**Proceedings of the International Symposium on Room Acoustics, ISRA 2010**

29-31 August 2010, Melbourne, Australia

# Prediction of Chinese speech intelligibility using useful to detrimental sound ratios based on auralization

**Peng Jianxin (1, 2), Bei Chengxun (1)**

(1) Department of Physics, School of Science, South China University of Technology, Guangzhou, China
(2) State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, China

**PACS:** 43.55.Hy, 43.71.Gv

## ABSTRACT

The subjective Chinese speech intelligibility scores are obtained by using the simulated binaural room impulse responses based on auralization technique. The simulated binaural room impulse responses are first convolved with the Chinese phonetically balanced test word lists signals recorded in an anechoic chamber, then reproduced over headphone. The relationship between subjective Chinese speech intelligibility scores and the objective acoustical parameter useful-to-detrimental sound ratio is studied and analysed in simulated rooms. There is high correlation between Chinese speech intelligibility scores and the useful-to-detrimental sound ratio. The useful-to-detrimental sound ratio can evaluate and predict Chinese speech intelligibility in rooms.

## 1. INTRODUCTION

The Speech Intelligibility (SI) describes the quality of the signal the listener receives and is mainly dependent on the Direct-Reverberant-Ratio (influenced by the Reverberation Time) and the Signal-to-Noise-Ratio (influenced by the sources sound level and background noise). Houtgast and Steeneken [1, 2] developed Speech Transmission Index (STI) and its simplification Rapid Speech Transmission Index (RASTI) based on the Modulation Transfer Function. Speech Transmission Index (STI) combines both a room acoustics and a signal-to-noise ratio component into a single objective index for speech intelligibility in rooms. Using loudspeakers to reproduce the signals recorded in anechoic chamber, Wang et al. [3] studied relations between SI and definition, reverberation index, clarity in reverberation chamber, and got some significant conclusions. Lochner and Burger indicated that speech intelligibility was related to the integrated sound energy over the first 50 ms after the arrival of the direct sound and intelligibility is reduced by sound energy arriving later than 50 ms after the direct sound [4]. They also created a parameter. It called “useful to detrimental sound ratio” This combines early-late ratio and speech-to-noise ratio into a single quality to predict the speech intelligibility in rooms. “Useful” sounds are the integrated energy of speech sounds arriving within the first 50 or 80 milliseconds after the direct sound, and “detrimental” sounds are the sum of later-arriving speech energy and ambient noise. In practice, both quantities may be found by integrating appropriate portions of the room impulse response. Latham [5] evaluated the Lochner and Burger procedure in a number of theaters, and only with minor modifications found it to be a quite successful predictor of speech intelligibility. Bradley also further studied the useful-to-detrimental sound ratio concept and indicated that in real rooms a simple unweighted sum of the useful (early) speech energy worked as well as Lochner and Burger’s original proposals [6]. The useful-to-detrimental sound ratio can be calculated as follows [7]:

$$U_t = 10\log\left\{\frac{E/L_t}{1+(E/L_t+1)N/S}\right\}, dB \qquad (1)$$

Here, t equals 50 ms or 80 ms, E is the early energy, L the late energy, S the speech level and N the ambient noise. Bradley showed that the simplified version of the useful-to-detrimental sound ratio (equation 1) was found to be a good predictor of subjective speech intelligibility in rooms for English[7]. However, a study about the relationship between the useful-to-detrimental sound ratio and subjective Chinese Mandarin speech intelligibility has not been reported so far.

On the other hand, room acoustical computer simulation and auralization technique are widely used in room acoustics. It also provides an alternative method for subjective speech intelligibility assessment. The speech intelligibility evaluation based on auralization technique with simulated binaural room impulse responses (BRIRs) is in agreement with reality and results from measured BRIRs[8] and Auralize speech-intelligibility tests were found to be reliable [9].When test word lists signals recorded in anechoic chamber are convolved with the simulated BRIRs to perform speech intelligibility tests, the obtained articulation scores will be a very good approximation of the real ones [8].

