https://doi.org/10.1016/j.buildenv.2022.109443

A laboratory study correlating serial recall performance and speech intelligibility of Chinese language in open-plan offices

Shengxian Kang [a], Cheuk Ming Mak [a]*, Dayi Ou [b,c,d] and Yuanyuan Zhang [b,c,d]

[a] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[b] School of Architecture, Huaqiao University, Xiamen, 361021, P.R. China

[c] Xiamen Key Laboratory of Ecological Building Construction, Xiamen, 361021, P.R. China

[d] State Key Laboratory of Subtropical Architecture Science, South China University of Technology, Guangzhou 510640, P.R. China

*Corresponding author: Tel.: +852 2766 5856; fax: +852 2765 7198.
E-mail address: cheuk-ming.mak@polyu.edu.hk (C.M. Mak).

Abstract

Speech noise can reduce occupants' work performance in open-plan offices. Some models have been created to predict the effect of speech of different intelligibility on work performance. However, few of them consider the effects of speech intelligibility in Chinese environments. This study aims to develop a model that evaluates how much work performance is decreased by speech noise with different intelligibility in Chinese open-plan offices. A laboratory experiment has been conducted in this paper to determine the effects of different speech intelligibility on occupants' objective performance of the serial recall task and perceived speech disturbance in Chinese open-plan offices. Then, a prediction model was developed by analyzing the data from this experiment and two previous studies. These two studies researched the effects of the Speech Transmission Index (STI, an objective parameter of speech intelligibility) on the serial recall performance in Chinese environments. According to the prediction model of serial recall performance in Chinese environments, performance decrease occurs within the STI range of 0.31-0.47. The comparison of curves between STI and DP with previous studies shows that the STI range for serial recall performance variation in Chinese environments is narrower than in non-Chinese language environments. Furthermore, the DP average change rate of serial recall tasks in Chinese environments is not less compared to non-Chinese environments, although the effect of speech noise on serial recall performance is lower in the Chinese environment.

Keywords: Open-plan Offices, Language Environments, Serial Recall Performance, Speech Transmission Index, Prediction Model

1. Introduction

The open-plan office is a typical office type in office buildings. Flexible layout and high economic efficiency make the open-plan offices popular [1, 2]. Large office space without partition walls is the main feature of open-plan offices. It allows occupants to work together and communicate conveniently, but also enables noise to transmit throughout the office with few obstacles. Limiting the noise disturbance between workstations or teams is viewed as one of the major acoustic challenges for all types of open-plan offices, according to ISO 22955:2021 [3]. Conversation noise (or speech noise), one of the most annoying noises in open-plan offices [4-6], has a strong effect on occupants' work performance [4, 7, 8], perception of acoustic environments [6, 9] and mental health [10-12]. The high speech intelligibility of speech noise is one of the main reasons causing the decrease in work performance [13-15] and perceived acoustic dissatisfaction [16-18] in open-plan offices.

1.1 Effects of speech intelligibility on work performance in open-plan offices

A large number of studies examined the effects of speech intelligibility on occupants' work performance [19-21]. The Speech Transmission Index (STI), an objective parameter of speech intelligibility, has been frequently used in previous studies to research the relationship between speech intelligibility and work performance [8, 14, 22]. For instance, Haka et al. [19] examined how STI conditions (STI=0.10, 0.35 and 0.65) affect occupants' work performance in Finnish environments. Jahncke et al. [20] tested how much work performance is reduced by speech noise with five levels of speech intelligibility (STI = 0.08, 0.16, 0.23, 0.34, and 0.71) in the Swedish environment. Renz et al. [23] conducted a laboratory experiment to study the effects of eight STI conditions (STI =0.10 to 0.33) on occupants' work performance and perceived sound annoyance in the German environment by changing the speech-to-noise ratio (SNR) and masking sounds. Ebissou et al. [24] tested occupants' work performance and mental workload in four STI conditions (STI = 0.25, 0.35, 0.45, and 0.65) in the French environment. Lou and Ou [14] explored how much work performance is reduced by five levels of speech intelligibility (STI = 0.08, 0.16, 0.23, 0.34, and 0.78) in Chinese open-plan offices. All of these studies mentioned above showed that work performance decreases with the increase of STI values.

Some studies [22, 25, 26] have proposed a prediction model to show how much work performance is decreased by speech with different intelligibility levels by summarising the results of experimental studies in which the effects of varying STI values on work performance were investigated. For instance, Ranz [26] proposed a prediction model in 2019 to show the relationship between STI values and the decrease in work performance (DP) in open-plan offices based on a series of experimental studies conducted in the German environment. Hongisto [25] in 2005 created an STI-DP model to predict the effects of irrelevant speech on work performance in open-plan offices. According to this STI-DP model, the work performance decrease occurs within the STI range of 0.2 and 0.5. In 2020, Haapakangas et al. [22] revised Hongisto’s STI-DP model through a systematic literature review. The work performance decrease is located within the STI range of 0.12-0.51. In the studies of Hongisto [25] and Haapakangas et al. [22], when developing the STI-DP model, all tests about the STI-performance relationship were analyzed without considering the effects of language environments. Moreover, the experimental data were almost collected from studies in Western language environments and rarely included experimental studies in Chinese environments.

1.2 Effects of language environments on speech intelligibility

Language is usually viewed as a critical factor when concerning speech intelligibility [27, 28]. The relationship between STI and speech intelligibility has been found to be different under different language environments [29, 30]. For example, the speech intelligibility score in English is much larger than that in Arabic, Chinese and Polish under the same STI condition [29]. Besides, Zhu et al. [30] revealed that the STI used to evaluate Chinese speech intelligibility could be calculated without considering the modulation of correction factor values, which is different from the recommendation of the International Standard for the objective rating of speech intelligibility, IEC 60268-16 [31]. As mentioned above, differences in languages may affect the relationship between STI and work performance [32]. Earlier studies [19, 33] in the Finnish environment showed that no significant differences in serial recall performance are found within the STI range of 0.0-0.35, and the performance decrease starts at the STI of 0.38 and reaches a maximum at the STI of 0.65. However, a previous study [20] in the Swedish environment indicated that the performance decrease of the serial recall task occurs within the STI range of 0.23-0.34. In addition, the study of Kang and Ou [8] found a difference between Chinese and non-Chinese environments in terms of the effects of STI values on work performance.

1.3 Task types

Some cognitive tasks (such as serial recall tasks, mental arithmetic, reading comprehension, operation span task, etc.) are usually used to quantify the negative effects of poor acoustic conditions in laboratory experiments researching the STI-performance relationship [34-36]. For these studies, the selection of task types is an important aspect of determining the negative effects of acoustic conditions since the effect of acoustic conditions on task performance varies by task type [37]. The study of Kang and Ou [8] shows a significant effect of STI on serial recall performance, while the effects of STI on the performance of mental arithmetic, reading comprehension and proofreading are not statistically significant. The study of Haapakangas et al. [36] indicates that there is a significant effect of acoustic conditions on serial recall performance, but the effects of acoustic conditions on the performance of creative thinking and proofreading tasks are not statistically significant.

Among the common cognitive tasks, the serial recall task is frequently used to test short-term memory efficiency [38, 39] and to evaluate the effects of environmental changes on work performance [21, 24, 34, 40]. This task requires subjects to recall a list of items in the order in which they appeared [38, 41]. Moreover, compared with other cognitive tasks, the serial recall task is more susceptible to speech noise of different intelligibility no matter in what kind of language environment.

1.4 Aim of this study

The following information can be summarised from the above literature review: (1) High speech intelligibility (or large STI value) is one of the main reasons causing the decrease in occupants' cognitive performance in open-plan offices. For determining the effects of STI on work performance, the serial recall task is frequently used in previous studies and is susceptible to the disturbance of speech noise. (2) Some STI-DP models have been created to evaluate work performance in different STI conditions and to guide the acoustic design in open-plan offices. The experimental data used to create the STI-DP model, however, rarely included experimental results in Chinese environments. (3) The language environment is an important factor when considering the relationship between perceived speech intelligibility and STI value. It may also be a non-negligible factor when determining the relationship between STI and DP.

