Article:

Zhu, P., Mo, F. and Kang, J. (2014) Relationship Between Chinese Speech Intelligibility and Speech Transmission Index Under Reproduced General Room Conditions. Acta Acustica united with Acustica, 100 (5). 880 - 887. ISSN 1610-1928

https://doi.org/10.3813/AAA.918767

Relationship between Chinese speech intelligibility and speech transmission index under reproduced general room conditions

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

[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, Western Bank, Sheffield S10 2TN, UK*

* Corresponding author

Summary: The subjective Chinese (Mandarin) articulation scores of a total of 50 sound conditions, namely at 12 receiver positions in four rooms, were obtained by expert listeners based on in-situ measured binaural room impulse responses (BRIRs) and binaural technology for headphone reproduction from a phonetically balanced test. The relationship between Chinese speech articulation scores and the speech transmission index (STI) according to IEC 60268-16 was established. The spectrum difference between the average spectrum of Chinese and that recommended by IEC may noticeably influence the measurement results with a STI difference of up to 0.054. However, the STI method can evaluate Chinese intelligibility without modulating the correction factor values recommended by IEC; the standard deviation between the articulation scores and the curve was 5.70%, which is relatively small.

2014 Acta Acustica united with Acustica
Date Received:23 December 2013 **Date Accepted**: 23 June 2014

1. Introduction

Speech intelligibility is an important metric that can be used to evaluate the sound transmission quality of auditorium or public address system. The assessment of speech intelligibility mainly includes subjective and objective evaluations [1, 2]. The objective evaluation metric speech transmission index (STI) can reflect the influence of the reverberation time (RT) and signal-to-noise ratio (SNR) on speech intelligibility and has certain ability to account for non-linear distortion [3, 4]. Therefore, it has been adopted and recommended by the IEC (International Electrotechnical Commission) standard [5] and is widely used in the field of building acoustics. According to the curve of speech intelligibility scores and STI, the subjective listening quality can be estimated directly after the objective metric STI is determined. However, the curves may depend on the situation, because subjective and objective evaluation methods, as well as the test conditions, may be different. Differences in language could affect the curve [6].

Chinese is a special monosyllabic language; initials (consonants), finals (vowels) and tones are the main components of the syllable information, which is considerably different from western languages, such as Dutch and English [7, 8, 9]. Few studies have examined the relationship between Chinese speech intelligibility scores and STI, such as Kang's [10] and Peng's [11] works. In their studies, Chinese phonetically balanced (PB) word (or monosyllable) lists were used as the test material; however, the STI method used has been developed specifically for Dutch nonsense consonant vowel consonant (CVC) words. In the objective speech intelligibility rating method recommended by IEC [5], the test signals correspond to specified speech spectrum and are corrected for the absolute speech reception threshold, auditory masking, weighting factor and redundancy factor in the later stage of the STI computation. The absolute speech reception threshold was used to correct the SNR if the background noise level was extremely low. Although the absolute speech reception thresholds of Chinese and Western languages differ, few opportunities exist to correct the threshold in practical applications. Auditory masking is related to the sound level; it only exerts influence when the sound level is high, while the weighting factor and redundancy factor are influenced by the speech materials. Therefore, the authentic objective speech intelligibility rating method of Chinese shall at least reflect the influence of the average spectrum of Chinese speech, and the specified weighting factor and redundancy factor of Chinese speech should be corrected for.

Therefore, this study aimed to obtain an authentic curve of speech intelligibility scores and STI for Chinese. Four different general rooms were selected to first measure STI over an average spectrum of Chinese speech, and speech intelligibility tests were then conducted with Chinese PB word lists to establish a curve of Chinese speech intelligibility scores and STI. In addition, the average spectrum of Chinese and that recommended by IEC, as well as the influence of the spectrum difference on the measurement results, were compared. The weighting factor and redundancy factor of Chinese speech were also discussed.

2. Methods

This section first conducts a speech intelligibility test with Chinese PB word lists in four different general rooms, it then measures STI with an average spectrum of Chinese speech.

