MDPI

*Article*

# Impact of Echo Interference on Speech Intelligibility in Extra-Large Spaces

**Wenkai Wang [1], Hui Ma [1,*] , Chao Wang [1] , Siyang Dong [1], Wenlin Hu [2,3] and Bin He [2,3]**

[1] School of Architecture, Tianjin University, Weijin Road 92, Nankai District, Tianjin 300072, China; wangwk2001@tju.edu.cn (W.W.); pdwangchao@tju.edu.cn (C.W.); dsy666@tju.edu.cn (S.D.)

[2] National Engineering Research Center for Digital Construction and Evaluation of Urban Rail Transit, China Railway Design Corporation, Tianjin 300308, China

[3] National-Local Joint Engineering Laboratory for Rail Transit Survey and Design, China Railway Design Corporation, Tianjin 300308, China

[*] Correspondence: mahui@tju.edu.cn

## Abstract

In extra-large spaces, the varying distances between distributed loudspeakers and listeners lead to sound delays in the loudspeakers' concentrated projection areas. When combined with the inherent long-delay reflected sounds in those spaces, this dual effect exacerbates the echo problems and poses challenges to maintaining speech intelligibility. To explore the influence mechanism of echo interference on speech intelligibility in extra-large spaces, a questionnaire survey was carried out in two representative extra-large buildings, and then listening experiments were conducted in the laboratory under different echo conditions and impulse characteristics. The results highlighted that (1) apparent echo problems existed in extra-large spaces and severely affected speech intelligibility; (2) the echo phenomenon can be classified into three groups—no echo (0 ms), short delay (100 or 200 ms), and long delay (≥300 ms)—with the detrimental effect on intelligibility increasing across the groups; and (3) a curve was established to describe the relationship between speech intelligibility and STI in extra-large spaces, and compared with the standard curve, the STI thresholds require further adjustment. These findings indicate that echoes in extra-large spaces significantly impair speech intelligibility and reduce the accuracy of its prediction, and therefore should not be neglected.

**Keywords:** echo; speech intelligibility; extra-large space

## 1. Introduction

Sound reinforcement systems are widely used in extra-large spaces, particularly in transportation buildings. Their performance determines whether users can accurately and clearly receive broadcast information, directly affecting their satisfaction and comfort [1]. Moreover, such systems play a vital role in providing guidance and facilitating safe evacuation during emergencies [2,3]. Therefore, maintaining clear speech intelligibility in these spaces is crucial [4].

Echo, apart from room acoustics [5,6] and the signal-to-noise ratio (SNR) [7], is also an important factor affecting speech intelligibility [8]. Haas [9] developed an apparatus for artificially generating echoes and conducted measurements with numerous observers under controlled conditions, demonstrating that both echo delay time and echo intensity affect the audibility of speech. In ordinary spaces (i.e., spaces other than the extra-large ones, which are defined in this study as having a volume smaller than 125,000 $m^3$ [10]),

late reflections generally reduce intelligibility, whereas reflections arriving shortly after the direct sound can be beneficial. This general trend has been consistently demonstrated in the literature. For instance, Lochner et al. [11] reported that a 30 ms delay can improve intelligibility equivalently to a 3 dB increase in the direct sound level, with the benefit diminishing and disappearing at around 95 ms. Similarly, Warzybok et al. [12] found that intelligibility remains stable for single early reflections up to approximately 25 ms, declines gradually at longer delays, and deteriorates substantially at 200 ms.

Most existing studies on reflections have focused on delays within 200 ms [13]. Unlike ordinary spaces, extra-large spaces are prone to long-path echoes due to their immense volumes [14]. In addition, commonly adopted multi-source sound reinforcement systems can introduce supplementary "echo-like" effects, as sound from neighboring loudspeakers may reach certain listening positions with delays of several hundred milliseconds [13]. Such delays require particular consideration in the acoustic design aimed at optimizing speech intelligibility in extra-large spaces.

