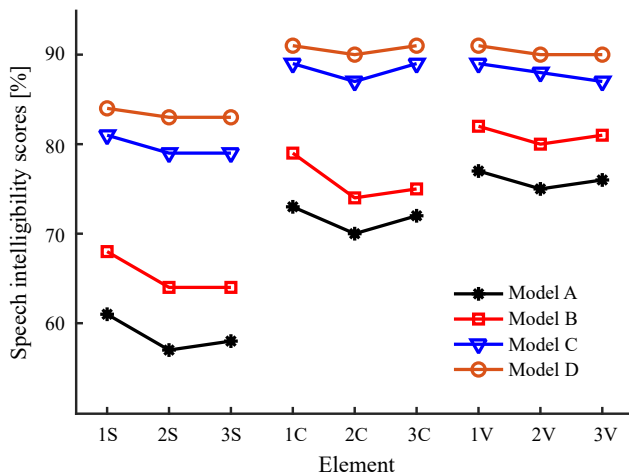


Fig. 5. Speech intelligibility scores of SCV. In the figure, 1S, 2S, and 3S denote syllable in the first, second, and third order, respectively; 1C, 2C, and 3C denote consonant in the first, second, and third order, respectively; 1V, 2V, and 3V denote vowel in the first, second, and third order, respectively.

The analyses of variance for average speech intelligibility scores of six receiving positions show that the model ($F(3, 468) = 614.176$), SCV ($F(2, 468) = 419.248$) and the orders ($F(2, 468) = 14.972$) have significant effects on the test ($p < 0.001$). Figure 5 shows that the scores of Model B are better than Model A, and Model D is better than Model C, which emphasized that the lower the low-frequency RT, the higher the speech intelligibility score. Due to the carrier phrase before the first syllables is fixed and slightly stopped, the first syllables will be less affected by the preceding syllables. Hence, the scores of syllables, consonants and vowels in first orders are significantly higher than the second and third orders ($p < 0.001$). The scores of vowels and consonants are significantly higher than syllables ($p < 0.001$), meanwhile the scores of vowels are significantly higher than consonants ($p < 0.001$).

There is a significant interaction between the model and SCV ($p < 0.001$), and the other interactions are not significant ($p > 0.05$). For the same element of SCV, there is a significant difference between each two models ($p < 0.05$), however, there is not significant difference between each two elements of SCV. For Mod-

els A and B, there is a significant difference between each two elements of SCV ($p < 0.001$). In Model A and Model B, the average score of consonants is lower than the average score of vowels in the same model (the consonant of Model A = 72%, the vowel of Model A = 76%, the consonant of Model B = 76%, the vowel of Model B = 81%). As both the average RTs of Model A and Model B at 500–1000 Hz are 2 s, while the reading speed is about 4 syllables per second, the RT is greater than the time interval between two syllables, which will bring serious interference. Since both the low frequencies and higher frequencies have a masking effect on the high frequencies, the front syllables of Model A and Model B have large masking effect on the higher frequency sound of the next syllables. Therefore, the scores of consonants are lower than vowels in Model A and Model B. For Model C and Model D, there is no significant difference between vowels and consonants ($p > 0.05$), but there is a significant difference between vowels and syllables ($p < 0.001$) or between consonants and syllables ($p < 0.001$). In Model C and Model D, the average scores of the initial consonant and vowel of the same model are basically the same (the consonant of Model C = 88%, the vowel of Model C = 88%, the consonant of Model D = 91%, the vowel of Model D = 91%). Owing to that the average RTs of Model C and Model D at 500–1000 Hz are only 0.8 s, the RT is too small to cause an obvious masking effect. Therefore, the scores of consonants are basically the same as that of vowels in Model C and Model D.

The above findings indicate that the vowel sounds have a certain masking effect on the consonant sounds, namely the low-frequency sound has a certain masking effect on the mid-high frequency sound, which confirmed the well-known “upward spread of masking” (OXENHAM, PLACK, 1998). The result of this study showed that too high reverberation time in the low frequency band significantly deteriorates the intelligibility of speech. To improve the speech intelligibility of a classroom, the average RT in the entire frequencies should be smaller to diminish the masking effect and the low-frequency RT is better not to be rising.

5. Conclusions

In this study, four models based on a large classroom were established by a hybrid method. The synthetic RIRs of the classroom models were obtained through the FDTD method modeling at low frequencies and the Odeon simulation at higher frequencies. The Chinese speech intelligibility listening test was conducted after verification. The average scores of each receiving position show that the scores of Model B are better than Model A, and Model D is better than Model C. The scores of syllables, consonants and vowels in first orders are significantly higher than the second and third orders. The scores of vowels and conso-

nants are significantly higher than syllables. The scores of consonants are lower than vowels in Model A and Model B, while the scores of consonants are basically the same as that of vowels in Model C and Model D. The above results indicate that RT in the entire frequencies should be smaller and the low-frequency RT is better to be flat to obtain better speech intelligibility.