In view of the research background, this study conducted a laboratory experiment to investigate the effects of STI conditions on occupants' work performance (i.e. serial recall performance) in Chinese environments. A total of seven STI conditions were concerned in this study. In addition, the effects of STI on serial recall performance in the Chinese environment from earlier studies were summarised to create an STI-DP model of the serial recall task applicable to the Chinese open-plan office. In particular, the following research questions are addressed:

1) What is the STI range that leads to changes in serial recall performance in Chinese environments?
2) Are there differences in the STI range between Chinese and non-Chinese environments for changes in occupants' serial recall performance?

2. Experiment

2.1 Methods and materials

Forty-one subjects were recruited from Huaqiao University in this experiment. There were 20 females and 21 males aged between 18 and 31 years, with an average of 22.4 years. All the subjects are native Chinese students and have no hearing problems. All were compensated for their participation.

The experiment was a one-way repeated measures design with seven levels of speech intelligibility according to the seven acoustic conditions during which the performance of the word serial recall task and subjective ratings of speech disturbance were tested. All seven acoustic conditions were presented using open-type headphones (Sennheiser HD 650 and 600). The tests were performed in a laboratory of Huaqiao University. Figure 1 shows the layout of the laboratory with the size of 6.6 x 6.5 x 3.0 m^3. Workstation E was arranged as the control console. Workstations R1 and R2 were used as test positions. A 1.5 m high partition was installed between R1 and R2 to avoid visual contact. The temperature was kept at 22-23 ℃. The vertical illumination level was around 530 lx on each table surface, and no glare problems were found on each table.

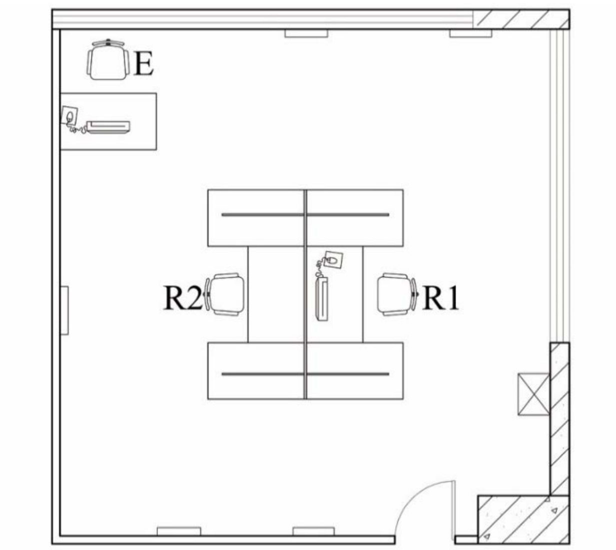

Figure 1 Layout of the test laboratory (E represents the position of the control console, and R1–R2 represent the test positions)

The word serial recall task was used to determine the impacts of STI on work performance. Each trial included ten sequences of Chinese words. For each sequence, seven words were extracted from one of 8 semantic categories (i.e., "fish and aquatic plants," "fruit and vegetable," "Chinese cities," "crop and trees," "human tissues and organs," "business and living goods," "food and beverages," and "animals"). The computer screen displayed one word (1 second on, 0.5 seconds off) at a time in a specific order. After all words were displayed, participants were required to recall the words in the order in which they appeared within 47 seconds.

The sound stimuli for auditory assessment were generated through the Odeon software. Firstly, a 3D model of the open-plan office was built using the SketchUp software. The dimension of this 3D model (37.2 x 109 x 3.6 m^3) was from the in-situ measurement in our recent study [42] (see Figure 2). Secondly, four office scenes were created by changing the sound absorption of office surfaces. Table 1 shows the information on 4 office scenes. As shown in Table 1, distraction distance r_D and comfort distance r_C of each office scene are given to show the acoustic performance of office scenes. The r_D describes the distance in meters from the speaker where the STI is below 0.5, which is used to predict the objective speech privacy of open-plan offices [43]. The r_C refers to the distance in meters in which the sound pressure level of speech is below 45 dBA, which is suggested to evaluate the effect of spatial attenuation of speech in open-plan offices [43-45]. Thirdly, one or two receiving positions (H1 and H2) were selected in each office scene (see Figure 2). Speech sounds and background noise were convoluted with the binaural room impulse at the selected position using the Odeon software. Finally, the convoluted speeches and background noise were played back through headphones (Sennheiser HD 600 and 650) and recorded via a binaural microphone (B & K, Type 4101) to ensure the sound pressure levels of distracting speech and background noise. Seven STI values were determined by calculating the EDT and SNR for each condition based on the method of Hongisto et al. [46] (see Table 1).

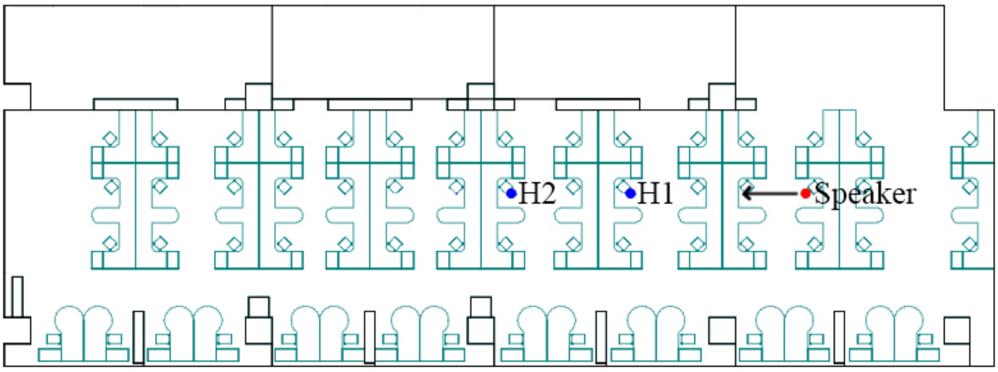

Figure 2 Layout of the simulated open-plan office (H1 and H2 present receiving positions)

Table 1 Information of seven acoustic conditions

No.	STI	$L_{A,S}$/dBA	$L_{P,A,B}$/dBA	$L_{Aeq,total}$/dBA	SNR/dBA	EDT/s	r_C/m	r_D/m	Office scenes
1	0.27	31.6	40.1	40.7	-8.5	0.20	3.8	3.5	1
2	0.36	35.7	39.1	40.7	-3.4	0.29	5.2	5.6	2
3	0.37	37.1	40.1	41.9	-3.0	0.11	3.8	3.5	1
4	0.39	38.9	36.8	41.0	2.1	0.40	6.7	8.5	3
5	0.41	39.2	39.1	42.2	0.1	0.22	5.2	5.6	2
6	0.50	42.2	32.1	42.6	10.1	0.75	11.9	13.5	4
7	0.54	46.4	32.6	46.6	13.8	0.64	11.9	13.5	4

Note:
SNR refers to the speech-to-noise ratio;
EDT was averaged over 250 to 4000 Hz octave bands.

The speech noise materials were 14 dry recordings of Chinese sentences, which were recorded in an anechoic room by female and male native Chinese speakers before the experiment. The background noise was a ventilation sound recorded from a working Gree vertical air conditioner (KFR-120LW/E(12568L)A1-N1) in an office. Our previous studies [8, 47] have used these speech and ventilation noises. The spectrum of the speech and ventilation noise is given in Figure 3.

