PAN ARCHIVES OF ACOUSTICS
Vol. **39**, No. 4, pp. 483–488 (2014)

DOI: 10.2478/aoa-2014-0052

Chinese Syllable and Phoneme Identification in Noise and Reverberation

Jianxin PENG

Department of Physics, School of Science, South China University of Technology
Guangzhou 510640, China; e-mail: phjxpeng@163.com

(*received April 26, 2014; accepted August 13, 2014*)

Chinese is a tonal language, which differentiates it from non-tonal languages in the Western countries. A Chinese character consists of an initial, a final, and a tone. In the present study, the effects of noise and reverberation on the Chinese syllable, initial, final, and tone identification in rooms were investigated by using simulated binaural impulse responses through auralization method. The results show that the syllable identification score is the lowest, the tone identification score is the highest, and the initial identification scores are lower than those of the final identification under the same reverberation time and signal-to-noise ratio condition. The Chinese syllable, initial, and final identification scores increase with the increase of signal-to-noise ratio and decrease of the reverberation time. The noise and reverberation have insignificant effects on the Chinese tone identification scores under most room acoustical environments. The statistical relationship between the Chinese syllable articulation and phoneme articulation had been experimentally proved under different noise and reverberation conditions in simulated rooms.

Keywords: syllable identification, phoneme identification, noise, reverberation time, signal-to-noise ratio.

1. Introduction

Chinese is a tonal language, which differentiates it from non-tonal languages in the Western countries. Each Chinese character is equal to one syllable and generally consists of an initial, a final, and a tone. Initials consist of consonants or semi-vowels; finals consist of vowels or vowels plus one of the two nasal sounds such as [n] or [ng]. The tone is superimposed over the entire syllable. Mandarin has four pitched tones and a “toneless” tone. Zhang (1974) has investigated the phonetic regularity and derived the expressions for the relation between syllable and phoneme articulation scores.

$$S = \frac{L^2}{L^2 + q_{12}\cdot 2L(1-L) + Lq_{12}\cdot (1-L)^2}, \qquad (1)$$

$$L = (I+F)/2, \qquad (2)$$

where q_{12} equals 0.87, L is the phoneme articulation scores. I, F, and S are initial, final, and syllable articulation scores, respectively. The above relationship has been validated by subjective Chinese speech intelligibility evaluation under different sound pressure levels at receiving positions, signal to noise ratios (SNRs), and filter conditions. However, normal speech communication often occurs in environments with noise and reverberation. Many researchers investigated the consonant and vowel recognition in noise, reverberation, and their combined conditions, and the effect of noise and reverberation on vowel and consonant recognition for Western languages (Helfer, 1994; Peissig, Kollmeier, 1997; Paglialonga *et al.*, 2011; Meyer *et al.*, 2013; Ozimek *et al.*, 2013). For Chinese, McLoughlin (2010) has evaluated Chinese vowel intelligibility using reproduced test signals in a noise free environment through a headphone. Liu *et al.* (2010) have investigated the relationship between Chinese syllable identification and SNR by using solely noise masking in laboratory. Peng (2005a; 2005b) has studied the effects of different types of noise on Chinese speech intelligibility using similar rhythm test lists under different noise and reverberation conditions. Kang (1998), Peng (2010) and Peng *et al.* (2011) have investigated the Chinese speech intelligibility under different objective acoustical conditions. At the same period of time, some researchers have explored the characteristics of Chinese speech perception under different acoustical environments (Zhang *et al.*, 1981; Zhang, Meng, 2013). However, from the best knowledge of the authors, there is still a lack of publications about the effects of the room acoustical factors (i.e., noise and reverberation) on the Chinese initial, final, and tone identification.

In the present study, 7 binaural room impulse responses (BRIRs) with different reverberation times (RTs) were obtained by acoustical simulation from 5 rooms with various dimensions. Chinese syllable, initial, final, and tone identification was conducted by using the auralization method under different SNR and reverberation-time conditions. The aim of the present study is to investigate the effects of noise and reverberation on Chinese syllable, initial, final, and tone identification. In addition, the relationship between the syllable and phoneme identification scores has been validated.