Recognizing the detrimental impact of long-path echoes, Cui et al. examined the effects of word familiarity [15] and inter-word pauses [13] on intelligibility under echo delays exceeding several hundred milliseconds. However, these studies did not account for room acoustics, which may compromise the validity of the Speech Transmission Index (STI) as a predictor of speech intelligibility. The STI, introduced by Houtgast and Steeneken [16–18], is the predominant objective measure used in ordinary spaces, but its applicability in large or complex spaces remains uncertain. Kang [19] found that, for a given STI value, intelligibility scores were lower in ordinary spaces than in long spaces (i.e., spaces where the length is greater than six times the width and the height [20]). Zhu [21] reported significant differences in the STI–intelligibility relationship across spaces with varying volumes and shapes. Liu et al. [22] further noted that STI values in large spaces should not be interpreted in the same way as in ordinary spaces, revising the STI–speech intelligibility relationship and proposing new rating thresholds.

Given the complex acoustic conditions in extra-large spaces, it is therefore necessary to examine the correspondence between STI and speech intelligibility under echo interference, and to consider the effects of long-delay echoes. The present study aims to: (1) Evaluate the impact of echoes on speech intelligibility from the listener's perspective; (2) Investigates the mechanisms through which echo interference influences intelligibility in extra-large spaces, and (3) explores the STI–intelligibility relationship under echo interference conditions.

## 2. Methods

### 2.1. Survey Site and Questionnaire

To assess the current status of the speech intelligibility in extra-large spaces, questionnaire surveys were conducted in Tianjin Railway Station and Tianjin Binhai International Airport. Both buildings serve as major transportation hubs in northern China and are representative of extra-large transportation buildings. The study focused on the second-floor waiting hall of Tianjin Railway Station and Terminal 2 at Tianjin Binhai International Airport (Figure 1), which are high-volume, high-traffic areas with dense loudspeaker coverage and are key to improving speech intelligibility.

A questionnaire (Figure 2) was designed to identify users' subjective evaluations of loudspeaker clarity and speech intelligibility in the target spaces. Two sections are contained in it: The first section examined respondents' perceptions of the acoustic environment. Q1–Q4 assessed their perceptions of the importance, satisfaction, clarity, and loudness of sound reinforcement systems using a 5-point Likert scale. Q5 investigated their perceptions of echoes or delayed sounds, and Q6 asked them to select and rank the factors
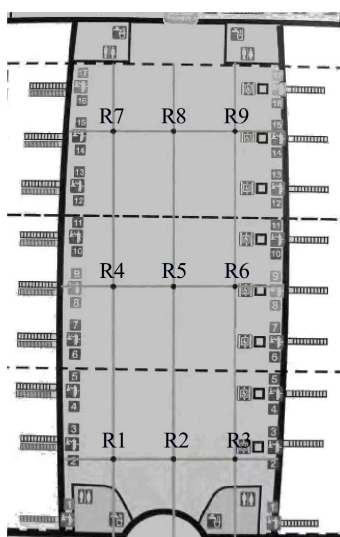
considered most influential to speech intelligibility. In the second section, respondents' demographic information, including gender, hearing condition, age, and the duration of stay in the space were collected.
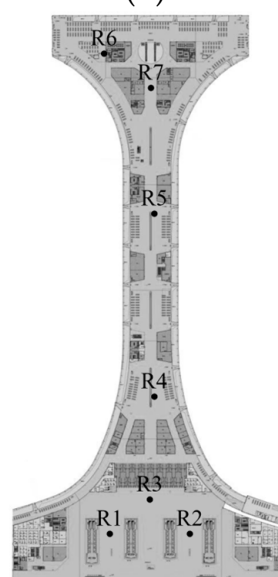


**Figure 1.** Site of the Questionnaire Survey. (**a**) The second-floor waiting hall of Tianjin Railway Station (26,700 m$^2$; maximum height approximately 19.9 m); (**b**) Terminal 2 of Tianjin Binhai International Airport (248,000 m$^2$; maximum height approximately 27.2 m); (**c**) floor plan of the second-floor waiting hall of Tianjin Railway Station with measurement points (R1–R9); and (**d**) floor plan of Terminal 2 at Tianjin Binhai International Airport with measurement points (R1–R7).

