Applied Acoustics 104 (2016) 50–56

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical Note

A study on the optimal English speech level for Chinese listeners in classrooms

Ming Qin, Xuhao Du, Jiancheng Tao *, Xiaojun Qiu

Key Laboratory of Modern Acoustics (MOE), Institute of Acoustics, Nanjing University, Nanjing 210093, China

ARTICLE INFO

Article history:
Received 6 June 2015
Received in revised form 30 August 2015
Accepted 15 October 2015
Available online 18 November 2015

Keywords:
Speech level
Chinese listeners
Signal to noise ratio
Reverberation time

ABSTRACT

Speech intelligibility in classrooms affects the learning efficiency of students directly, especially for the students who are using a second language. The speech intelligibility value is determined by many factors such as speech level, signal to noise ratio, and reverberation time in the rooms. This paper investigates the contributions of these factors with subjective tests, especially speech level, which is required for designing the optimal gain for sound amplification systems in classrooms. The test material was generated by mixing the convolution output of the English Coordinate Response Measure corpus and the room impulse responses with the background noise. The subjects are all Chinese students who use English as a second language. It is found that the speech intelligibility increases first and then decreases with the increase of speech level, and the optimal English speech level is about 71 dBA in classrooms for Chinese listeners when the signal to noise ratio and the reverberation time keep constant. Finally, a regression equation is proposed to predict the speech intelligibility based on speech level, signal to noise ratio, and reverberation time.

1. Introduction

Speech intelligibility (SI) is defined as the measure of the comprehensible quality of speech, which can be evaluated by the percentage of correctly understood words or sentences in a specific list under controlled conditions [1]. SI can be used to quantify the speech perception in classrooms, which directly affects the learning efficiency of students, especially younger students and the students who are working on a second language [1–4]. Poor perception in classrooms has been reported to cause perceptual and cognitive problems among students [1]. With a number of linguistic and phonetic reasons, the perception results of native and non-native listeners might differ with each other [5]. It has been indicated that SI is affected by a number of acoustic parameters such as signal to noise ratio (SNR), speech level (L_s), and reverberation time (T_{60}) [2,6–27].

L_s affects the perception of listeners directly, especially in noisy environment. It was shown that in the 70% occupied classrooms, the L_s of lecturers varied from 46 to 51 dBA in the front, and 42 to 47 dBA at the back [2]. As early as in 1922, it was found that the syllable articulation decreases with L_s when L_s is relatively high or low [6]. Later in 1946, it was illustrated that SI increases with L_s at different rates, depending on the background noise level (L_{BN}) and the SNR of the environment [7]. In 1956, it was also found that SI deteriorates with extremely weak or strong vocal forces [8], and SI decreases when L_s increases from 80 dB regardless to the variance of SNR [9]. It was also found that both the SI of the low-frequency speech in low-frequency maskers and the SI of the high-frequency speech in high-frequency maskers decrease with the increase of L_s when SNR keeps constant [10]. In recent studies, SI and listening difficulty ratings were used simultaneously to found the optimum L_s or acceptable range of L_s [11,12].

SNR is defined as the difference between L_s and L_{BN}, where the background noise in classrooms is mainly contributed by ventilation systems, students' activities, and noise sources outside classrooms [13]. It has been demonstrated that SI increases with SNR, and the detrimental effect of interfering noise can be eliminated if SNR is larger than 15 dB in classrooms [13–15]. Considering both L_s and SNR, the criteria for the maximum allowed L_{BN} in different kinds of classrooms have been proposed [16–21]. For instance, according to the criterion proposed by American National Standard Institute (ANSI), L_{BN} should not exceed 35 dBA in an unoccupied classroom smaller than 566 m^3 [19].

T_{60} depends on the room geometry and the amount of absorption in the room. SI is usually negatively related to T_{60}, and the optimal T_{60} for SI should be closed to 0 s [22–24]. The maximum

* Corresponding author.
E-mail address: jctao@nju.edu.cn (J. Tao).

http://dx.doi.org/10.1016/j.apacoust.2015.10.017
0003-682X/

allowed T_{60} for different kinds of classrooms has been proposed in some standards and acoustic design guides [19–21]. For example, according to the criterion of ANSI, T_{60} should not exceed 0.6 s in an unoccupied classroom that is smaller than 566 m^3 [19].

Besides L_s, SNR, and T_{60}, other acoustic parameters such as useful to detrimental sound ratio (UDR), speech transmission index (STI), the Definition (*D*), and the Clarity (*C*) have also been used to predict SI in rooms because they are positively related to SNR but negatively related to T_{60} [2,13,15,19,25–27].

It has been found that SI is affected by some acoustic parameters simultaneously. For instance, it is demonstrated that SI increases with L_s at different rates when L_{BN} changes [7,9,11,12,28], and the relationship between SI and SNR also changes when T_{60} varies [13,17,22,28]. This paper will investigate the prediction of the SI in classrooms considering L_s, SNR and T_{60} together when Chinese listeners are listening English. The methodology used in the subjective test will be described first, and then the relationships between the SI and L_s, SNR and T_{60} will be analyzed separately. Finally, a regression equation involving all of the parameters will be proposed for predicting SI.