2. Methods

Two classrooms, two auditoriums, and a church have been modeled using a software of ODEON. The rooms' geometries and dimensions are shown in Table 1. In each model, a vocal source was located at the middle-front of the room and a receiver was located at the middle-back of the room. The sound absorbing and scattering coefficients of the materials were assigned according to the materials on the surfaces in the actual rooms. The sound absorption coefficients of the materials on some surfaces in two classrooms were adjusted (i.e., increase of sound absorption coefficients) to obtain a shorter RT. In total, seven reverberation conditions were simulated. The average RTs in mid-frequency 500 Hz and 1000 Hz octave bands, average definition D_{50} in 500 Hz∼4000 Hz octave bands, and speech transmission index (STI) at the receiver position in each room are shown in Table 1. The BRIRs at the receiver positions were also obtained from the acoustical simulation in these simulated rooms.

Table 1. Dimensions of room models and the objective acoustical parameters at the receiver's position.

Room model	Shape	Dimension [m^3]	RT [s]	D_{50}	STI
Classroom	rectangle	8.0×8.4×5.1	0.4	0.90	0.84
Classroom	rectangle	8.0×8.4×5.1	1.8	0.34	0.47
Classroom	rectangle	16.0×8.4×5.1	1.4	0.43	0.53
Classroom	rectangle	16.0×8.4×5.1	2.2	0.30	0.43
Auditorium	trapezoid	19.0×18.6×5.0	0.6	0.79	0.73
Auditorium	trapezoid	30.8×24.0×6.0	1.0	0.58	0.63
Church	rectangle	64.5×31.2×20.7	5.0	0.21	0.34

The Mandarin phonetically balanced word lists as specified by GB 15508-1995 were used for Chinese syllable, initial, final, and tone identification tests for all testing conditions. Each list consists of 25 three-syllable rows and total 75 syllables. The syllables were designed with consideration of balance of the level of difficulty and phonemic characteristics. The three syllables in each row were randomly arranged without related meaning and construction of combined meaningful words. The test characters were embedded in the carrier phrases: "The – row is × × ×", where "–" denotes the row number and "× × ×" stands for the three syllables randomly selected from the lists. Syllables did not repeat in the tests. All speech signals were recorded at a rate of about 4.0 words per second with two male and two female speakers in an anechoic chamber. The recording was edited by CoolEdit Pro. Ten-second intervals of silence between two adjacent carrier sentences for the listeners to write down their solution had been inserted. The initial, final, and tone identification scores are the percentage of the initials, finals, and tones that are correctly identified in each list by the subjects, respectively. The initial, final, and tone in a syllable correctly identified by the subjects were regarded as a correct syllable. The syllable identification score is the percentage of the correct syllable in each list.

Based on the average speech spectrum from two male and two female speakers, two speech-shaped noise spectra have been developed for the use in these experiments. By using a speech-shaped noise with a frequency spectrum equivalent to the long-term speech spectrum, the SNR was balanced for all selected frequency bands (Steeneken, Houtgast, 2002). Both the testing word lists signals and speech-shaped noises were convolved with simulated BRIRs obtained from ODEON after headphone equalization, respectively. Afterwards, the signals were mixed with various SNR (−5 dB, 0 dB, 5 dB, 10 dB, 15 dB, and a noiseless case) and reproduced by a headphone (Sennheiser HD580). During the test, level adjustment was applied. The level estimation was based on the overall A-weighted RMS value and corrected for the effect of silent periods by the application of a threshold (Steeneken, Houtgast, 2002). Previous studies indicated that a 70 dBA speech level was the suitable hearing level for speech intelligibility test (Brachmanski, 2002; Peng, 2010). In the present study, the level of speech signals was about 70dBA at the ears of the listeners for all tests.

The subjects were undergraduate students from 19 to 24 years old. They could speak native Mandarin and had a normal hearing ability. All the subjects were trained and have passed a pretest that requires them to recognize spoken words in a quiet condition with at least 95% accuracy. Seven different reverberation times and six different SNR conditions were applied. The total of 42 test conditions were evaluated. For each testing condition, two speech signals (one male and one female voice) were presented. The subjects were asked to write down the spellings of the key words that they heard. The averages of the subjective Chinese syllable, initial, final, and tone identification scores across eight test word lists were determined accordingly for each test condition.