For Q2–Q4, responses rated above moderate were regarded as positive feedback, while those rated moderate or below were regarded as negative feedback. For Q6, the four selected factors were weighted with scores of 4 to 1, corresponding to the first through fourth choices.

Questionnaires were randomly administered to respondents at various locations within the survey area to capture their assessments of the building's overall acoustic environment. A total of 193 questionnaires were issued, and 189 valid responses were collected after excluding those from participants with hearing impairments. The demographic characteristics of the respondents are summarized in Table 1.

**Subjective Assessment**

Q1. Do you think the clarity of the sound reinforcement system is important?

☐ Not at all ☐ Not important ☐ Moderately ☐ Important ☐ Very important

Q2. Are you satisfied with the current clarity of the sound reinforcement system?

☐ Not at all ☐ Not satisfy ☐ Moderately ☐ Satisfy ☐ Very satisfy

Q3. Do you find the current sound reinforcement system clear?

☐ Not at all ☐ Not clear ☐ Moderately ☐ Clear ☐ Very clear

Q4. Do you find the current sound reinforcement system loud enough?

☐ Not at all ☐ Not loud ☐ Moderately ☐ Loud ☐ Very loud

Q5. Do you notice any echo or delayed sound from different loudspeakers in the sound reinforcement system?

☐No ☐Uncertainly ☐Yes

Q6. According to your actual experience, please select in order the four factors that most affect the clarity of the sound reinforcement system：_____

a. People talking b. Echoes within the space c. Low sound level of the system

d. Interference from multiple sound sources e. Noise from air conditioning

f. Noise from airplanes/trains g. Noise from shop advertisements

h. Noise from elevators i. Others: _____

**Demographic Information and Length of Stay**

Q7. Gender

☐ Male ☐ Female

Q8. Is your hearing normal?

☐ Yes ☐ No

Q9. Age:

☐ <20 ☐ 20–40 ☐ 40–60 ☐ >60

Q10. Length of stay

☐ <15 min ☐ 15–60 min ☐ >60 min

**Figure 2.** The questionnaire used in the survey.

**Table 1.** Statistics of Respondents' Gender, Age, and Length of Stay.

| Survey Locations | Gender | | Age | | | | Length of Stay (min) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | <20 | 20–40 | 40–60 | >60 | <15 | 15–60 | >60 |
| Tianjin Railway Station | 55.5% | 44.5% | 38.8% | 55.7% | 4.4% | 1.1% | 12.3% | 57.7% | 30.0% |
| Binhai International Airport | 52.9% | 47.1% | 25.4% | 58.8% | 13.8% | 2.0% | 10.0% | 36.1% | 53.9% |
| Total | 54.5% | 45.5% | 31.6% | 56.9% | 10.0% | 1.5% | 10.8% | 46.2% | 43.0% |

In addition, multiple measurement points were selected in two buildings. At each point, measurements were conducted both with and without broadcast sound. The results showed that the background noise levels ranged from 59.6 to 68.5 dB (A), and the SNR ranged from 0.6 to 11.3 dB. The measurement results and the distribution of points are presented in Table 2.

**Table 2.** Field measurement results and distribution of measurement points.

| Measurements Locations | | SPL | | SNR |
|---|---|---|---|---|
| | | with Broadcast Sound | Without Broadcast Sound | |
| Tianjin Railway Station | R1 | 67.2 | 75 | 7.0 |
| | R2 | 68.2 | 71.8 | 1.1 |
| | R3 | 68 | 72.8 | 3.1 |
| | R4 | 68.5 | 76.3 | 7.0 |
| | R5 | 67.3 | 71.2 | 1.6 |
| | R6 | 67 | 78.6 | 11.3 |
| | R7 | 61.2 | 70.5 | 8.8 |
| | R8 | 59.6 | 64.6 | 3.3 |
| | R9 | 60.2 | 69.3 | 8.5 |
| Binhai International Airport | R1 | 66.2 | 70.7 | 2.6 |
| | R2 | 68.2 | 71.8 | 1.1 |
| | R3 | 66.2 | 69.5 | 0.6 |
| | R4 | 61.6 | 67.8 | 5.0 |
| | R5 | 60.1 | 68.7 | 8.0 |
| | R6 | 59.7 | 68.6 | 8.3 |
| | R7 | 63.1 | 69.2 | 4.9 |

## 2.2. Experiment Design

To further examine the impact of echoes on speech intelligibility in extra-large spaces, a listening experiment was designed in which participants performed tasks under various conditions. The following sections describe the speech materials, experimental conditions, and detailed procedures.