2.1 Experimental arrangement

Four rooms were selected as the test rooms for the subjective and objective evaluation of Chinese speech intelligibility, including an office, a laboratory, a multimedia lecture hall and a semi-anechoic chamber (with one desk and four chairs inside). Of these rooms, the office, the laboratory and the semi-anechoic chamber were rectangular, and the multimedia lecture hall was an octagon. Two receiver positions were arranged in the office, three receiver positions were arranged in the laboratory, six receiver positions were arranged in the multimedia lecture hall, and one receiver position was arranged in the semi-anechoic chamber. The layout of the receiver positions and the sound sources are shown in Figure 1, and the characteristics of the four rooms are shown in Table 1, which also provides the early decay time (EDT), reverberation time (RT30) and clarity (C80).

To obtain a wide range of STI, a noise source (PYRITE dodecahedral sound source) was arranged at a distance of 0.5 m from the signal source. This source reproduced a pink noise, which was modulated according to the average spectrum of Chinese speech at four different sound levels. The sound level at 1 m directly ahead of the signal source was set to 60 dB [5] in the semi-anechoic chamber, and the sound level of the noise source was simultaneously adjusted to ensure that the position 1.5 m away from the two sound sources correspond to four different SNRs: -5 dB, 0 dB, 10 dB and 20 dB. To obtain more sound conditions with high intelligibility, at receiver position R12 in the semi-anechoic chamber, SNRs of 5 dB, 15 dB were added. These SNRs does not represent the actual SNRs at the receiver positions on site. Subsequently, the signal source and the noise source, which were pre-set in the anechoic chamber, were placed in the corresponding sound source positions in the test rooms, and the measurements were carried out at each receiver position.

Table 1. Characteristics of the four test rooms.

Room Type	Capacity (m^3)	Receiver position	*EDT*（s）	T_{30}(s)	C_{80}（dB）
Office	108	R_1	0.52	0.63	9.58
		R_2	0.59	0.67	8.22
Laboratory	238	R_3	1.39	1.50	3.60
		R_4	1.52	1.48	0.57
		R_5	1.65	1.56	-0.18
Multimedia lecture hall	1674	R_6	0.82	0.82	10.54
		R_7	0.73	0.80	7.06
		R_8	0.66	0.78	6.76
		R_9	0.54	0.73	9.79
		R_{10}	0.53	0.73	9.09
		R_{11}	0.59	0.77	8.27
Semi-anechoic chamber	550	R_{12}	0.05	0.03	64.26

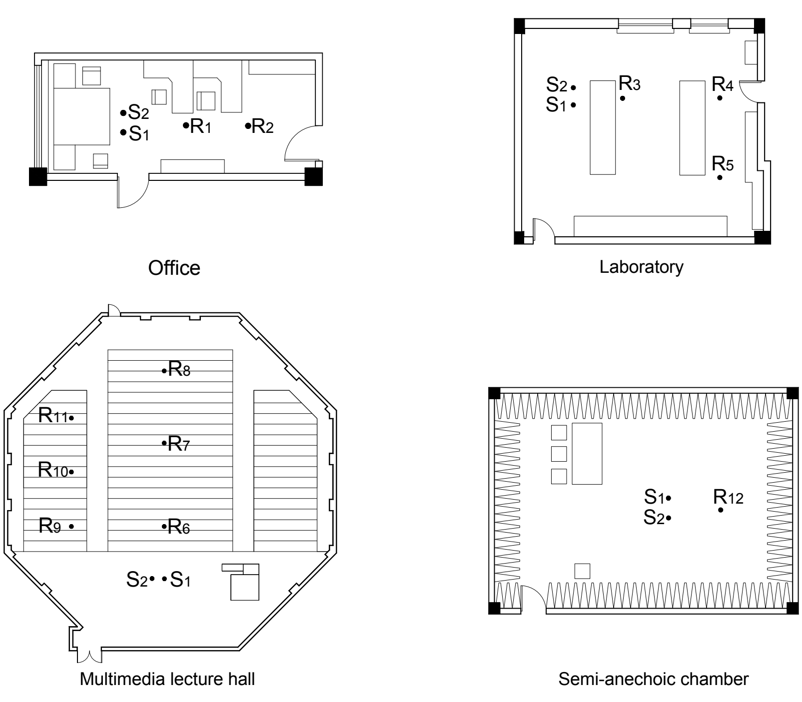

Fig. 1 The layout of the receiver positions and the sound sources in the four rooms. S_1 is the signal source and S_2 is the noise source, R_1-R_{12} are the receiver positions; the height of the receiver positions is 1.2 m, and the height of the loudspeaker is 1.5 m.