3. Results and discussions

Figure 1 shows the subjective syllable, initial, final, and tone identification scores under different RT and SNR conditions. It can be seen from Fig. 1 that the syllable identification score is the lowest, the tone identification score is the highest, and the initial identification scores are lower than the final identification scores under the same reverberation and SNR conditions. The syllable, initial, and final identification scores increase with an increase of SNR and a decrease of RT, except for the tone identification scores. The syllable, initial, and final identification scores increase with an increase of SNR under the same RT conditions, or decrease with an increase of RT under the same SNR conditions. The syllable, initial, and final identification scores are lower under longer RT and (/or) lower SNR conditions. The tone identification scores are less than 90% correct with only 8 out of the 42 test conditions.

The analysis of variance has been conducted to compare the influence of RT and SNR on syllable, initial, final, and tone identification scores under the 42 test conditions. The results are given in Table 2. It can be seen that both RT and SNR at the receiver's

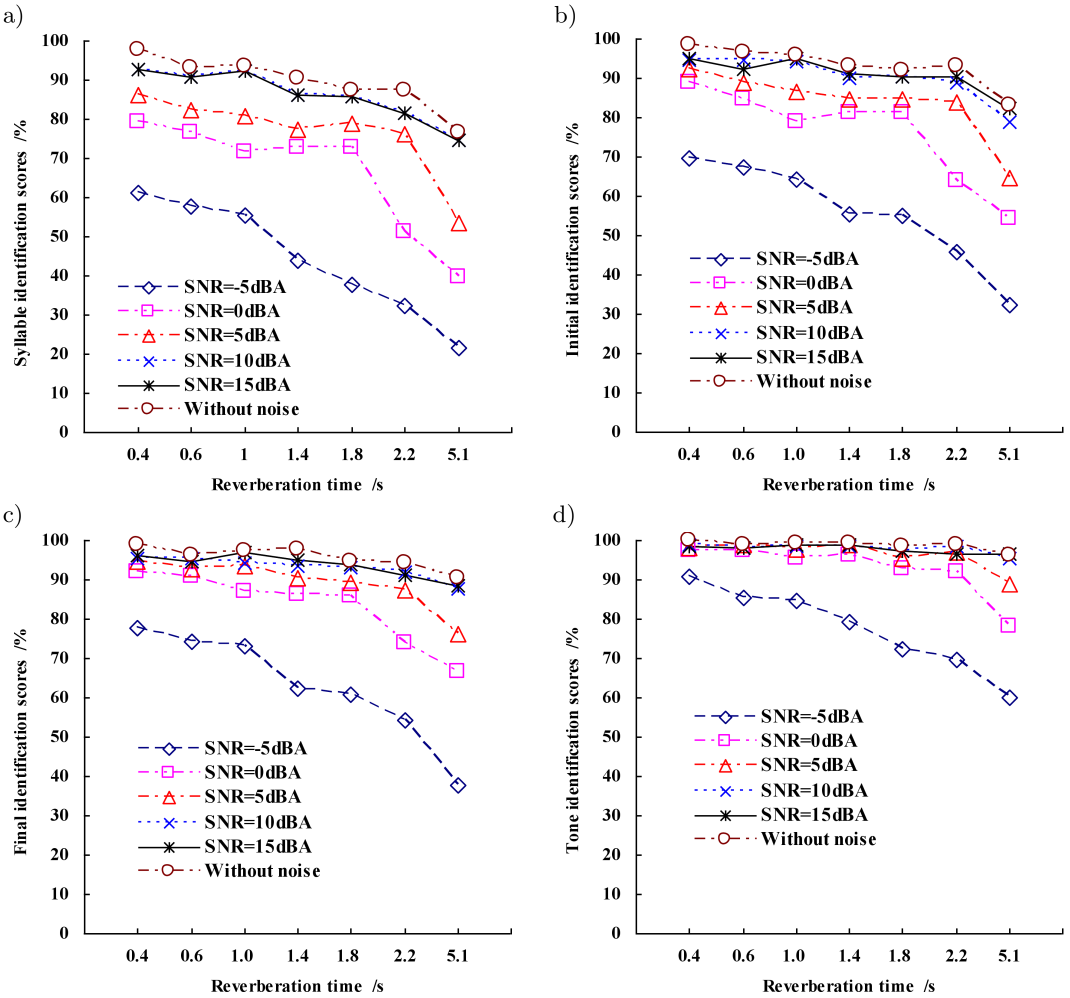

Fig. 1. Syllable, initial, final, and tone identification scores under different reverberation time and SNR conditions a) syllable, b) initial, c) final, d) tone.

Table 2. Analysis of variance for the syllable, initial, final, and tone identification scores under different reverberation time and SNR conditions.

Factors	df	F(syllable)	F(initial)	F(final)	F(tone)
SNR	5	320.1 ($p < 0.001$)	327.6 ($p < 0.001$)	257.3 ($p < 0.001$)	131.1 ($p < 0.001$)
RT	6	87.3 ($p < 0.001$)	89.1 ($p < 0.001$)	56.5 ($p < 0.001$)	25.5 ($p < 0.001$)
SNR * RT	30	4.1 ($p < 0.001$)	4.7 ($p < 0.001$)	5.5 ($p < 0.001$)	4.7 ($p < 0.001$)

position have significant effects on syllable, initial, final, and tone identification scores ($p < 0.001$). F value of the SNR is more than that of the reverberation time. This shows that the SNR has a more significant effect on the syllable, initial, final, and tone identification scores than RT. There is a significant interaction between SNR and RT for the identification scores ($p < 0.001$) but the F value of the interaction is smaller than that of RT and SNR individually. SNR has a more significant impact on the syllable identification scores than the initial, final, and tone identification scores, while RT has a more significant impact on the initial identification scores than the syllable, final, and tone identification scores.