In this article, the speech intelligibility evaluation based on auralization is employed. The binaural room impulse responses obtained from the room acoustical simulation software Odeon [10] are first convolved with the speech intelligibility test signals recorded in anechoic chamber, then reproduced through the headphone. The subjective Chinese speech intelligibility scores are obtained and the relationship between Chinese speech intelligibility scores and the useful to detrimental sound ratios are studied and analyzed. Goal is to built the relationship between Chinese speech intelligibil-

ity scores and useful to detrimental sound ratios and predict Chinese speech intelligibility in rooms.

## 2. EXPERIMENT METHOD

### 2.1 Rooms and sound fields

Four classrooms, three report halls and a church are modelled using room acoustical simulation software ODEON. In each model, a source and two or three receivers are set. In order to obtain a wide-range acoustical condition, 33 different listening positions are selected and simulated through adjusting materials on the rooms' surfaces. The BRIRs and objective acoustical parameters, such as early decay time (EDT), reverberation time ($T_{30}$), Definition (D,) Clarity ($C_{80}$) and STI at these positions are obtained from ODEON. Table 1 is the statistic of objective acoustical parameters at these listening positions. In present work, the useful-to-detrimental sound ratio with the 50ms and 80ms early time limit are studied. The $U_{50}$ and $U_{80}$ are calculated from D and $C_{80}$ according equation (2) and (3) under different signal to noise ratios (SNRs) conditions, respectively.

$$U_{50} = 10\log\left\{\frac{D}{1-D+10^{(-SNR/10)}}\right\} \quad (2)$$

$$U_{80} = 10\log\left\{\frac{10^{(C_{80}/10)}}{1+(1+10^{(C_{80}/10)})10^{(-SNR/10)}}\right\} \quad (3)$$

**Table 1**. The statistic of objective acoustical parameters at 33 listening positions

| | Min | Max | Average |
|---|---|---|---|
| EDT(500-1000Hz), s | 0.28 | 5.95 | 1.82 |
| T30(500-1000Hz), s | 0.36 | 5.16 | 1.79 |
| D(1000Hz) | 0.04 | 0.89 | 0.48 |
| $C_{80}$(1000Hz), dB | -11.4 | 14.0 | 2.43 |
| STI | 0.27 | 0.86 | 0.57 |

Based on the average speech spectrum from two male and two female speakers two speech shaped noise spectra are selected for use in the experiments. By using a speech-shaped noise with a frequency spectrum equivalent to the long-term speech spectrum, the SNR is equal for all selected octave band frequencies [11~13].

### 2.2 Speech intelligibility test

Chinese speech intelligibility scores are strongly dependent on the type subject-based speech test used. The rhyme test with five alternative forced choice based on initial consonants embedded in a carrier phrase is a good test and simple to apply for Chinese. However, the scores saturate easily and there is a ceiling effect for rhyme test [14, 15]. A phonetically balanced word test is more sensitive to changes in reverberation time and SNRs than the modified rhyme test [16] and is therefore more likely to highlight differences between test conditions. It should be noted that phonetically balanced intelligibility testing requires more training of listeners than other statistical tests. In this study, Chinese Mandarin phonetically balanced word lists as specified by GB 15508-1995[17] are used for Chinese speech intelligibility test for every testing condition. Every list consists of 25 three-syllables rows, total 75 syllables which have keep the same balance of the level of difficulty and phonemic characteristic. The three syllables in each row are randomly arranged and nonsense. The test words are embedded in the carrier phrase, "The –row is ×××". The "–" stands for row number and "×××" stands for three syllables selected from one list randomly and without repetition. All word lists are recorded at a rate of 4.0 words per second spoken by two male and two female speakers in an anechoic chamber.

The long-term frequency spectrum is different for male and female speech. Both the testing word lists signals recorded in anechoic chamber and speech-shaped noises are convolved respectively with simulated BRIRs obtained from ODEON after headphone equalization, then mixed according a given value for SNR and reproduced by headphone (Sennheiser HD580) at about 70 dBA level for each listening condition. During the test, a level adjustment is applied. The level estimation is based on the overall A-weighted RMS value and is corrected for the effect of silent periods by the application of a threshold [11~13]. A total of 136 different conditions are selected for test.