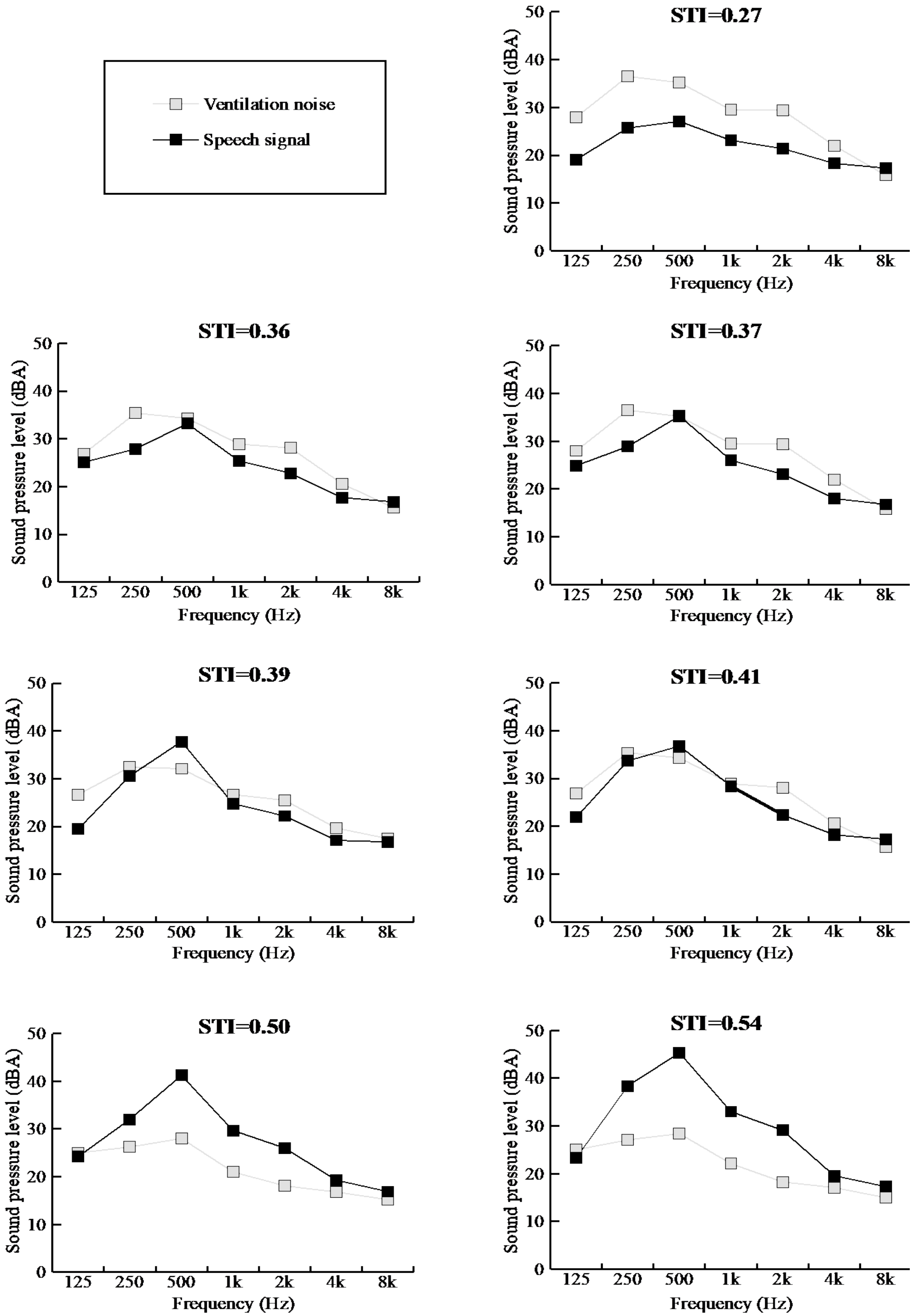

Figure 3 Average sound pressure levels of speech and ventilation noise for each STI condition

The experiment lasted about 2h20min. One or two subjects took part in the experiment each time. At the beginning of the experiment, subjects were informed of the experimental purposes and practised the serial recall task to familiarize themselves with the requirements of the task within 10 min in a quiet environment. During the formal testing stage, each subject performed one trial under each of the 7 STI conditions. The perceived speech disturbance of each subject was collected by the question answered on a 5-point scale (from 1 not at all to 5 very high) after the word serial recall task. Each STI condition took about 12 min. A 4-minute break was given after each condition. The presentation order of the STI conditions was randomized.

2.2 Results

Subjects' objective performance (accuracy rates of word serial recall tasks) and perceived speech disturbance under seven STI conditions are shown in Figures 4 and 5. For objective performance, participants' objective performance seems to decrease with the increase of STI values (see Figure 4). Repeated measures of analysis of variance (RMANOVA) tests reveal there are statistically significant differences among average accuracy rates under the 7 conditions (($F_{6,240}$ = 13.085, P-value = 0.000, partial η^2 = 0.246). F refers to F-ratio, which is the test statistic [48]. Partial η^2, a measure of the effect size, expresses the ratio of the sum of squares of the effect to the sum of squares of the effect and the error sum of squares [49]. The post-hoc contrast comparisons with Bonferroni adjustment were conducted to further examine the difference between any particular two conditions. This post hoc analysis can be summarized as follows: 1) the objective performance at STI = 0.54 is significantly lower than at STI = 0.27 ($P<0.01$), STI = 0.36 ($P<0.01$), STI = 0.37 ($P<0.01$), STI = 0.39 ($P<0.01$), and STI = 0.41 ($P<0.01$); 2) the objective performance at STI = 0.50 is significantly lower than at STI

= 0.27 (P<0.01) and STI = 0.36 (P<0.05).

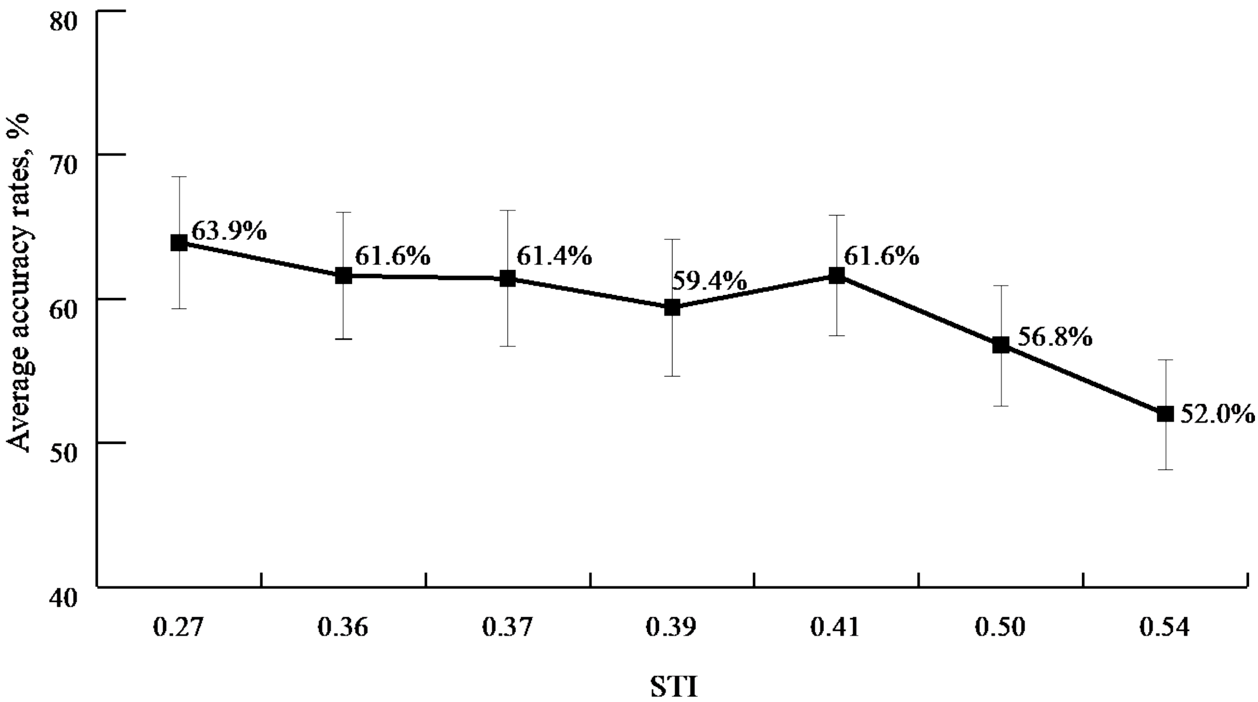

Figure 4 Subjects' average accuracy rates of the serial recall tasks under the 7 STI conditions (error bars define confidence intervals)