References

1. ADELMAN-LARSEN N.W. (2015), Possible acoustic design goals in very large venues hosting live music concerts, *Auditorium Acoustics*, **37**(3): 308–316.
2. BARRON M. (2010), *Auditorium Acoustics and Architectural Design*, 2nd ed., Spon Press, London and New York.
3. BERANEK L.L. (1962), *Music, Acoustics and Architecture*, Wiley, New York.
4. BERANEK L.L. (1996), *Concert and Opera Halls: How They Sound*, Acoustical Society of America, New York.
5. BERANEK L.L. (2010), Listening to the acoustics in concert halls, [in:] *Proceedings of the International Symposium on Room Acoustics (ISRA 2010)*, Melbourne.
6. BOTTELDOOREN D. (1995), Finite-difference time-domain simulation of low-frequency room acoustic problems, *The Journal of the Acoustical Society of America*, **98**(6): 3302–3308, doi: 10.1121/1.413817.
7. FUCHS H.V., STEINKE G. (2015), Requirements for low-frequency reverberation in spaces for music: Part 2: Auditoria for performances and recordings, *Psychomusicology: Music, Mind, and Brain*, **25**(3): 282–293, doi: 10.1037/pmu0000089.
8. GB (1995), GB 15508-1995, Acoustics – Speech articulation testing method [in Chinese], *Standard of PR China*.
9. GB (2010), GB 50118-2010, Code for design of sound insulation of civil buildings [in Chinese], *Standard of PR China*.
10. GB/T (2005), GB/T 50356-2005, Code for architectural acoustical design of theater, cinema and multi-use auditorium [in Chinese], *Standard of PR China*.
11. GELFAND S.A. (1998), *Hearing: An Introduction to Psychological and Physiological Acoustics*, Marcel Dekker, New York.
12. JGJ/T (2012), JGJ/T 131-2012, Specification for acoustical design and measurement of gymnasium and stadium [in Chinese], *Standard of PR China*.
13. KOWALCZYK K., VAN WALSTIJN M. (2008), Modeling frequency-dependent boundaries as digital impedance filters in FDTD and K-DWM room acoustics simulations, *Journal of the Audio Engineering Society*, **56**(7/8): 569–584, <https://www.aes.org/e-lib/browse.cfm?elib=14401>.
14. OXENHAM A.J., PLACK C.J. (1998), Suppression and the upward spread of masking, *The Journal of the Acoustical Society of America*, **104**(6): 3500–3510, doi: 10.1121/1.423933.
15. OXNARD S. (2018), Investigating the stability of frequency-dependent locally reacting surface boundary conditions in numerical acoustic models, *The Journal of the Acoustical Society of America*, **143**(4): EL266–EL270, doi: 10.1121/1.5030917.
16. PENG J., LAU S.K., ZHAO Y. (2020), Comparative study of acoustical indices and speech perception of students in two primary school classrooms with an acoustical treatment, *Applied Acoustics*, **164**: 107297, doi: 10.1016/j.apacoust.2020.107297.
17. SAKAMOTO S., NAGATOMO H., USHIYAMA A., TACHIBANA H. (2008), Calculation of impulse responses and acoustic parameters in a hall by the finite-difference time-domain method, *Acoustical Science and Technology*, **29**(4): 256–265, doi: 10.1250/ast.29.256.
18. SOUTHERN A., SILTANEN S., MURPHY D.T., SAVIOJA L. (2013), Room impulse response synthesis and validation using a hybrid acoustic model, *IEEE Transactions on Audio, Speech, and Language Processing*, **21**(9): 1940–1952, doi: 10.1109/TASL.2013.2263139.
19. WU S., PENG J., BI Z. (2014), Chinese speech intelligibility in low frequency reverberation and noise in a simulated classroom, *Acta Acustica united with Acustica*, **100**(6): 1067–1072, doi: 10.3813/AAA.918786.
20. WU Z. (1964), The spectrographic analysis of the vowels and consonants in standard colloquial Chinese [in Chinese], *Acta Acustica*, **1**(1): 33–40, doi: 10.15949/j.cnki.0371-0025.1964.01.006.
21. XU S., PENG J., XIAO Y., HUANG W. (2021), The effect of low frequency reverberation on Chinese speech intelligibility in two classrooms, *Applied Acoustics*, **182**: 108241, doi: 10.1016/j.apacoust.2021.108241.
22. ZHA X., LYU H. (2020), Analysis and improvement of classroom acoustic environment [in Chinese], *Technical Acoustics*, **39**(4): 461–467, doi: 10.16300/j.cnki.1000-3630.2020.04.014.