### 2.2 Subjects

The subjects are chosen from undergraduate students aged from 20-24 years old. They can speak and hear standard Mandarin Chinese and have no known hearing problems. All subjects are trained, familiarized with the test words lists and passed a pretest which required them to recognize test word in clean condition at an identification rate of at least 95%. For each listening conditions, two test word lists (one is man speaker, the other woman speaker) and four subjects are used. The subjects are asked to write down the key words which he or she heard. Irrespective of the grapheme, only if the vowel, consonant and tone are all correct the response is regarded as true. The subjective Chinese Mandarin intelligibility scores are averaged across these subjects.

## 3. RESULTS

Table 2 gives the multiple correlation coefficients (R) and associated standard deviations (SD) for third-order polynomial fits to measure in each octave band from 125 to 8000 Hz. To avoid one octave band resulting in a too high or too low correlation coefficient and standard deviation, the combination with more octave bands, such as 500, 1000, 2000 and 4000 Hz, are included. When multi-octave-bands are selected, the $U_{50}$ and $U_{80}$ values are simply arithmetic averaged from these octave bands [6]. From table 2, it can be seen that the multiple correlation coefficients gradually increase with the increase of octave band frequency and standard deviations decrease for both $U_{50}$ and $U_{80}$ when only one octave band is considered from 125 to 8000 Hz. The multiple correlation coefficient in 8000 Hz octave-band frequency is the highest and standard deviation the smallest among seven single octave band frequency. In general, the higher octave band frequency is, the more acoustical absorption and air absorption for sound energy occurs and the shorter the reverberation time is in rooms. The average values of $U_{50}$ and $U_{80}$ in low octave band frequency for all test conditions is smaller than those in low octave band frequency. These factors may lead to a higher multiple correlation coefficient and lower standard deviation with the increase of the octave-band frequency.

In the case more octave bands, such as 500-2000 Hz, 500-4000 Hz, 1000-4000 Hz and 125-8000 Hz, the multiple correlation coefficients have minor different, and all of them are greater than 0.8. When average $U_{50}$ and $U_{80}$ values are calculated from higher octave bands frequency, the multiple correlation coefficients are generally greater and standard deviations smaller. The octave bands included 1000, 2000 and 4000 Hz produced higher multiple correlation coefficients and lower standard deviations than those in other combined with multi-octave bands. Figure 1 and 2 show the curve of

the best third-order polynomial fit between Chinese speech intelligibility scores and $U_{50}$(1000-4000) and $U_{80}$(1000-4000), respectively. $U_{50}$(1000-4000) and $U_{80}$(1000-4000) are the useful-to-detrimental sound ratios with simply arithmetic averaged in 1000 to 4000 Hz octave bands for the 50 ms and 80 ms early time limit. There appears to be a high correlation between Chinese speech intelligibility scores and the useful to detrimental sound ratio. The multiple correlation coefficient is 0.86, 0.90 and standard deviation is 7.9, 6.7% for $U_{50}$ and $U_{80}$, respectively.

**Table 2**. The correlation coefficients and standard deviations for third-order polynomial fits in each octave band and combined octave bands

| Octave-band frequency, Hz | $U_{50}$ | | $U_{80}$ | |
|---|---|---|---|---|
| | R | SD, % | R | SD, % |
| 125 | 0.756 | 10.15 | 0.794 | 9.42 |
| 250 | 0.760 | 10.08 | 0.798 | 9.35 |
| 500 | 0.776 | 9.78 | 0.816 | 8.97 |
| 1000 | 0.795 | 9.42 | 0.839 | 8.43 |
| 2000 | 0.866 | 7.76 | 0.903 | 6.67 |
| 4000 | 0.889 | 7.09 | 0.911 | 6.41 |
| 8000 | 0.922 | 6.01 | 0.939 | 5.35 |
| 500-2000 | 0.815 | 8.98 | 0.858 | 8.00 |
| 500-4000 | 0.841 | 8.38 | 0.886 | 7.21 |
| 1000-4000 | 0.861 | 7.89 | 0.903 | 6.66 |
| 125-8000 | 0.849 | 8.21 | 0.894 | 6.94 |