For subjective perceptions, the mean score of speech disturbance seems to increase with the increase of STI (see Figure 5). Friedman tests reveal that there is a statistically significant difference among the seven STI conditions in terms of speech disturbance (Friedman's Q(6) =116.07, P<0.01, W=0.472). Friedman's Q is the test statistic that summarizes how differently SIT conditions were rated in a single number. W is the effect size (Kendall's W value). Pairwise comparisons were subsequently performed with adjustment for multiple comparisons using a Bonferroni correction. The analysis of post hoc tests can be summarized as follows: 1) subjects' perceived speech disturbance in STI=0.27 is significantly lower than in any other STI condition (P<0.01 for all comparisons). 2) Perceived speech disturbance at STI=0.54 is significantly higher than at STI=0.36 (P<0.01), STI=0.37 (P<0.01), and STI=0.41 (P<0.05). 3) A borderline-significant difference (P=0.069) is shown between perceived speech disturbance at STI=0.50 and STI=0.36. It is worth mentioning that perceived speech disturbance at STI=0.27 is 1, which means that there is little speech disturbance affecting subjects when performing the serial recall task under the condition of STI=0.27. A possible explanation for this could be that the sound pressure level (SPL) of speech at STI=0.27 is too low (only 31.6 dBA, see Table 1). The SPL of ventilation noise was 40.1 dBA which was higher than speech. Considering the masking effects of ventilation noise, it is not surprising that all subjects could not hear speeches during testing in the condition of STI 0.27 and rated the disturbance level of speech as "not at all," despite the STI being calculated as 0.27 based on the SNR and early decay time (EDT).

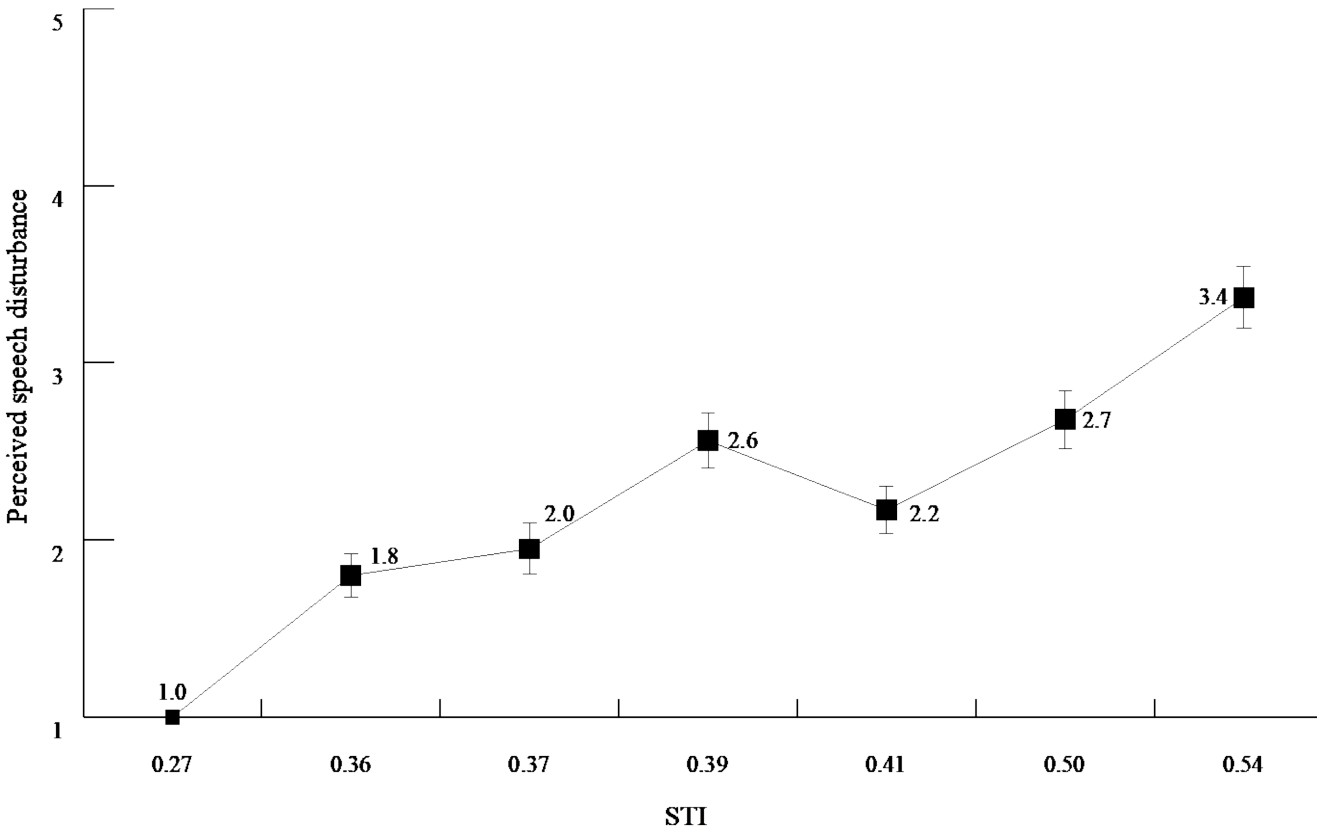

Figure 5 Subjects' perception of the speech disturbance under the 7 STI conditions (error bars define standard errors)

3. Relationships between STI and work performance in Chinese open-plan offices

3.1 Prediction model

The relationship between STI and work performance was one of the focuses of studies on open-plan offices' acoustic environments. The sigmoidal function is usually used to simulate the relationship between STI and the decrease in work performance (DP) in open-plan offices by previous studies [22, 25, 26]. The normal equation is:

$$F(x) = A - \frac{B}{1+\exp[(x-x_0)/k]} \tag{1}$$

in which A (%), B (%), x_0, and k are constants to be optimized.

This study also selected the sigmoidal function to simplify the relationship between STI and DP in Chinese environments. The model is built depending on equation (1) using Solver (Microsoft Office Package, Excel). The optimization task was to minimize the sum of squares of residuals. The constraints of the constants were 2-30 for A and B, 0.1-0.9 for x_0, and 0.03-0.1 for k. The value of R^2 describes the model's goodness of fit, and the Pearson correlation coefficients r_{xy} is calculated to compare with the other STI-DP models proposed by previous studies in this field.

3.2 Relationships between STI and DP in Chinese environments

There are the other four experimental studies about the STI-performance relationship in Chinese environments. Table 2 shows the detailed information of these studies. As seen in Table 2, the effects of STI conditions depended on the task type. Of the eight tests of the STI-performance relationship, three tests (i.e., tests 1, 2, and 4) found strongly significant effects of STI conditions on the accuracy rate of the serial recall task. Only one test (i.e., test 3) revealed strong effects of STI conditions on the accuracy rate of the reading comprehension task. The remaining tests (i.e., tests 5-8) found no statistically significant effects of STI conditions on the accuracy rates of tasks. The STI-DP model was created based on tests 1, 2, and 4 for the serial recall task.