2.2 Subjective intelligibility evaluation

Subjective Chinese speech intelligibility tests are usually conducted in two ways: a Diagnostic Rhyme Test (DRT), specified by GB/T 13504-2008 [12] and SJ2467-84 [13], or a PB word dictation test, specified by GB/T 15508-1995 [14]. DRT is a type of closed set test, for which higher scores are more easily obtained [11, 15]. 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 [14] 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 and a female 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 [14], to guarantee all the subjects were familiarised with the PB word lists.

Eight testing juries, a total of 64 subjects (8 subjects each jury, four juries listened male voice speech material and another four juries listened female voice speech material), participated in the subjective test, 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 the subjects prior to the test to ensure that they understand the entire test process and master the key points. The experiment results indicated that there is no significant relationship between the HL and the subjective evaluation scores in this study. This is perhaps because there are many factors influencing the subjective evaluation scores, such as personal characters, cultural background and the degree of concentration for testing. The influence of the HL might be relatively less important compared with that of other factors.

Under general room conditions, the authentic speech intelligibility test should be conducted in real rooms using an artificial mouth to reproduces the test word lists and at the same time the subjects at the receiver position write down the word they heard. However, performing such a test in real rooms has many limitations, for example, it is difficult to perform a speech intelligibility test simultaneously with a large number of subjects at a single receiver position, and there is also no guarantee that the repeated test conditions are exactly the same. To avoid these limitations, the test was conducted via reproduction in headphones in this study to produce the same listening effect as actual listening on site, and thus, the BRIRs measured at the receiver position using a binaural microphone B&K 4101-A with the ear canals blocked and pointing to the signal source namely artificial mouth GRAS 44AA were used. The speech material was the signals convolving the in-situ measured BRIRs with the PB word lists recorded in an anechoic chamber. The headphone used in the test was a Sennheiser HD-600 model, the corresponding power amplifier was a Rane-HC4s model, and the audio interface was a B&K ZE-0948 model. HD-600 is a type of open-air headphone that is easily affected by ambient noise. Therefore, the test was conducted in a semi-anechoic chamber. Many studies have debated the ability of headphones to produce the same listening effect as actual listening on site [16, 17]. 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 sound level of the reproduction was exactly the same as the sound level on site; the equalisation processing of reproduction system; whether the pre-set SNRs can be truly reflected, etc. In this study, the conditions can be realised via the signal processing program compiled based on binaural technology [18]. The recording signal reproduced through the headphone placed on the artificial head and the actual signal reproduced through the artificial mouth recorded by the artificial head were compared. The comparison results are shown in Table 2.

The sound level 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 sound level is based on the removal of the silent parts of the speech signals, i.e., the pauses between the words [19], which were used to calibrate the word lists in this study. Naturally, the sound level of all other test signals, such as the noise signal and test signal for the direct method, etc., were accurately controlled at the beginning of signal generation by filtering to ensure the complete equivalence of the sound level between the test signals and speech signals.

To reduce the inaccuracy due to listening sequence and memory, 48 of 50 test conditions (SNRs of 5 dB, 15 dB test conditions at receiver position R_{12} were not included) were designed into four incomplete Latin squares, each including 12 test conditions (3 receiver positions multiply 4 SNRs), for listening by one jury member. Tests for male voice and female voice were conducted respectively. The repeated use of word lists may significantly influence the measurement results [20]. Therefore, 12 different word lists were used to ensure that one exclusive word list was used in each of the 12 test conditions in one incomplete Latin square, i.e., each subject can hear 12 different word lists. In each incomplete Latin square, one jury member performed the test in order. With SNRs of 5 dB, 15 dB test conditions at receiver position R_{12} another two different word lists were used (thus a total of 14 word lists were used in this study), and the two tests were carried out by two juries separately.