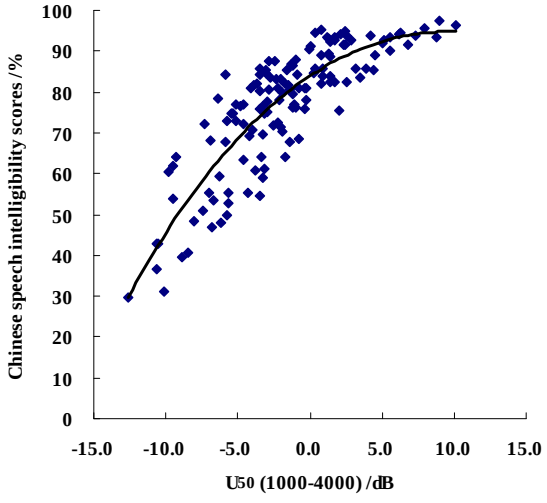


**Figure 1**. Subjective Chinese speech intelligibility scores versus $U_{50}$(1000-4000) values, best-fit third-order polynomial curve.

Attempts have been done to calculate $U_{50}$ and $U_{80}$ using the different octave band weighting factors, such as those used by STI [18] and SII [19] procedure, and proposed by Pavlovic [20], Marshall [21] and Zhang and Ma [22]. The weighting factors of each octave band frequency for different weighting methods are given in Table 3. Broadband $U_{50}$ and $U_{80}$ are calculated. Table 4 gives the multiple correlation coefficients and standard deviations for third-order polynomial fits between the subjective speech intelligibility scores and useful-to-detrimental sound ratio using different octave band weighting factors.

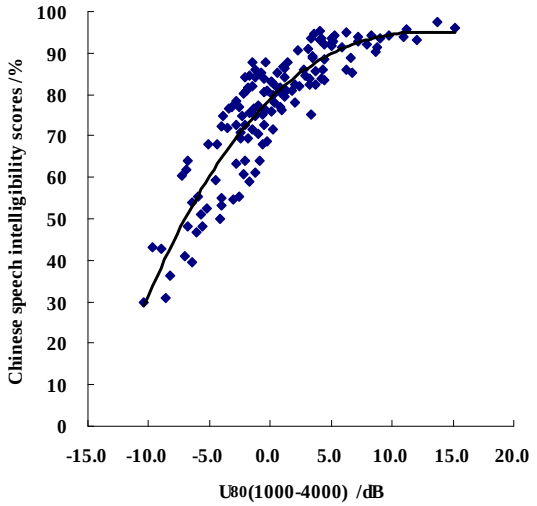


**Figure 2**. Subjective Chinese speech intelligibility scores versus $U_{80}$(1000-4000) values, best-fit third-order polynomial curve.

**Table 3**. Relative weighting factors for different octave bands

| Bands, Hz | Weighting factors | | | | | |
|---|---|---|---|---|---|---|
| | STI [18] | Pavlovic [20] | SII [19] | Marshall [21] | GB/T 15485 [23] | Zhang and Ma [22] |
| 125 | 0.129 | | | | | |
| 250 | 0.143 | 0.044 | 0.155 | | 0.072 | 0.035 |
| 500 | 0.114 | 0.129 | 0.156 | 0.15 | 0.144 | 0.105 |
| 1000 | 0.114 | 0.202 | 0.216 | 0.25 | 0.218 | 0.26 |
| 2000 | 0.186 | 0.312 | 0.277 | 0.35 | 0.327 | 0.325 |
| 4000 | 0.171 | 0.258 | 0.149 | 0.25 | 0.234 | 0.223 |
| 8000 | 0.143 | 0.055 | 0.047 | | | 0.052 |

**Table 4**. The correlation coefficients and standard deviations for third-order polynomial fits using the procedure with different octave band weights

| Octave-band Frequency, Hz | $U_{50}$ | | $U_{80}$ | |
|---|---|---|---|---|
| | R | SD, % | R | SD, % |
| STI | 0.856 | 8.03 | 0.901 | 6.74 |
| Pavlovic | 0.858 | 8.00 | 0.901 | 6.72 |
| SII | 0.837 | 8.50 | 0.882 | 7.31 |
| Marshall | 0.850 | 8.16 | 0.889 | 6.71 |
| GB/T 15485 | 0.845 | 8.30 | 0.889 | 7.11 |
| Zhang and Ma | 0.855 | 8.05 | 0.899 | 6.80 |