### 2.2.1. Speech Materials

In this experiment, sentence lists were selected as the speech material, as they better reflect real listening conditions in extra-large spaces where public address systems typically deliver information in sentences rather than isolated words. The sentences were derived from the Chinese Word Matrix (Table 3) [23], with each sentence following the fixed syntactic structure 'name + verb + numeral + adjective + noun'. Based on this matrix, a total of 120 semantically unpredictable target sentences were randomly generated. Participants' accuracy in repeating the target sentences was recorded as a measure of speech intelligibility. Furthermore, to make the speech materials more representative of real conditions in public spaces, each target sentence was randomly preceded by different lead-in sentences (Figure 3), which consisted of phrases such as 'Next, you will hear…'.

**Table 3.** Base matrix of the CMN sentence test. The 120 target sentences used in the experiment were randomly generated accordingly.

| | Name | Verb | Numeral | Adjective | Noun |
|---|---|---|---|---|---|
| 0 | 郭毅 | 带走 | 一个 | 彩色的 | 板凳 |
| | Guoyi | took away | one | colourful | stool |
| 1 | 李锐 | 借来 | 两个 | 大号的 | 茶杯 |
| | Lirui | borrowed | two | large-sized | cup |
| 2 | 沈悦 | 看见 | 三个 | 很旧的 | 灯笼 |
| | Shengyue | looked | three | very old | lantern |
| 3 | 王石 | 留下 | 四个 | 便宜的 | 饭盒 |
| | Wangshi | kept | four | cheap | lunch-box |
| 4 | 徐敏 | 买回 | 五个 | 漂亮的 | 花瓶 |
| | Xumin | bought | five | beautiful | vase |

**Table 3.** *Cont.*

|  | Name | Verb | Numeral | Adjective | Noun |
|---|---|---|---|---|---|
| 5 | 杨硕 | 拿起 | 六个 | 普通的 | 戒指 |
|  | Yangshuo | picked up | six | ordinary | ring |
| 6 | 张伟 | 弄丢 | 七个 | 奇怪的 | 闹钟 |
|  | Zhangwei | lost | seven | strange | alarm-clocks |
| 7 | 郑贤 | 收好 | 八个 | 全新的 | 书包 |
|  | Zhengxian | put away | eight | brand-new | school-bag |
| 8 | 周明 | 需要 | 九个 | 特别的 | 水壶 |
|  | Zhouming | needed | nine | special | kettle |
| 9 | 朱婷 | 找出 | 十个 | 用过的 | 玩具 |
|  | Zhuting | found | ten | used | toy |