Since a syllable is accepted as correct only when all of the initial, final, and tone in a syllable were correctly identified by the subjects, it is expected that the syllable identification scores are lower than the initial, final, and tone identification scores. It can be seen from Figure 1 that the Chinese initial identification score is lower than the final identification score under the same test conditions. In this study, the interfered noise is a speech-shape noise which had a similar spectrum with the speech test signals. The masking effect of the speech-shape noise on the Chinese initial and final mainly is a energetic masking but the energy of the Chinese initial is lower than that of the final. Therefore, the Chinese initial can be more easily masked by noise as compared to the final. On the other hand, the initial of a syllable can be masked by the final of the prior syllable, which smears the speech signals due to the reverberation effect. Syllable identification depends more on the initial than on the finial. Moreover, in the time domain, the reverberation smears the time gap between syllables, which also reduces the speech intelligibility (Tillery *et al.*, 2012). When reverberation and noise are combined, especially with long RT, reverberation does not only strongly mask the initial, final, and tone, but it also enhances the noise level that increases the masking effect of noise on the initial, final, and tone.

In Fig. 1, when the RT is 5.0 s with the SNR 0dB and 5dB, and the SNR is −5 dB except for RT 0.4 s, the tone identification scores are less than 90%. This is due to the fact that speech signal frequency, amplitude envelope in the time domain, envelope information, and the fine signal structure in the time and frequency domains all had the effect on Chinese tone identification (Kong, Zeng, 2006). In quiet conditions, subjects with normal hearing may be able to recognize tones only with the fine signal structure in the time and frequency domains. In noise conditions, even though the envelope information is more sensitive to noise than other factors for the tone perception of speech, listeners can identify a tone by using its fine signal structure (Kong, Zeng, 2006). However, for the combination of reverberation and noise, the fine signal structure of the tone is damaged, so that the tone identification score is reduced, especially under the long reverberation time conditions. Therefore, noise and reverberation have effects on the tone identification score under the longer RT and lower SNR, and the score is less than 90% correct. However, they have an insignificant effect on the tonal identification score under other conditions. These results were also consistent with the results by Ma and Shen (2004).