No significant differences were observed in the intelligibility scores resulting from the male or female speaker in this study although there are differences in the speech spectrum and personal pronunciations which may exert influence. The speech intelligibility scores given by the 16 subjects for each test condition were averaged. Finally, 8 speech intelligibility scores were obtained in the office, 12 scores in the laboratory, 24 scores in the multimedia lecture hall and 6 scores in the semi-anechoic chamber. In total, 50 Chinese speech intelligibility scores were obtained.

Table 2. Comparison results between the recording signal reproduced through the headphone that is put on the artificial head (Sig.1) and the actual signal reproduced through artificial mouth recorded by the artificial head (Sig.2). Receiver position R_4, 10 dB SNR. 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	49.7	54.53	54.85	52.86	47.57	52.6	45.88	59.56
	Right ear	49.32	54.76	55.07	53.41	48.82	53.93	47.47	60.44
Sig.2 (dB)	Left ear	50.65	54.9	55.33	52.99	47.74	53.39	49.11	60.18
	Right ear	50.04	54.92	54.44	53.03	48.11	53.96	49.95	60.37
Difference between Sig.1 and Sig.2 (dB)	Left ear	-0.95	-0.37	-0.48	-0.13	-0.17	-0.79	-3.23	-0.62
	Right ear	-0.72	-0.16	0.63	0.38	0.71	-0.03	-2.48	0.07

2.3 Objective intelligibility evaluation

Two STI measurement methods are recommended by IEC 60268-16 [5], namely the direct method based on signal modulation and the indirect method based on impulse response. With the exception of some commercial STIPA meters adopted in the direct measurement method [21], the available STI measurement platforms, such as Dirac, Aurora, etc., all adopt indirect measurement methods [22, 23]. The direct measurement method is usually time-consuming, and the data processing is relatively taxing on the hardware. Moreover, the full STI is more difficult to measure using the direct method, which is usually used in research according to the suggestion of the IEC standard [5]. In this study, the full STI test signals modulated according to the average spectrum of Chinese speech and computing programs of direct measurement method were compiled. Correction factors including weighting and redundancy factors used in the computing programs were based on the IEC standard [5].

The measurement system used in this paper included a signal source GRAS 44AA, noise source PYRITE, power amplifier AMPHION (corresponding to PYRITE), audio interface B&K ZE-0948, microphone B&K 4189 (power supply is B&K 1704), sound recording software Cool Edit Pro and binaural microphone B&K 4101-A. To reduce the inaccuracy caused by the measurement system, loop calibration was carried out to ensure that the system is linearly time-invariant (LTI) without harmonic distortions. Considering that the artificial mouth and noise source have poor frequency response, the system was equalised in the semi-anechoic chamber using an inverse filter system, and the sound level of the two sound sources were also calibrated in the semi-anechoic chamber.

The measurement program used in this study is also the key to the reliability of the measurement result. The measurement signal generated according to the standard, whose modulation depth was 0.8, was directly processed as a received signal (loop calibrated), and it was important to note that the 98 values were not 0.8 anymore. Many factors may cause a loss in signal, such as the pink noise pseudo-random signal, the filtering and modulation that are required during signal generation, and the filtering and envelope extraction that are also required for the received signal. To evaluate these disparities, one-sample T test method in statistics was used to examine any significant differences between the population mean of the 98 figures and the designated value of 0.8. The p-value was shown to be 0.819, which indicated the absence of any significant variations. 98 modulation transmission values after passing the loop and 98 modulation transmission values of 0.8 were also input into the program to simultaneously calculate the STI. The STI difference under the two conditions was only 0.001, which was far smaller than a JND (Just Noticeable Difference, approximately 0.03) [24]. This finding shows that the inaccuracy introduced by the measuring program is very small.

3. Experimental results

This section first establishes and validates the relationship between Chinese speech articulation scores and STI according to IEC 60268-16, and this is then compared with the curves obtained by others and with the curve for English. The section subsequently examines the effect of the test rooms and spectrum difference. The weighting factor and redundancy factor of Chinese speech are also discussed.