As can be seen from Table 4, the multiple correlation coefficients are generally similar in magnitude, and all of them are greater than 0.83 for $U_{50}$ and 0.88 for $U_{80}$. The standard deviations are also only slightly different for both $U_{50}$ and $U_{80}$. The multiple correlation coefficient obtained from STI, Pavlovic and Zhang and Mas' weights are higher and standard deviations are lower than those using other weighting methods. However, compared with the results from simply arithmetic averaged in 1000 to 4000 Hz octave bands, useful-to-detrimental sound ratio $U_{50}$ and $U_{80}$ obtained from weighting methods can not significantly improve the prediction of Chinese speech intelligibility in rooms. Useful-to-detrimental sound ratio using simply arithmetic averaged in the1000,

2000 and 4000 Hz three octave bands has the highest multiple correlation coefficient and the lowest standard deviation.

## 4. DISCUSSIONS

As mentioned above, $U_{50}$ and $U_{80}$ have been calculated from D and $C_{80}$ obtained from ODEON simulation according equation (2) and (3) using different signal to noise ratios, respectively. The SNR is assumed the same in each octave band frequency in equation (2) and (3). During the subjective testing, speech-shaped noise is used. Actually, the SNR for different octave bands may be different because the frequency spectrum of different test word list signals is not the same. The SNRs are only approximative estimate in equation (2) and (3). This may be another reason that the multiple correlation coefficients gradually increase with the increase of frequency and standard deviations decrease for both $U_{50}$ and $U_{80}$ when only one octave band is considered from 125 to 8000 Hz.

It can be seen from Table 2 and 4, there are some differences for $U_{50}$ and $U_{80}$ from only one octave band frequency, combination of multi-octave band frequencies and weights method. The multiple correlation coefficients are greater and standard deviations are lower for $U_{80}$ than those for $U_{50}$. This seems to show that useful-to-detrimental sound ratio for 80ms early time limit predicts Chinese Mandarin speech intelligibility more accurately than that for 50 ms early time limit from present work.

A comparison of the present result with Bradley's result [24] shows that there are some discrepancies. Figure 3 shows the relationship between speech intelligibility and $U_{50}$ in 1000 Hz octave band frequency for both Chinese and English. In Bradley's study, a English Fairbanks Rhyme Test was used [6]. In present work, a Chinese Mandarin Phonetically Balanced (PB) test is used. It can be seen from figure 3 that the curve of the relationship between speech intelligibility scores for English and $U_{50}$(1000 Hz) falls above the curve for the Chinese. The Mandarin Chinese is different from the western language, which is a kind of tone-language. The differences between two curves can be attributed to test method (Fairbanks Rhyme Test and PB Test) and language differences (accent and tone) between English and Chinese.

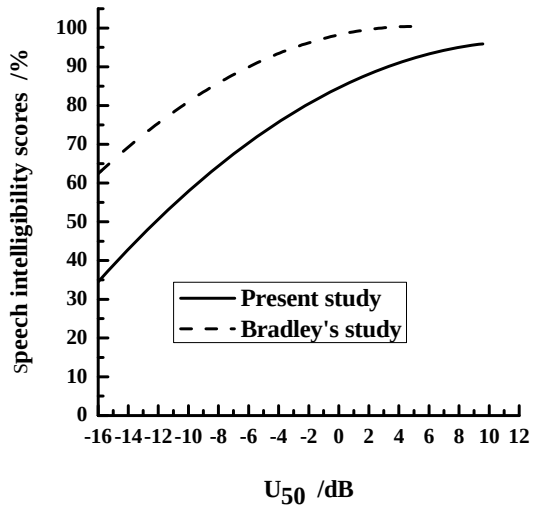


**Figure 3**. Relations between $U_{50}$(1000 Hz) and speech intelligibility for two types of speech intelligibility tests

## 5. CONCLUSIONS

The relationships between Chinese speech intelligibility scores and useful-to-detrimental sound ratios are studied based on simulated BRIRs through subjective test of speech intelligibility using auralization techniqe for Mandarin Chinese. The results show that there is high correlation between Chinese speech intelligibility scores and useful-to-detrimental sound ratio. The multiple correlation coefficient is greater and standard deviation lower for $U_{80}$ than that for $U_{50}$. Using simply arithmetic averaged in the three most critical octave bands for speech (1000, 2000 and 4000 Hz) is a good predictor for Chinese speech intelligibility.