Table 2 Descriptive information on eight tests of the STI-performance relationship in Chinese open-plan offices

Test No.	Study	No. of participants	Task type	Range in STI	Range in $L_{Aeq,total}$, dBA	Are there statistically significant differences?
1	This work	41	Serial recall task	0.27-0.54	40.7-46.6	Yes
2	Zhang et al. [47]	30	Serial recall task	0.00-0.66	54.6-55.2	Yes
3	Lou and Ou [14]	20	English literature reading comprehension	0.08-0.78	49.5-50.5	Yes
4	Kang and Ou [8]	38	Serial recall task	0.00-0.67	40.4-47.1	Yes
5	Kang and Ou [8]	38	Mental arithmetic	0.00-0.67	40.4-47.1	No
6	Kang and Ou [8]	38	Reading comprehension	0.00-0.67	40.4-47.1	No
7	Kang and Ou [8]	38	Proofreading	0.00-0.67	40.4-47.1	No
8	Yu et al. [50]	32	Simulated X-ray screening task	0.00-0.69	65.0	No

The DP in this study was obtained by the difference in the average accuracy rate between the reference and speech intelligibility conditions. For this work, the SPL of speech in STI 0.27 is very low (31.6 dBA, see Table 1). Moreover, as mentioned in Section 2.2, the mean score of perceived speech disturbance of STI 0.27 is 1, which means there is little speech disturbance affecting subjects when performing the serial recall task under the condition of STI 0.27. Thus, the STI of 0.27 was regarded as the reference condition to calculate the DP. In the studies of Kang and Ou [8] and Zhang et al. [47], the quiet condition (i.e., the condition without masking sound and speech noise) was viewed as the reference condition to calculate the DP of the serial recall task in different STI conditions. Zhang et al. [47] studied word serial recall performance under different STI conditions with four masking sounds. The DP in different STI conditions masked by pink noise and ventilation sound was calculated because these two sounds were also used in tests 1 and 4, respectively, to vary the STI value. The data from the experimental studies mentioned above are shown as individual points in Figure 6, and the S-shaped curve in Figure 6 is the result of the sigmoidal function based on the non-linear least-square fitting method.

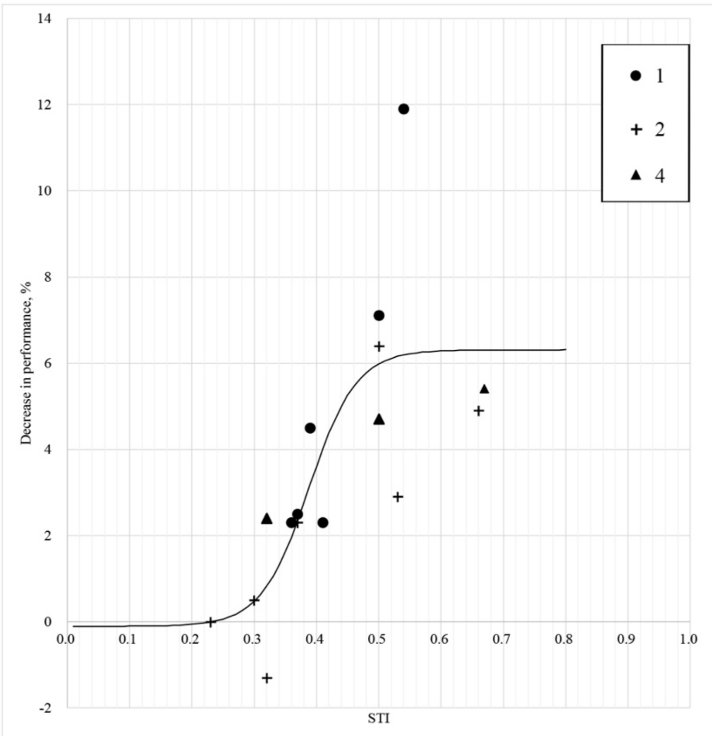

Figure 6 STI-DP model of the serial recall task in Chinese environments.

The task-specific model in Chinese environments was created based on the tests listed in Table 3.

Haapakangas et al. [22] introduced STI_{10} and STI_{90} to describe the STI range where work performance changes drastically. In their study, STI_n refers to the STI value in which n% of the maximum decrease in performance is reached. Thus, STI_{10} and STI_{90} were computed in this study to show the STI range for changing work performance. Table 3 shows the variables and correlation coefficients of the STI-DP model. As shown in Table 3, the performance decrease occurs within the STI range of 0.31-0.47 in the STI-DP model. The Pearson's correlation coefficient was calculated to show the relationship between observed and predicted DP. The larger coefficient, the better the model fit. As seen in Table 3, the Pearson's correlation coefficient STI-DP model is 0.79.

Table 3 Results of variables in the DP-STI model

Variables	A (%)	B (%)	x_0	k	R^2	r_{xy}	DP_{10}(%)	DP_{90}(%)	STI_{10}	STI_{90}
Model	6.4	6.3	0.4	0.04	0.63	0.79	0.6	5.7	0.31	0.47

Note:
r_{xy}: Pearson's correlation coefficient for the relationship between observed and predicted DP;
DP_{10}: The decrease in performance at STI_{10};
DP_{90}: the decrease in performance at STI_{90}.

4. Discussion

4.1 The effect of STI

The experiment results reveal that speech noise has a significant effect on serial recall performance and perceived speech disturbance. The accuracy rates significantly decrease with the increase of STI, while speech disturbance significantly increases with the increase of STI. These results are consistent with previous studies [8, 19, 24]. It is worth mentioning that the overall trends of serial recall performance and perceived speech disturbance have local changes at STI=0.41 (See Figures 4 and 5). A possible explanation for these changes could be the difference in SNR between conditions of STI=0.39 and STI=0.41. Both the EDT and SNR are two key parameters for determining the STI condition in open-plan offices. In this study, the STI conditions were controlled by changing the EDT and SNR, not just the SNR. As seen in Table 1, although the objective speech intelligibility in the condition of STI 0.39 is lower than in the condition of STI 0.41 due to the contribution of EDT, the SNR at STI=0.39 (2.1 dBA) was larger than that at STI=0.41 (0 dBA).

Moreover, as can be seen from Figure 6, there is an interesting situation to underline. For both the small and medium STI conditions (i.e., STI<0.5), the DP of the serial recall task in this study is close to that in the studies of Zhang et al. [47] and Kang and Ou [8]. For the large STI conditions (i.e., STI≥0.5), however, the DP of the serial recall task in the present study is much larger than that in the other two studies. For instance, at STI=0.50, the DP of the serial recall task in this study is 7.1%, which is greater than that in the studies of Zhang et al. [47] (6.4%) and Kang and Ou [8] (4.7%). The maximum DP in this study is as high as 11.9%, which is larger than that in the other two studies. One possible explanation for this situation could be the differences in SNR and EDT among the three tests. In the condition of STI=0.50, SNR in this study was 10.1 dBA, which is much larger than in Zhang et al. [47] (4.8 dBA) and Kang and Ou [8] (5.5 dBA). In addition, Zhang et al. [47] and Kang and Ou [8] altered the STI value by adjusting the SNR. The EDT in these two studies remained unchanged (0.55 s and 0.33 s, respectively). However, the EDT at STI=0.50 and STI=0.54 in this study was 0.75 s and 0.64 s, respectively. Braat-Eggen et al. [51] revealed that the perceived speech disturbance increases when working in office scenes with a longer reverberation time. Thus, it is possible that speech in the condition with a long reverberation and large SNR may cause more disturbance to occupants' serial recall performance.

4.2 Comparison of relationships between STI and serial recall performance with other studies

A comparison of the curves between STI values and the decrease in serial recall performance was carried out to obtain the difference between this study and the studies of Haapakangas et al. [22] and Renz [26]. Figure 7 shows the three curves and corresponding DP_{10}, DP_{90} and r_{xy} values. In 2020, Haapakangas et al. [22] proposed 4 STI-DP models based on 34 laboratory tests investigating the change in cognitive performance with the increase of STI. Of these 34 tests, 19 were carried out in Finnish, 6 in German and 7 in Swedish environments. The remaining two tests were conducted in French and Chinese environments, respectively. All speech sounds used for these 34 tests were recorded in the participants' native language to ensure participants could understand. In the study of Haapakangas et al. [22], model 3 presents the relationship between STI and DP of serial recall task, which was obtained through a summary analysis of 11 experimental studies conducted in 4 language environments (i.e. Finnish, German, Swedish, and French environments). Renz [26] proposed an STI-DP model based on six laboratory studies that analyzed the change in the serial recall performance with the STI increase in the German environment.