2. Methodology

The SI test was carried out with professional headphones (Type: AKG 518). The test material was generated by convoluting the sections of Coordinate Response Measure (CRM) corpus with the measured room impulse responses and then mixing them with the background noise. In the test, the listeners were asked to listen to the test material and answer the questions on what they had heard. After the test was completed, the SI results were obtained by collecting and grading the answer sheets.

2.1. Processing of the test material

The widely used CRM corpus is adopted to generate the test material, which consists of 256 short phrases in English, and each phrase is made up of 3 parts: a signal, a color and a single number [29]. There are 8 English signals (arrow, baron, charlie, eagle, hopper, laker, ringo and tiger), 4 colors (blue, green, red and white) and 8 single numbers (1–8) in total. With different combinations, 256 different short phrases were made following the structure: "ready (a signal) go to (a color) (a number)". These phrases were spoken by 8 native English speakers (4 males and 4 females) and recorded in an anechoic chamber.

The test material was made up of 64 different sections, and each section was constituted by 20 phrases which were randomly chosen from the CRM corpus. There is a 2.5-s time interval between adjacent phrases and a 5-s pause between adjacent sections. The L_s, the SNR, and the T_{60} of each section were processed to be different from each other, and there were 4 different L_s (50, 60, 70 and 80 dBA), 4 different SNRs (0, 5, 15 and 20 dB) and 4 different reverberation conditions (T_{60} = 0.77, 1.07, 1.29 and 1.52 s) in total. In order to prevent listeners from adapting, the order of the 64 sections was randomly arranged. Since the whole test material lasts about 96 min, it was divided into 4 parts with equal lengths. During the listening tests, each volunteer only takes one part. Because the order of the 64 sections was randomly arranged, dividing them into 4 parts did not introduce new errors.

The reverberation effect of each processed section was implemented by convoluting the standard CRM section with the room impulse responses measured in a real room with dimensions 6.0 m × 6.6 m × 4.5 m. As shown in Fig. 1, a loudspeaker was placed at the corner of the room to excite as many modes as possible, and the measurement microphone (Type: BAST 100291) was placed 5.7 m away from the loudspeaker (about 1.5 m, 1.4 m and 3.2 m away from the walls). In the measurements, a B&K Pulse 3560D system generated pink noise to drive the loudspeaker and recorded the data with a sampling frequency of 65,536 Hz. Each impulse response was measured three times and averaged, and then the corresponding T_{60} was calculated from them. The T_{60} values used in the tests were the average values of 500 Hz, 1000 Hz and 2000 Hz octaves.

By changing the amount of sound absorption panels and foams decorated in the room, the impulse response and the corresponding T_{60} of the room were adjusted. A previous survey on 26 different classrooms at Xianlin Campus of Nanjing University showed that the T_{60} varies from 0.5 to 1.6 s, and the L_{BN} varies from 30 to 40 dBA when the classrooms are unoccupied [30]. Therefore, the T_{60} was controlled to vary from 0.77 to 1.52 s in the measurements.

The CRM sections with reverberation effects were then mixed by using the *Adobe Audition* software with the typical noise measured in a classroom at Xianlin Campus of Nanjing University during a class break. The L_s and SNR of the playback sound were adjusted by changing the amplitudes of the input signals of the CRM sections and the typical noise and monitored by using a B&K Head and Torso Simulator (HATS, Type: 4128).

2.2. Listening test

Sixty normal-hearing volunteers with English as a second language (24 females, 36 males, aged between 20 and 24) participated the listening test. All listeners had passed the College English Test Band 4, which is an authoritative English test in China. The 60 volunteers were randomly divided into 4 groups to listen to the 4 parts of the test material, respectively, so there were 15 complete samples after all volunteers completed the test.

To help the volunteers to get familiar with the test procedure, they were asked to listen to a standard section of CRM corpus before the formal test. This standard section doesn't contain any background noise or reverberation information and the sound pressure level is adjusted to 70 dBA during the playback. During the test, the volunteers were asked to choose the color and the number that they had heard. The L_{BN} in the test room was lower than 35 dBA, which was at least 15 dB lower than the playback sound pressure level. So the background noise effect of the playback environment can be neglected. After the volunteers completed the test, the answer sheets were collected and graded.

3. Data analysis and discussion

The percentages of correct answers (SI) of the 15 samples are averaged and presented in Table 1 with the corresponding standard deviations (SD) in Table 2. Compared with the results in Refs. [17,22], the perception accuracy shown in Table 1 is a little lower, while the corresponding SDs in Table 2 are larger. The reason is that the volunteers who participated the test are non-native listeners, whose perception accuracy might be poorer than native listeners under the same acoustic conditions [3,4]. Previous studies have found that the phonological factors such as some features of the first language and the phonetic characteristics of some sound of the second language may significantly reduce the speech intelligibility and increase listening difficulty of non-native listeners [3,5,31,32].

In the following sections, the contributions of the 3 major factors (T_{60}, SNR, and L_s) on SI are analyzed individually first, and then their combined effect is considered. In the following figures, the data points of SI are the average values, and the error bars represent the associated SDs.