The phoneme identification scores can be calculated from the initial and final identification scores according to Eq. (2) under different reverberation and SNR conditions. Figure 2 shows the relationship between the measured syllable identification scores and calculated phoneme identification scores from Eq. (2) under different RT and SNR conditions. The statistical curve described by Eq. (1) (Zhang, 1974) is also plotted in Fig. 2. It can be seen from Fig. 2 that the relationship between the syllable and phoneme identification scores generally matches the results of the theoretical curves from Zhang (1974).

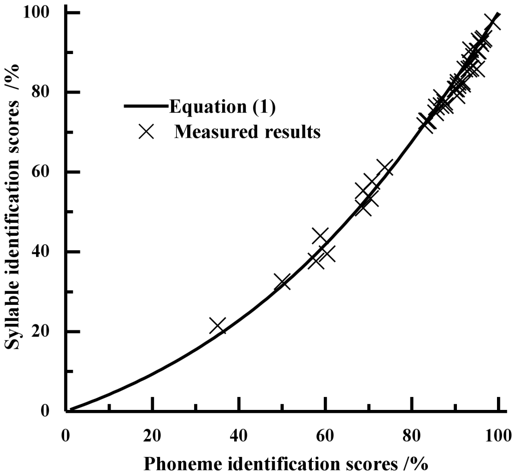

Fig. 2. Relationship between the syllable and phoneme identification under different reverberation and SNR conditions.

Figure 3 shows the fitting line between syllable identification scores under different noise and reverberation conditions and the calculated syllable identification score from Eq. (1). The correlation coefficient was 0.99 and the standard deviation is 2.4%. The fitting equation is:

$$\mathrm{SI} = \mathrm{SI}' - 0.9, \tag{3}$$

where SI is the Chinese syllable identification score measured for different noise and reverberation conditions in this study, SI′ is the Chinese syllable articulation score calculated from Zhang (1974) by using Eq. (1). Figure 2 and Fig. 3 show the measured Chinese syllable identification scores accorded with the predicted Chinese syllable articulation scores from Eq. (1). Equation (1) can be also used for predicting

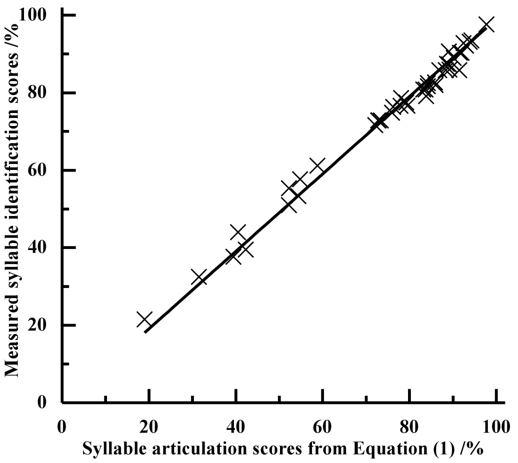

Fig. 3. Fitting line between syllable identification scores under different noise and reverberation conditions and the calculated syllable articulation scores from Eq. (1).

the syllable articulation in rooms under noise and reverberation conditions.

4. Conclusions

Seven BRIRs with different RTs were obtained from room acoustical simulations. The Chinese syllable, initial, final, and tone identifications have been evaluated by using the auralization method under different RT and SNR conditions. The effects of noise and reverberation on Chinese syllable, initial, final, and tone identifications have been investigated. The results show that the syllable identification score is the lowest among those scores, while the tone identification score is the highest, the initial identification scores are lower than the final identification scores under the same RT and SNR conditions. The syllable, initial, and final identification scores increase with an increase of SNR and with a decrease of RT except for the tone identification score. The noise and reverberation have insignificant effects on the Chinese tone identification scores under most of the test conditions. Only when the RT is 5.0 s with the SNR 0 dB and 5 dB, and the SNR is -5 dB except for RT 0.4 s, the tone identification scores are less than 90%. The relationship between the syllable and phoneme articulation scores described by Eq. (1) in Zhang (1974) has been validated by the experimental results in simulated rooms under different noise and reverberation conditions.