The combined sentences were evenly divided into six groups, with one group used for familiarization and the remaining five for the formal experiment. The five formal groups were arranged in six different random orders, yielding 30 lists with varied content or sentence sequences. For each experimental condition, one list was randomly selected and convolved with the corresponding impulse response to generate the stimuli.
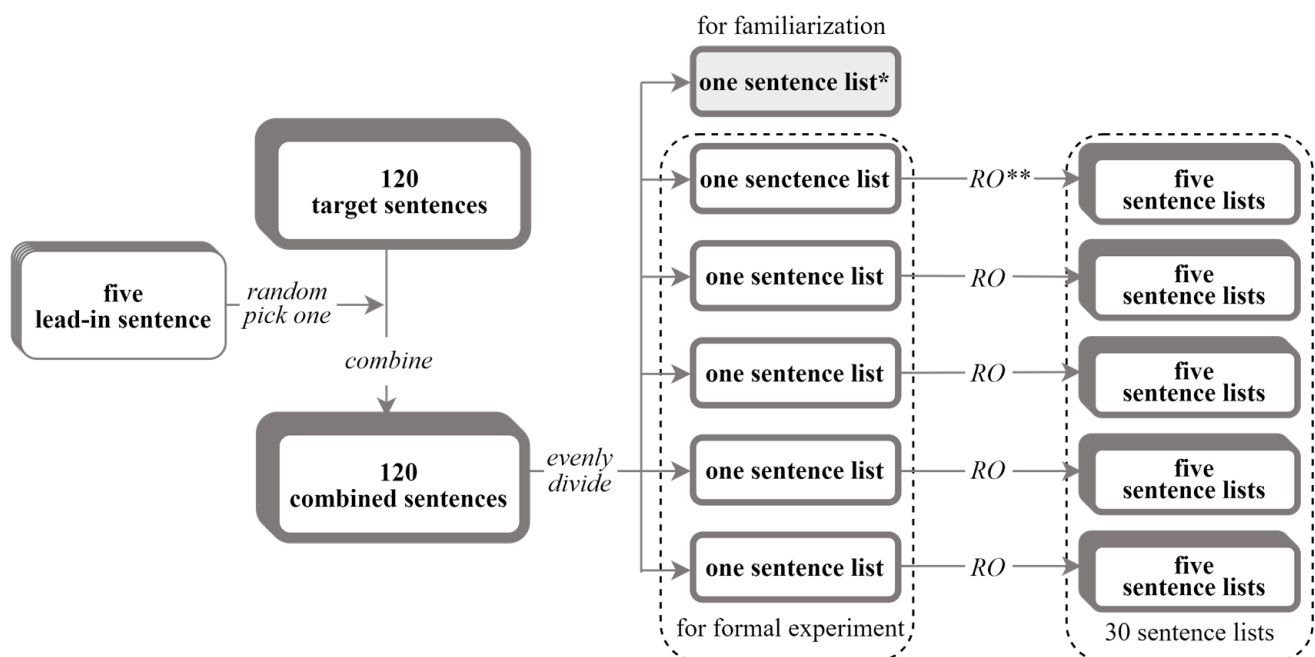


**Figure 3.** Generation of speech materials for the experiment. * One sentence list consisting of 20 sentences; ** Random Order, with the sentences in the list presented in a randomized sequence.

### 2.2.2. Experimental Conditions

A total of 160 experimental conditions (8 × 4 × 5) were tested by manipulating three variables: echo delay time (8 levels), echo strength (4 levels) and impulse responses (5 receiver positions).

On-site measurements revealed echo delays of up to approximately 1000 ms, leading us to test eight delay times (0, 100, 200, 300, 400, 500, 800, and 1000 ms) to fully capture this range. Echo strength was set to four levels (−10, −5, 0, and +5 dB) to account for potential sound-focusing effects caused by curved surfaces or domes in extra-large spaces [24], including cases where the echo was stronger than the direct sound. Impulse responses from five receiver positions in the two buildings (Figure 4) were used to reproduce realistic lis-