## ACKNOWLEDGEMENTS

The authors thank all the students who participated in subjective evaluation of Chinese speech intelligibility. This work was support by National Natural Science Foundation of China (Grant No. 10774048) and the Opening Project of State Key Laboratory of Subtropical Building Science, South China University of Technology, China (Grant No. 2008KB32).

## REFERENCES

1 T. Houtgast and H. J. M.Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica. 28, 6–73 (1973).
2. H. J. M. Steeneken and T.Houtgast, "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. 67, 318-326(1980)..
3. J.Wang, L.Shao, "An experimental study of acoustic ratio and speech intelligibility," Proceedings of the 12th International Congress on Acoustics, Vol E, 10-14(1989)
4. J P A.Lochner, J F Burger, "The influence of reflections on auditorium acoustics," J. Sound Vib. **1**, 426-454(1964)
5. H G Latham, "The signal-to-noise ratio for speech intelligibility ---- an auditorium acoustics design index," Applied Acoustics, **18** , 252-320(1979)
6. J S Bradley, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am., **80**, 837-845(1986).
7. J. S. Bradley, "Relationships among measures of speech intelligibility in rooms," J. Audio Eng. Soc. **46**, 396–405 (1998).
8. Peng Jianxin, "Feasibility of subjective speech intelligibility assessment based on auralization," Applied Acoustics, **66**, 591-601(2005).
9. W.Yang, M.Hodgson, "Validation of the Auralization Technique: Comparative Speech Intelligibility Tests in Real and Virtual Classrooms," ACTA ACUSTICA UNITED WITH ACUSTICA, **93**, 991-999(2007).
10. C. L. Christensen, "Odeon Room Acoustics Program User Manual," Odeon A/S, Lyngby, Denmark, (2009).
11. H. J. M.Steeneken and T.Houtgast "Mutual dependence of the octave-band weights in predicting speech intelligibility," Speech Commun. **28,** 109-123(1999).
12. H. J. M. Steeneken and T. Houtgast, "Phoneme-group specific octave-band weights in predicting speech intelligibility," Speech Commun. **38**, 399-411(2002).
13. H. J. M.Steeneken and T. Houtgast, "Validation of the revised STIr method," Speech Commun. **38**, 413-425 (2002).
14. J. Peng "Relationship between Chinese speech intelligibility and speech transmission index using diotic listening," Speech Commun. **49,** 933-936(2007).
15. J. Peng "Relationship between Chinese speech intelligibility and speech transmission index in rooms using dichotic listening," Chinese Sci. Bull. **53**, 2748-2752(2008).

16. K.Kruger, K.Gough, and P. Hill, “A comparison of subjective speech intelligibility tests in reverberant environments,” Can. Acoust. **19,** 23–24(1991).
17. GB/T 15508. “Acoustics--Speech articulation testing method,” Standard of P. R. China, 1995.
18. IEC 60268-16 Ed. 3.0, “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index”. International Electrotechnical Commission, 2003
19. ANSI/ASA S3.5-1997 (R2007), “Methods for Calculation of the Speech Intelligibility Index,” American National Standard, 2007
20. C.V,Pavlovic, “Derivation of primary parameters and procedures for use in speech intelligibility predictions,” J. Acoust. Soc. Am., **82**, 413-422(1987)
21. L. Gerald Marshall, “An acoustics measurement program for evaluating auditoriums based on the early/late sound energy ratio,” J. Acoust. Soc. Am., **96**, 2251-2261(1994).
22. Y. T Zong, B.Bai and G. R,Sun “Application of modulation transfer function in communication system,” Audio Eng., **13**,1-5(1989).
23. GB/T 15485-1995， “Acoustics--Methods for the calculation of the articulation index of speech, ” Standard of P. R. China, 1993.
24. J. S. Bradley and S. R. Bistafa, “Relating speech intelligibility to useful-to- detrimental sound ratios”, J. Acoust. Soc. Am. **112**, 27–29 (2002).