3.1 Relationship between Chinese speech intelligibility and STI

Data based on 50 subjective and objective evaluations and their best-fitting third-order polynomial curve are given in Figure 2. The curve essentially covers a wide range of speech intelligibility scores from 6% to 98% and STI values from 0.049 to 0.972. The coefficient of determination, R^2, of the curve is 0.964, and the standard deviation between the speech intelligibility scores and the curve is 5.70%, which demonstrate a high correlation between Chinese speech intelligibility scores and STI. The relationship between Chinese speech intelligibility scores and STI is defined in Eq. (1):

$$y = 54.989x^3 - 238.51x^2 + 287.35x - 8.056 \tag{1}$$

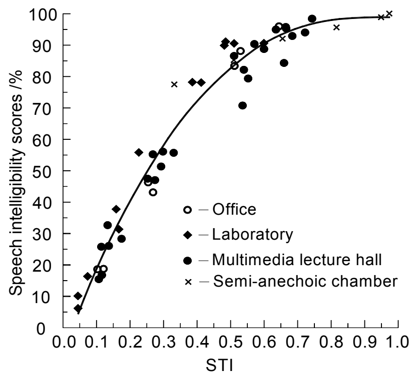

Fig. 2 Relationship between Chinese speech intelligibility scores and STI

3.2 Comparison with other curves

Comparisons of the present study with Kang's [10] and Peng's [11] curves are given in Figure 3. All these curves are from dichotic listening conditions. Chinese speech intelligibility scores strongly depend on the speech test used. As the figure shows, the differences between curves for the PB word test and DRT are considerable.

The curve obtained in the present study is similar to that published by Kang [10] when the STI is in the vicinity of 0.64, but the difference becomes larger when the STI is lower or higher. The difference with Kang's curve is perhaps due to differences in the spectrum and objective speech intelligibility metric used. The accuracy of RASTI may be affected by the frequency characteristics of RT [25]; the spectrum difference may also influence the measurement results (see Section 3.4). In addition, the test rooms were less representative than those in this study: one test room was a corridor, i.e., a typical rectangular long enclosure, while another was a seminar room with a volume of only 110 m^3, and the speech intelligibility of the two test rooms were also limited. The difference between the present study and Peng's [11] curve for the PB word test is smaller compared with that between the present study and Kang's [10] curve when STI is lower than 0.56, and larger when STI increases. In Peng's study, the curve was obtained based on simulation, which is rather different from the in-situ measurement.

A comparison of Chinese speech intelligibility scores and English speech intelligibility scores [2] corresponding to STI is shown in Figure 4. Although there are differences in test conditions and measurement methods for the two curves, and the English speech intelligibility scores, which were obtained from the diotic listening condition, might be lower than those from the dichotic listening condition [11], the two curves are approximately similar, and the largest score difference, about 4.5%, appears when STI is in the vicinity of 0.4, despite the fact [20] that Chinese is a special monosyllabic language whose syllable structure is more constrained, Chinese tones are difficult to interfere with, and the loss of tone articulation is very small under noise masking, frequency distortion and nonlinear distortion conditions, which is helpful to enhance intelligibility [15, 26]. In addition, the average spectrum of Chinese speech was used in this study, possibly resulting in higher STI values (see Section 3.4).

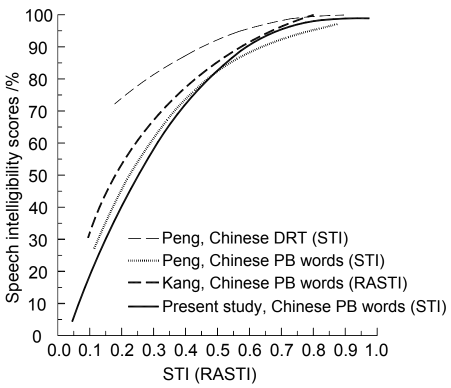

Fig. 3 Comparison of the present study with other curves

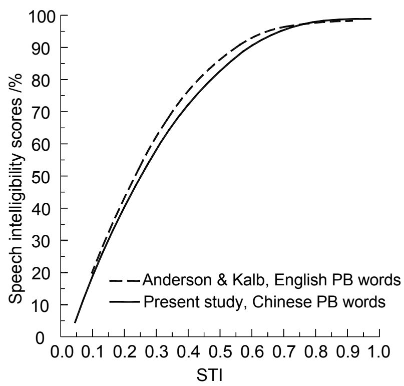

Fig. 4 Comparison of Chinese speech intelligibility scores with English speech intelligibility scores corresponding to STI