tening experiences in these spaces, while balancing the influence of architectural design differences among the positions.
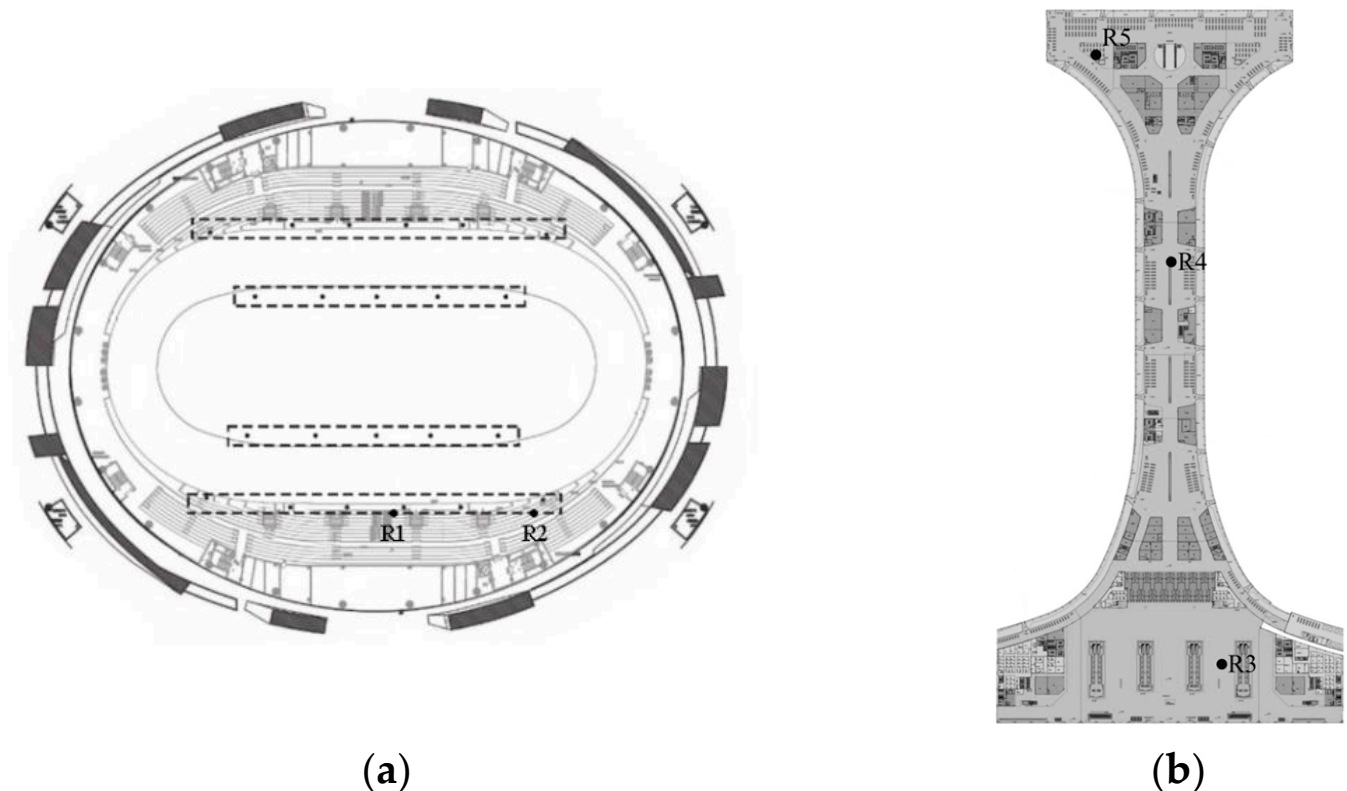


**Figure 4.** Impulse responses generated using Odeon 14, based on (**a**) the Tuanbo Lake Velodrome (R1, R2) and (**b**) Terminal 2 of Tianjin Binhai International Airport (R3–R5), both representative large-scale public buildings with spatial volumes exceeding 125,000 m$^3$.

Impulse responses were obtained using ODEON 14 with its built-in hybrid calculation method, and the acoustic characteristics of the building model were validated against in situ measurements, with deviations below one just noticeable difference (Appendix B). For each position, two nearby sound sources were selected, and their delays and overall gains were adjusted to generate room impulse responses that satisfied the acoustic requirements for the experimental stimuli. It should be noted that the impulse responses were employed solely to ensure that the sound stimuli more closely reflected realistic listening conditions. Subsequent analyses focused exclusively on echo delay time and echo strength.

All signals were presented at a sound pressure level of 65 dB (A), with a fixed signal-to-noise ratio (SNR) of 0 dB to avoid ceiling effects [13,15], which is consistent with our measurement results. The background noise was created by filtering and modulating pink noise according to field-measured noise in an extra-large space.

### 2.2.3. Subjects

The experimental participants were graduate students from Tianjin University, aged between 18 and 30 years. All were native Mandarin speakers with normal listening and speaking abilities, and they passed a pure-tone audiometric test. A total of 20 participants (9 males and 11 females) took part in the experiment.

The 160 experimental conditions were evenly divided into two groups, and each participant completed only one group. To ensure group homogeneity, 10 conditions were repeated across groups, resulting in 85 listening conditions per participant.