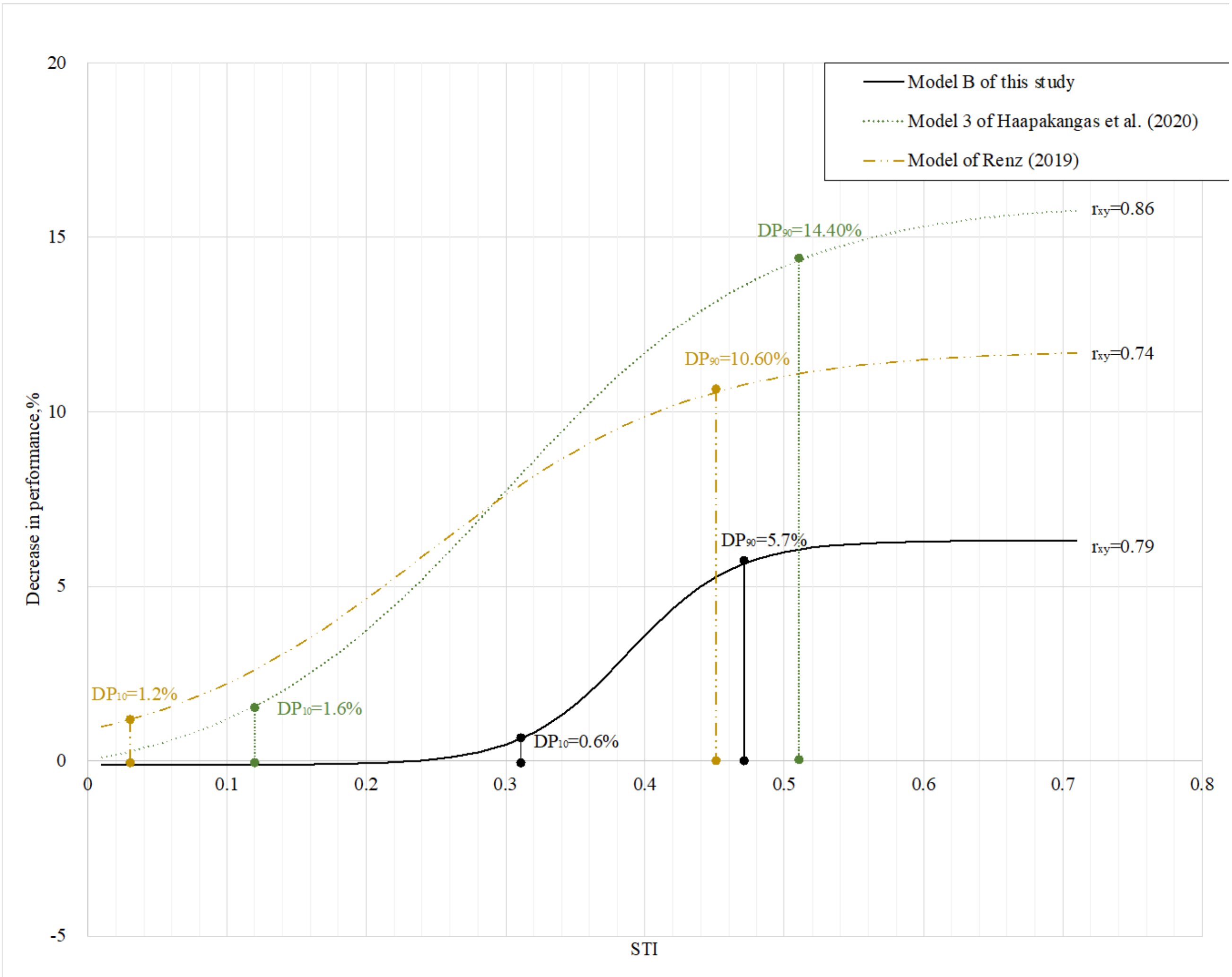

Figure 7 A comparison of models expressing the relationship between STI and DP of the serial recall task. The models of Haapakangas et al. [22], Renz [26] and this study are included.

Based on STI_{10} and STI_{90} values in Figure 7, the decrease in serial recall performance is located within the STI range of 0.31 and 0.47 in Chinese environments, while it occurs within a larger range (0.03-0.45) in German environments (the model of Ranz). According to model 3 of Haapakangas et al. [22], serial recall performance decreases within the STI range of 0.12 to 0.51, which is much larger than the STI range in Chinese environments. Additionally, the DP gradient of the STI-DP model in this study from STI_{10} to STI_{90} is close to that of model 3 of Haapakangas et al. [22] and steeper than that of Ranz [26] (see Figure 7). The DP of the serial recall task in this study is always lower than those in the other two models under the same STI condition. These results mentioned above imply that: 1) the decrease in serial recall performance in Chinese environments occurs within a narrower STI range than in Western language environments. 2) The effect of speech noise in the Chinese environment on serial recall performance is lower than that in Western languages, while the average change rate of the DP is not less in Chinese environments than in Western language environments within the STI range of 0.31-0.47.

It is worth noting that the maximum differences in the STI_{10} and STI_{90} among the three curves are up to 0.27 and 0.06, respectively (see Figure 4). Bradley et al. [52] found that a just noticeable difference (JND) of STI is 0.03, which means that the differences in the STI_{10} and STI_{90} are large. One possible explanation could be the difference in language environments. Chinese is a monosyllabic language that belongs to the Sino-Tibetan language family. Its syllable information consists of consonants, finals and tones. Unlike many Western languages such as Finnish, German and Swedish, the tone in Chinese is used to distinguish the meanings of words. Furthermore, a difference in the average spectrum is found between Chinese and Western languages [30]. The differences in pronunciation and vocal characteristics mentioned above may cause the difference in perceived speech intelligibility in varying language environments and thus impact the relationship between speech intelligibility and work performance. In addition, previous studies [29, 53] have demonstrated that the perceived speech intelligibility differs among language environments even under the same STI value. That is to say, the STI value representing the same speech intelligibility may differ between Chinese and Western language environments such as Finnish, Swedish and German. Overall, differences in STI_{10} and STI_{90} show that the STI range for work performance changing varies across language environments. Another possible explanation for the difference in STI_{90} could be the differences in the measurement of STI. The STI values reported in Haapakangas et al. [22], Ranz [26], and this study have not been calculated in the same method. In Haapakangas et al. [22] and Ranz [26], some STI values have been obtained based on the method recommended in IEC 60268-16, which has been revised three times in the past 20 years [31, 54-56]; while some STI values have been determined based on the method described by previous studies [46, 57] by considering the result of reverberation time (or EDT) and the energy-equivalent sound pressure levels of speech and background sounds. In this study, the determination of all STI values was based on the method of Hongisto et al. [46].

4.3 Limitations of this study

The STI-DP model of serial recall task involves inevitable uncertainty because of the determination of STI conditions. In this study and the other two studies, the reported STI values have been determined based on the method described by Hongisto et al. [46] by using EDT and SNR values. Although similar methods of determining STI conditions are usually applied in previous studies [19, 20, 23], it is different from the international standard ISO 60268-16.

Another source of uncertainty is the calculation of the DP. The performance in a quiet environment is usually viewed as an ideal reference to calculate the DP. In this study, however, the minimum STI value (0.27) was used as a reference condition to calculate DP. Although there was no speech disturbance at STI=0.27 based on subjects' rating results, the background noise (i.e. ventilation sound) may also exert adverse effects on subjects than in a quiet environment.

This study was limited to the effects of various STI conditions on serial recall performance. The relationship between STI and work performance may vary by task type. For instance, Kang and Ou [8] found that STI conditions have no main effects on the objective performance of mental arithmetic, proofreading, and reading comprehension (in Chinese) tasks, with the exception of the serial recall performance. Lou and Ou [14] revealed that STI conditions have a significant effect on the objective performance of English literature reading comprehension, which was inconsistent with findings on the relationship between STI and reading comprehension performance (in Chinese) [8]. More studies are needed to explore the STI-DP model for specific work activities. In addition, inter-individual variability (such as gender, noise sensitivity, etc.) is a critical factor that impacts participants' serial recall performance [24, 58]. In this study, the difference in the effects of speech between Chinese and non-Chinese environments (see Figure 7) may also be affected by inter-individual variability. Further studies are needed to examine the impact of non-acoustic factors (such as inter-individual variability, language environment, etc.) on serial recall performance in open-plan offices.