3.3 Effect of room

The curves of the four rooms are given in Figure 5. These curves are close to each other, but they still differ. In these four rooms, the RT is the longest in the laboratory and the shortest in the semi-anechoic chamber. For a given STI value, the speech intelligibility scores in the laboratory were relatively higher; the scores in the semi-anechoic chamber were relatively lower when the STI value was high, while they increased as the STI value decreased. The RT in the office was close to that in the multimedia lecture hall, and the curves were also similar. These results indicate that the curves may also depend on the room conditions, and with similar STI values, the speech articulation scores may be different.

This difference is likely due to the variations in sound field between the four rooms, as sound field characteristics are highly related with intelligibility [10, 27]. The measurement of STI was carried out using a single omnidirectional microphone and thus, the influences of sound field characteristics such as direction, diffusion and early-to-late arriving sound ratio (C50) need to be considered.

Fig. 5 Relations between speech intelligibility scores and STI of the four rooms

3.4 Effect of speech spectrum

Average spectrum of Chinese and that recommended by IEC [5] is given in Table 3. The octave band levels of Chinese in the table were obtained by calculating the power spectrum density in GBT 7347-87 [28]. The table shows that the level of Chinese is lower than that recommended by IEC at 125Hz and 500Hz, i.e., it is almost the same at 1000Hz, while it is higher at frequencies of 2000Hz, 4000Hz and 8000Hz.

To accurately compare the influence of the test signals of the average spectrum of Chinese and test signals of the spectrum recommended by IEC on STI, the STI of 5 receiver positions in the office and laboratory were measured for a total of 20 listening environments. Because identical conditions are preferred for the two spectra, pink noise (the frequency response was equalised for a PYRITE noise source) without spectrum modulation was used as a noise signal, and all the other conditions were the same. The result is given in Figure 6. The STI difference was small at -5 dB and 0 dB SNR, the STI measured for the Chinese spectrum was higher at 10 dB and 20 dB SNR, and the STI difference was maximised at 0.054 for 10 dB SNR at receiver position R_3 in the laboratory, which exceeds a JND; the difference did not increase further at 20 dB. Generally, the STI values of the Chinese spectrum may be higher. In combination with the data given in Table 3 discussed above, which show that the energy of Chinese speech is higher than that recommended by IEC at high frequencies, and high frequencies have more contributions than that of low frequencies to speech intelligibility, this result is reasonable.

Table 3. Octave band levels (dB) relative to the A-weighted speech level for Chinese and that recommended by IEC.

Frequency band(Hz)	125	250	500	1000	2000	4000	8000	
Chinese	-3.0	3.8	-2.2	-6.9	-10.9	-14.8	-18.8	0.0
Recommended by IEC	2.9	2.9	-0.8	-6.8	-12.8	-18.8	-24.8	0.0

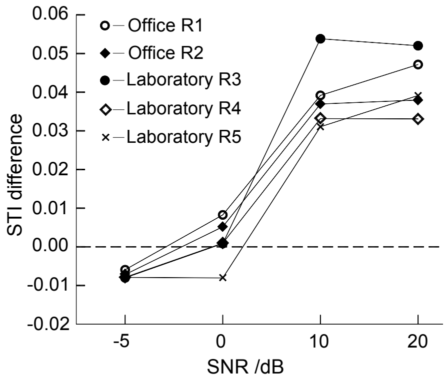

Fig. 6 STI difference between with Chinese speech spectrum and with that recommended by IEC

3.5 Weighting factor and redundancy factor of Chinese speech

The weighting and the redundancy factor in IEC 60268-16 are defined by computing the minimum standard deviation between the subjective speech intelligibility scores and regression curve, as described by Steeneken and Houtgast [29]. In this paper, the standard deviation was 5.70%, which is smaller than the 6.9% figure reported by Anderson and Kalb for English [2] and larger than the 4.73% figure reported by Steeneken and Houtgast [29] for Dutch. This finding indicates that the weighting factors and redundancy factors of western languages and Chinese may not differ significantly, which corresponds to the results of a previous study [30], which showed that the precision of STI for assessing Chinese speech intelligibility using the weighting factors recommend by IEC is within a good range.Given that the data obtained in this study are relatively limited, the weighting factor and redundancy factor of Chinese speech was not computed.