Acknowledgments

The authors express their gratitude to the students who participated in subjective evaluation of Chinese speech identification. Special thanks are given to Dr. Siu-Kit Lau for his helpful suggestions and for his kind revising of the manuscript. This work is supported by the National Natural Science Foundation of China (Grant No. 10774048, 11374106).

References

1. Brachmanski S. (2002), *Effect of additive interference on speech transmission*, Archives of Acoustics, **27**, 95–108.
2. Helfer K.S. (1994), *Binaural cues and consonant perception in reverberation and Noise*, Journal of Speech & Hearing Research, **37**, 429–438.
3. Kong Y.Y., Zeng F.G. (2006), *Temporal and spectral cues in Mandarin tone recognition*, J. Acoust. Soc. Am., **120**, 2830–2840.
4. Liu H., Zhang S.Y., Meng Z.H. (2010), *Test on Mandarin Monosyllable Clarity and Speech Intelligibility Estimation With Low SNR*, Audio Engineering, **34**, 60–3.
5. Mao D.Y., Shen H. (2004), *Handbook of acoustics*, Science press, Beijing.
6. McLoughlin I. (2010), *Vowel intelligibility in Chinese*, IEEE Transactions on Journal Audio, Speech, and Language Processing, **18**, 117–125.
7. Meyer J., Dentel L., Meunier F. (2013), *Speech recognition in natural background noise*, PLoS One, **8**, e79279.
8. Kang J. (1998), *Comparison of speech intelligibility between English and Chinese*, J. Acoust. Soc. Am, **103**, 1213–1216.
9. Ozimek E., Kociński J., Kutzner D., Sęk A., Wicher A. (2013), *Speech intelligibility for different spatial configurations of target speech and competing noise source in a horizontal and median plane*, Speech Communication, **55**, 1021–1032.
10. Paglialonga A., Tognola G., Grandori F. (2011), *SUN-test (Speech Understanding in Noise): a method for hearing disability screening*, Audiology Research, **1**, e13.
11. Peissig J., Kollmeier B. (1997), *Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners*, Journal of the Acoustical Society of America, **101**, 1660–1670.
12. Peng J.X. (2005a), *Effects of different kinds of noise sources on Chinese speech intelligibility*, Journal of Vibration and Shock, **24**, 98–101.
13. Peng J.X. (2005b), *Study of Chinese Speech Intelligibility under Noise of Chinese Average Frequency Spectrum Condition*, Journal of south China University of Technology, **33**, 71–74.
14. Peng J.X. (2008), *Relationship between Chinese speech intelligibility and speech transmission index in rooms using dichotic listening*, Chinese Science Bulletin, **53**, 2748–2752.
15. Peng J.X. (2010), *Chinese speech intelligibility at different speech sound pressure levels and signal-to-noise*

ratios in simulated classrooms, Appl. Acoust., **71**, 386–390.

16. Peng J.X., Bei C.X., Sun H.T. (2011), *Relationship between Chinese speech intelligibility and speech transmission index in rooms based on auralization*, Speech Communication, **53**, 986–990.

17. Steeneken H.J.M., Houtgast T. (2002), *Phoneme-group specific octave-band weights in predicting speech intelligibility*, Speech Commun., **38**, 399–411.

18. Tillery K.H., Brown C.A., Bacon S.P. (2012), *Comparing the effects of reverberation and of noise on speech recognition in simulated electric-acoustic listening*, J. Acoust. Soc. Am., **131**, 416–423.

19. Zhang J.L. (1974), *On the statistical relation between the syllable articulation and the phoneme articulation*, Acta Phys. Sin., **23**, 315–320.

20. Zhang J.L., Qi S.J., Lv S.N. (1981), *A preliminary study of the perceptual configurations of Chinese consonants*, Acta Psychologica Sinica, **13**, 78–87.

21. Zhang S.Y., Meng Z.H. (2013), *The experimental analysis on perceptual features of putonghua with reverberation*, Acta Acustica, **38**, 85–91.