5. Conclusion

A laboratory experiment was conducted to determine the effects of STI conditions on occupants' serial recall performance in Chinese open-plan offices. Both the occupants' objective performance (accuracy rates of word serial recall tasks) and perceived speech disturbance were tested under seven STI conditions (i.e., STI = 0.27, 0.36, 0.37, 0.39, 0.41, 0.50 and 0.54).

Based on the results of the laboratory experiment and analysis of experimental data from two experimental studies, an STI-DP model of the serial recall task for Chinese open-plan offices was first proposed by the non-linear least square fitting method. This model can be used to evaluate the effects of speech on serial recall performance in different STI conditions in Chinese open-plan offices.

The main findings can be drawn from the analysis results:

1) The experimental results show that with the increase of the STIs in Chinese open-plan offices, occupants' performance of the serial recall task decreases, and their perceived speech disturbance increases.

2) According to the STI-DP model, in the Chinese environment, serial recall performance starts to decrease when STI is above 0.31, while the decrease in performance reaches the maximum when STI exceeds 0.47. This STI range for changes in serial recall performance is much narrower than in non-Chinese environments.

3) The effect of speech noise in the Chinese environment on serial recall performance is lower in comparison with non-Chinese environments, but the average change rate of the DP in Chinese is not lower than in non-Chinese environments in the range of STI = 0.31 to 0.47.

Acknowledgement

The work was supported by a PhD studentship funded by Hong Kong Polytechnic University, Natural Science Foundation of Fujian Province of China (2018J01070), Social Science Planning Project of Fujian Province of China (FJ2021B075) and State Key Lab of Subtropical Building Science, South China University of Technology (2022ZA02).

Reference

[1] J. Pejtersen, L. Allermann, T. S. Kristensen, and O. M. Poulsen, "Indoor climate, psychosocial work environment and symptoms in open-plan offices," *Indoor Air,* vol. 16, no. 5, pp. 392-401, 2006, doi: 10.1111/j.1600-0668.2006.00444.x.

[2] A. Shafaghat, A. Keyvanfar, M. S. Ferwati, and T. Alizadeh, "Enhancing staff's satisfaction with comfort toward productivity by sustainable Open Plan Office Design," *Sustainable Cities and Society*, vol. 19, pp. 151-164, 2015, doi: 10.1016/j.scs.2015.08.001.

[3] ISO 22955:2021, *Acoustics - Acoustic quality of open office spaces*, International Organization for Standardization. Geneva, Switzerland; 2021.

[4] C. M. Mak and Y. P. Lui, "The effect of sound on office productivity," *Building Services Engineering Research and Technology,* vol. 33, no. 3, pp. 339-345, 2011, doi: 10.1177/0143624411412253.
[5] J. Kim and R. de Dear, "Workspace satisfaction: The privacy-communication trade-off in open-plan offices," *Journal of environmental psychology,* vol. 36, pp. 18-26, 2013, doi: 10.1016/j.jenvp.2013.06.007.
[6] S. Kang, D. Ou, and C. M. Mak, "The impact of indoor environmental quality on work productivity in university open-plan research offices," *Building and Environment*, vol. 124, pp. 78-89, 2017, doi: 10.1016/j.buildenv.2017.07.003.
[7] M. Rashid, J. Wineman, and C. Zimring, "Space, behavior, and environmental perception in open-plan offices: A prospective study," *Environment and Planning. B, Planning & Design*, vol. 36, no. 3, pp. 432-449, 2009, doi: 10.1068/b33034.
[8] S. Kang and D. Ou, "The effects of speech intelligibility on work performance in Chinese open-plan offices: A laboratory study," *Acta Acustica United with Acustica,* vol. 105, no. 1, 2018.
[9] A. Haapakangas, V. Hongisto, M. Eerola, and T. Kuusisto, "Distraction distance and perceived disturbance by noise—An analysis of 21 open-plan offices," *Journal of the Acoustical Society of America,* vol. 141, no. 1, p. 127, 2017.
[10] C. B. Danielsson and L. Bodin, "Office type in relation to health, well-being, and job satisfaction among employees," *Environment and Behavior,* vol. 40, no. 5, pp. 636-668, 2008, doi: 10.1177/0013916507307459.
[11] M. Pierrette, E. Parizet, P. Chevret, and J. Chatillon, "Noise effect on comfort in open-space offices: Development of an assessment questionnaire," *Ergonomics,* vol. 58, no. 1, pp. 96-106, 2015, doi: 10.1080/00140139.2014.961972.
[12] P. J. Lee, B. K. Lee, J. Y. Jeon, M. Zhang, and J. Kang, "Impact of noise on self-rated job satisfaction and health in open-plan offices: A structural equation modelling approach," *Ergonomics,* vol. 59, no. 2, pp. 222-234, 2016, doi: 10.1080/00140139.2015.1066877.
[13] T. L. Smith-Jackson and K. W. Klein, "Open-plan offices: Task performance and mental workload," *Journal of environmental psychology,* vol. 29, no. 2, pp. 279-289, 2009, doi: 10.1016/j.jenvp.2008.09.002.
[14] H. Lou and D. Ou, "The effects of speech intelligibility on English scientific literature reading in Chinese open-plan offices," *Journal of the Acoustical Society of America,* vol. 147, no. 1, pp. EL1-EL6, 2020, doi: 10.1121/10.0000497.
[15] A. Kaarlela-Tuomaala, R. Helenius, E. Keskinen, and V. Hongisto, "Effects of acoustic environment on work in private office rooms and open-plan offices – longitudinal study during relocation," *Ergonomics,* vol. 52, no. 11, pp. 1423-1444, 2009, doi: 10.1080/00140130903154579.
[16] M. Zhang, J. Kang, and F. Jiao, "A social survey on the noise impact in open-plan working environments in China," *The Science of the Total Environment,* vol. 438, pp. 517-526, 2012, doi: 10.1016/j.scitotenv.2012.08.082.
[17] S. P. Banbury, W. J. Macken, S. Tremblay, and D. M. Jones, "Auditory distraction and short-term memory: Phenomena and practical implications," *Human Factors,* vol. 43, no. 1, pp. 12-29, 2016, doi: 10.1518/001872001775992462.
[18] J. E. Marsh, R. W. Hughes, and D. M. Jones, "Interference by process, not content, determines

semantic auditory distraction," *Cognition,* vol. 110, no. 1, pp. 23-38, 2009, doi: 10.1016/j.cognition.2008.08.003.

[19] M. Haka, A. Haapakangas, J. Keränen, J. Hakala, E. Keskinen, and V. Hongisto, "Performance effects and subjective disturbance of speech in acoustically different office types - a laboratory experiment," *Indoor Air,* vol. 19, no. 6, pp. 454-467, 2009, doi: 10.1111/j.1600-0668.2009.00608.x.

[20] H. Jahncke, V. Hongisto, and P. Virjonen, "Cognitive performance during irrelevant speech: Effects of speech intelligibility and office-task characteristics," *Applied Acoustics*, vol. 74, no. 3, pp. 307-316, 2013, doi: 10.1016/j.apacoust.2012.08.007.

[21] S. J. Schlittmeier, J. Hellbrück, R. Thaden, and M. Vorländer, "The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance," *Ergonomics,* vol. 51, no. 5, pp. 719-736, 2008, doi: 10.1080/00140130701745925.

[22] A. Haapakangas, V. Hongisto, and A. Liebl, "The relation between the intelligibility of irrelevant speech and cognitive performance—A revised model based on laboratory studies," *Indoor Air,* vol. 30, no. 6, pp. 1130-1146, 2020, doi: 10.1111/ina.12726.

[23] T. Renz, P. Leistner, and A. Liebl, "Auditory distraction by speech: Can a babble masker restore working memory performance and subjective perception to baseline?," *Applied Acoustics,* vol. 137, pp. 151-160, 2018, doi: 10.1016/j.apacoust.2018.02.023.