4. Conclusions

In this paper, a curve of Chinese speech intelligibility scores and STI that complies with the current IEC 60268-16 was built based on in-situ measured BRIRs and binaural technology for headphone reproduction. A comparison with previous curves indicates that the adoption of a spectrum of Chinese and revised STI method may significantly influence the presented curve. The curves may also depend on the room conditions, and the analysis of different curves indicates that with similar STI values, the speech articulation scores may be different.

The difference between average spectrum of Chinese and that recommended by IEC may noticeably influence the measurement results with a STI difference of up to 0.054. However, the STI method can evaluate Chinese intelligibility without the modulation of correction factor values, for which the standard deviation between articulation scores and the curve was 5.70%, which is relatively small.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 51278078, 11274243, and 51308087) and Fundamental Research Funds for the Central Universities (Grant No. DUT13RW306).

References:

[1] N. R. French, J. C. Steinberg: Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am. 19 (1947) 90–119.
[2] B. W. Anderson, J. T. Kalb: English verification of the STI method for estimating speech intelligibility of a communications channel. J. Acoust. Soc. Am. 81 (1987) 1982–1985.
[3] T. Houtgast, H. J. M. Steeneken: The modulation transfer function in room acoustics as a predictor of speech intelligibility.Acustica 28 (1973) 66–73.
[4] T. Houtgast, H. J. M. Steeneken: A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am. 77 (1985)1069–1077.
[5] IEC 60268-16: Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. IEC, 2011.
[6] T. Houtgast, H. J. M. Steeneken: A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria. Acustica 54 (1984) 185–199.
[7] J. L. Zhang: Speech. – In: Handbook of Acoustics. D. Y. Maa, H. Shen (eds.). Science Press, Beijing, 2004, second ed. Chap.19, 404–435.
[8] X. P. Yin: On common language and English pronunciation (in Chinese). J. Shanxi Agric. Univ. 6 (2007) 144–145.
[9] G. Y. Xu, Z. G. Wang: On the contrastive study of Chinese-English phonetics in China in the 20th century (in Chinese).Journal of Zhejiang University (Humanities and Social Sciences)5 (2002) 50–55.
[10] J. Kang: Comparison of speech intelligibility between English and Chinese. J. Acoust. Soc. Am. 103 (1998) 1213–1216.
[11] J. X. Peng: Relationship between Chinese speech intelligibility and speech transmission index in rooms based on auralization. Speech Communication 53 (2011) 986–990.
[12] GB/T 13504-2008: Diagnostic rhyme test (DRT) method of Chinese articulation. Standard of PR China, 2008.
[13] SJ2467-84: Test methods for Chinese articulation in communication equipment. Standard of PR China, 1984.
[14] GB/T 15508-1995: Acoustics – Speech articulation testing method. Standard of PR China, 1995.
[15] Z. H. Meng, L. Dai: Test study on relationship between clarity of Mandarin monosyllable and STIPA caused by reverberation (in Chinese). Audio Engineering 2 (2009) 4–8.
[16] D. Rao, S. X. Wu: Auralization differences between individualized and non-individualized binaural room impulse responses (in Chinese). Journal of South China University of Technology (Natural Science Edition) 8 (2008) 123–127.
[17] D. Rao: Analysis of non-individualized compensation for headphone transfer function (in Chinese). Audio Engineering 3 (2007) 41–43.
[18] M. Henrik: Fundamentals of binaural technology. Applied Acoustics 36 (1992) 171–218.
[19] H. J. M. Steeneken, T. Houtgast: Comparison of some methods for measuring speech levels. Report IZF 1986-
20. TNO-HFRI Human Factors Research Institute, Soesterberg,The Netherlands, 1986.
[20] J. L. Zhang: Chinese human-machine communication system.First ed. (in Chinese). Shanghai scientific & Technical Publishers, Shanghai, 2002, 500–574.
[21] P. Mapp: Is STIPa a robust measure of speech intelligibility performance. AES 118th Convention, Barcelona, Spain,2005.
[22] Brüel & Kjær Sound & Vibration Measurement A/S: Usermanual, v5.0, dirac room acoustics software. Lyngby, Denmark,2010.
[23] A. Farina: User manual of Aurora43. University of Parma,Parma, Italy A/S, 2012.
[24] J. S. Bradley, R. Reich, S. G. Norcross: A just noticeable difference in c50 for speech. Applied Acoustics 58 (1999)99–108.
[25] H. Shen: Experimental study of modulation transfer functions for Mandarin (in Chinese). Audio Engineering 1(1993) 2–7.
[26] G. L. Shen: Discussion on speech sound recording (in Chinese).Motion Picture Video Technol. 5 (1997) 3–11.
[27] S. K. Tang: Speech related acoustical parameters in classrooms and their relationships. Applied Acoustics 69 (2008) 1318–1331.

[28] GBT 7347-87: The standard spectrum of Chinese speech.Standard of PR China, 1987.
[29] H. J. M. Steeneken, T. Houtgast: Mutual dependence of the octave-band weights in predicting speech intelligibility.Speech Communication 28 (1999) 109–123.
[30] H. H. Zhang, H. Xie: Applicability of weighting methods in calculating speech transmission index for assessing speech intelligibility of Chinese in rooms (in Chinese). Journal of Zhejiang University (Engineer Science) 3 (2012) 463–469.

Appendix A. STI values and speech intelligibility scores.

Room Type	Receiver position	SNR (dB)	STI	Score (%) (male voice)	Score (%) (female voice)	Score (%) (average)
Office	R_1	-5	0.120	20.1	17.9	19.0
		0	0.269	44.2	42.4	43.3
		10	0.529	89.2	87.0	88.1
		20	0.664	93.5	98.7	96.1
	R_2	-5	0.108	18.3	18.9	18.6
		0	0.255	47.3	45.5	46.4
		10	0.512	83.2	83.8	83.5
		20	0.646	94.3	97.7	96.0
Laboratory	R_3	-5	0.075	6.2	26.4	16.3
		0	0.227	48.2	63.6	55.9
		10	0.483	84.2	96.6	90.4
		20	0.598	89.2	92.4	90.8
	R_4	-5	0.046	11.3	8.3	9.8
		0	0.168	36.2	26.4	31.3
		10	0.411	82.1	74.3	78.2
		20	0.511	90.1	91.3	90.7
	R_5	-5	0.049	8.1	4.1	6.1
		0	0.159	46.2	29.6	37.9
		10	0.388	83.2	73.2	78.2
		20	0.485	91.1	91.3	91.2
Multimedia lecture hall	R_6	-5	0.177	23.2	33.8	28.5
		0	0.330	58.2	53.0	55.6
		10	0.596	88.2	89.2	88.7
		20	0.743	98.1	97.7	97.9
	R_7	-5	0.119	25.2	26.4	25.8
		0	0.275	47.2	46.6	46.9
		10	0.538	84.2	79.6	81.9
		20	0.665	95.1	94.5	94.8
	R_8	-5	0.108	12.3	18.9	15.6
		0	0.254	44.1	50.9	47.5
		10	0.512	87.2	86.0	86.6
		20	0.637	98.1	91.3	94.7
	R_9	-5	0.140	28.1	24.3	26.2
		0	0.298	65.2	46.6	55.9
		10	0.569	92.1	88.1	90.1
		20	0.720	94.2	93.4	93.8
	R_{10}	-5	0.136	40.3	25.3	32.8
		0	0.292	60.2	42.4	51.3

		10	0.553	81.1	77.5	79.3
		20	0.683	99.2	86.0	92.6
	R_{11}	-5	0.117	22.3	11.5	16.9
		0	0.269	59.1	50.9	55.0
		10	0.534	72.3	68.9	70.6
		20	0.660	91.1	77.5	84.3
Semi-anechoic chamber	R_{12}	-5	0.332	74.1	80.7	77.4
		0	0.510	81.2	89.2	85.2
		5	0.654	90.2	93.4	91.8
		10	0.816	93.1	97.7	95.4
		15	0.948	97.0	97.8	97.4
		20	0.972	98.8	97.6	98.2