[24] A. Ebissou, E. Parizet, and P. Chevret, "Use of the Speech Transmission Index for the assessment of sound annoyance in open-plan offices," *Applied Acoustics,* vol. 88, pp. 90-95, 2015, doi: 10.1016/j.apacoust.2014.07.012.

[25] V. Hongisto, "A model predicting the effect of speech of varying intelligibility on work performance," *Indoor Air,* vol. 15, no. 6, pp. 458-468, 2005, doi: 10.1111/j.1600-0668.2005.00391.x.

[26] T. Renz, "Comparison of models to predict the effect of background speech on work performance in open-plan offices," in *Proceedings of the International Congress on Acoustics*, Aachen, German, 2019, vol. 2019, pp. 2415-2422, doi: 10.18154/RWTH-CONV-239819.

[27] K. Lee and A. R. Bradlow, "The role of first-language production accuracy and talker-listener alignment in second-language speech Intelligibility," *Journal of the Acoustical Society of America,* vol. 132, no. 3, pp. 2078-2078, 2012, doi: 10.1121/1.4755666.

[28] M. Cooke and M. L. G. Lecumberri, "The intelligibility of Lombard speech for non-native listeners," *Journal of the Acoustical Society of America,* vol. 132, no. 2, pp. 1120-1129, 2012, doi: 10.1121/1.4732062.

[29] L. Galbrun and K. Kitapci, "Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions," *Applied Acoustics,* vol. 114, pp. 79-91, 2016.

[30] P. Zhu, F. Mo, and J. Kang, "Relationship between Chinese speech intelligibility and speech transmission index under reproduced general room conditions," *Acta Acustica United with Acustica,* vol. 100, no. 5, pp. 880-887(8), 2014.

[31] IEC 60268-16:2020, *Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. ED.5*, International Electrotechnical Commission, Geneva, Switzerland, 2020.

[32] S. Kang, D. Ou, and Y. Zhang, "The effects of speech intelligibility and masking sound on work

performance in open-plan offices: an experimental review," *Building Science*, vol. v.35;No.267, no. 10, pp. 179-184, 2019 [in Chinese].

[33] V. Hongisto, J. Varjo, H. Leppämäki, D. Oliva, and J. Hyönä, "Work performance in private office rooms: The effects of sound insulation and sound masking," *Building and environment*, vol. 104, pp. 263-274, 2016, doi: 10.1016/j.buildenv.2016.04.022.

[34] A. Haapakangas, V. Hongisto, J. Hyönä, J. Kokko, and J. Keränen, "Effects of unattended speech on performance and subjective distraction: The role of acoustic design in open-plan offices," *Applied acoustics*, vol. 86, pp. 1-16, 2014, doi: 10.1016/j.apacoust.2014.04.018.

[35] N. Venetjoki, A. Kaarlela-Tuomaala, E. Keskinen, and V. Hongisto, "The effect of speech and speech intelligibility on task performance," *Ergonomics*, vol. 49, no. 11, pp. 1068-1091, 2007, doi: 10.1080/00140130600679142.

[36] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, and E. Keskinen, "Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices," *Acta Acustica United with Acustica*, vol. volume 97, no. 4, pp. 641-655(15), 2011.

[37] J. Reinten, P. E. Braat-Eggen, M. Hornikx, H. S. M. Kort, and A. Kohlrausch, "The indoor sound environment and human task performance: A literature review on the role of room acoustics," *Building and Environment*, vol. 123, pp. 315-332, 2017, doi: 10.1016/j.buildenv.2017.07.005.

[38] W. E. Michael and T. K. Mark, *Cognitive psychology: A student's handbook*. London: London: Taylor and Francis, 2015.

[39] M. Botvinick and L. M. Bylsma, "Regularization in short-term memory for serial order," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 2, pp. 351-358, 2005, doi: 10.1037/0278-7393.31.2.351.

[40] L. Brocolini, E. Parizet, and P. Chevret, "Effect of masking noise on cognitive performance and annoyance in open plan offices," *Applied Acoustics*, vol. 114, pp. 44-55, 2016, doi: 10.1016/j.apacoust.2016.07.012.

[41] A. K. Haberlandt, "Serial Recall," in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan Eds. New York, NY: Springer New York, 2011, pp. 2265-2266.

[42] S. Kang, C. M. Mak, D. Ou, and Y. Zhang, "An investigation of acoustic environments in large and medium-sized open-plan offices in China," *Applied Acoustics*, vol. 186, p. 108447, 2022, doi: 10.1016/j.apacoust.2021.108447.

[43] ISO 3382-3:2022, *Acoustics - Measurement Of Room Acoustic Parameters - Part 3: Open Plan Offices*, International Organization for Standardization. Geneva, Switzerland; 2022.

[44] V. Hongisto and J. Keränen, "Comfort distance—a single-number quantity describing spatial attenuation in open-plan offices," *Applied Sciences*, vol. 11, no. 10, p. 4596, 2021, doi: 10.3390/app11104596.

[45] V. Hongisto, J. Keränen, L. Labia, and R. Alakoivu, "Precision of ISO 3382-2 and ISO 3382-3 – A Round-Robin test in an open-plan office," *Applied Acoustics*, vol. 175, 2021, doi: 10.1016/j.apacoust.2020.107846.

[46] V. Hongisto, J. Keränen, and P. Larm, "Simple Model for the Acoustical Design of Open-Plan Offices," *Acta Acustica United with Acustica*, 2004.

[47] Y. Zhang, D. Ou, and S. Kang, "The effects of masking sound and signal-to-noise ratio on work performance in Chinese open-plan offices," *Applied Acoustics*, vol. 172, 2021, doi:

10.1016/j.apacoust.2020.107657.

[48] C. L. Stamm and M. J. Safrit, "Comparison of Significance Tests for Repeated Measures ANOVA Design," *Research Quarterly of the American Association for Health, Physical Education and Recreation,* vol. 46, no. 4, pp. 403-409, 1975, doi: 10.1080/10671315.1975.10616696.

[49] J. M. Maher, J. C. Markey, and D. Ebert-May, "The Other Half of the Story: Effect Size Analysis in Quantitative Research," *CBE Life Sci Educ,* vol. 12, no. 3, pp. 345-351, 2013, doi: 10.1187/cbe.13-04-0082.

[50] R. Yu, L. Yang, and C. Liu, "Effects of target prevalence and speech intelligibility on visual search performance," *Measurement and Control (London),* vol. 48, no. 3, pp. 87-91, 2015, doi: 10.1177/0020294015569265.

[51] E. Braat-Eggen, M. K. v. d. Poll, M. Hornikx, and A. Kohlrausch, "Auditory distraction in open-plan study environments: Effects of background speech and reverberation time on a collaboration task," *Applied acoustics,* vol. 154, pp. 148-160, 2019, doi: 10.1016/j.apacoust.2019.04.038.

[52] J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in C50 for speech," *Applied acoustics,* vol. 58, no. 2, pp. 99-108, 1999, doi: 10.1016/S0003-682X(98)00075-9.

[53] J. Kang, "Comparison of speech intelligibility between English and Chinese," *Journal of the Acoustical Society of America,* vol. 103, no. 2, pp. 1213-1216, 1998.

[54] IEC 60268 - 16:2011, *Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. Ed.4,* International Electrotechnical Commission, Switzerland, 2011.

[55] IEC 60268 - 16:1998, *Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. Ed.2,* International Electrotechnical Commission, Geneva, Switzerland, 1998.

[56] IEC 60268 - 16:2003, *Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. Ed.3,* International Electrotechnical Commission, Geneva, Switzerland, 2003.

[57] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *The Journal of the Acoustical Society of America,* vol. 77, no. 3, pp. 1069-1077, 1985, doi: 10.1121/1.392224.

[58] Ellermeier, W., & Zimmer, K. "Individual differences in susceptibility to the “irrelevant speech effect”, " *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2191–2199, 1997. https://doi.org/10.1121/1.419596.