### 2.2.4. Experimental Procedure

Before the formal experiment, participants completed a hearing test and a familiarization session. During the formal experiment, 20 sentences (one sentence list) were presented under each condition. For each sentence, participants repeated the words they heard, and the experimenter recorded their responses. Speech intelligibility scores for each condition were derived from the proportion of correctly reproduced words. Upon completion of each condition, participants rated the listening difficulty on a 0–3 scale. The overall process is shown in Figure 5.
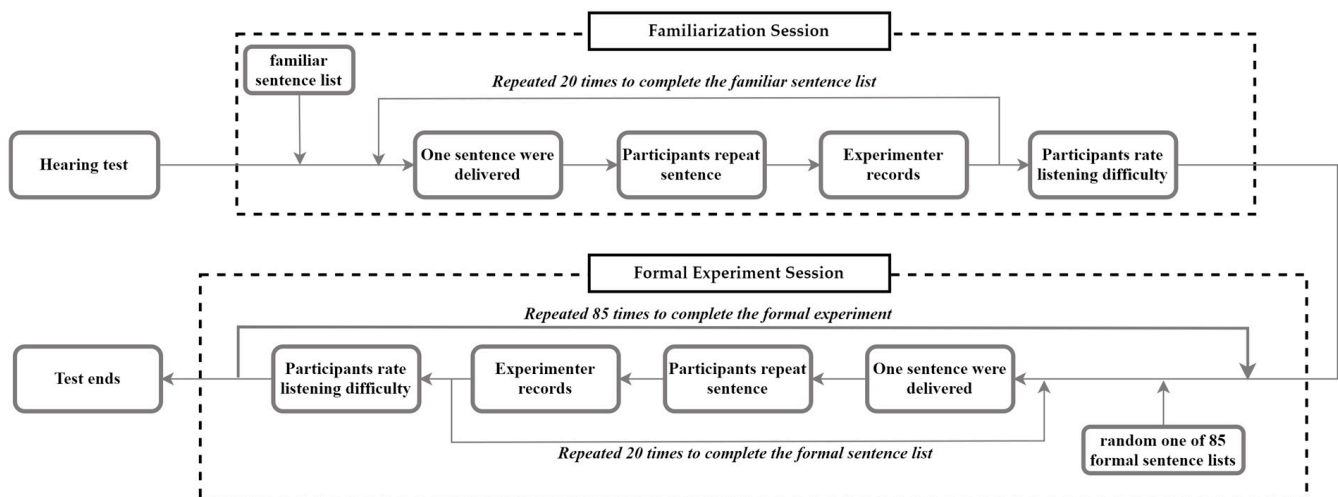


**Figure 5.** Overall experimental process.

The 85 sound stimuli were tested across three days. Each session lasted about 2.5 h and included 25–30 conditions, with breaks every 20–30 min to reduce fatigue. In total, the entire experiment took about 7 h per participant.

## 3. Results

### 3.1. Subjective Perceptions of the Acoustic Environment in Extra-Large Spaces

Figure 6 summarizes the responses to questionnaire. More than 90% of respondents rated the clarity of sound reinforcement systems as important or very important. However, dissatisfaction with the acoustic environment in the surveyed areas was prevalent: 47.7% of the respondents were not satisfied with the clarity of sound reinforcement systems, 40.8% indicated that announcements were not clear, and 49.6% considered the announcements insufficiently loud. These findings indicate a substantial gap between the perceived importance of announcement clarity and its actual performance, underscoring the need for targeted acoustic improvements in extra-large spaces.

According to responses of Q5, more respondents reported being affected by echo interference (35.2%) than those who did not (23.8%). Correlation analysis further revealed significant negative relationships between echo perception and the subjective ratings of announcement clarity ($r = -0.414$, $p < 0.01$) and satisfaction ($r = -0.374$, $p < 0.01$). Similarly, comparison of the Q6 scores (4–1 points assigned to the first through fourth choices) showed that inherent spatial reflections and multi-source loudspeaker interference were the most frequently perceived factors affecting announcement clarity, along with crowd conversations and low loudness. These results suggest that echo phenomena are indeed present in extra-large spaces and have a noticeable impact on speech intelligibility. More detailed data are presented in Appendix A.
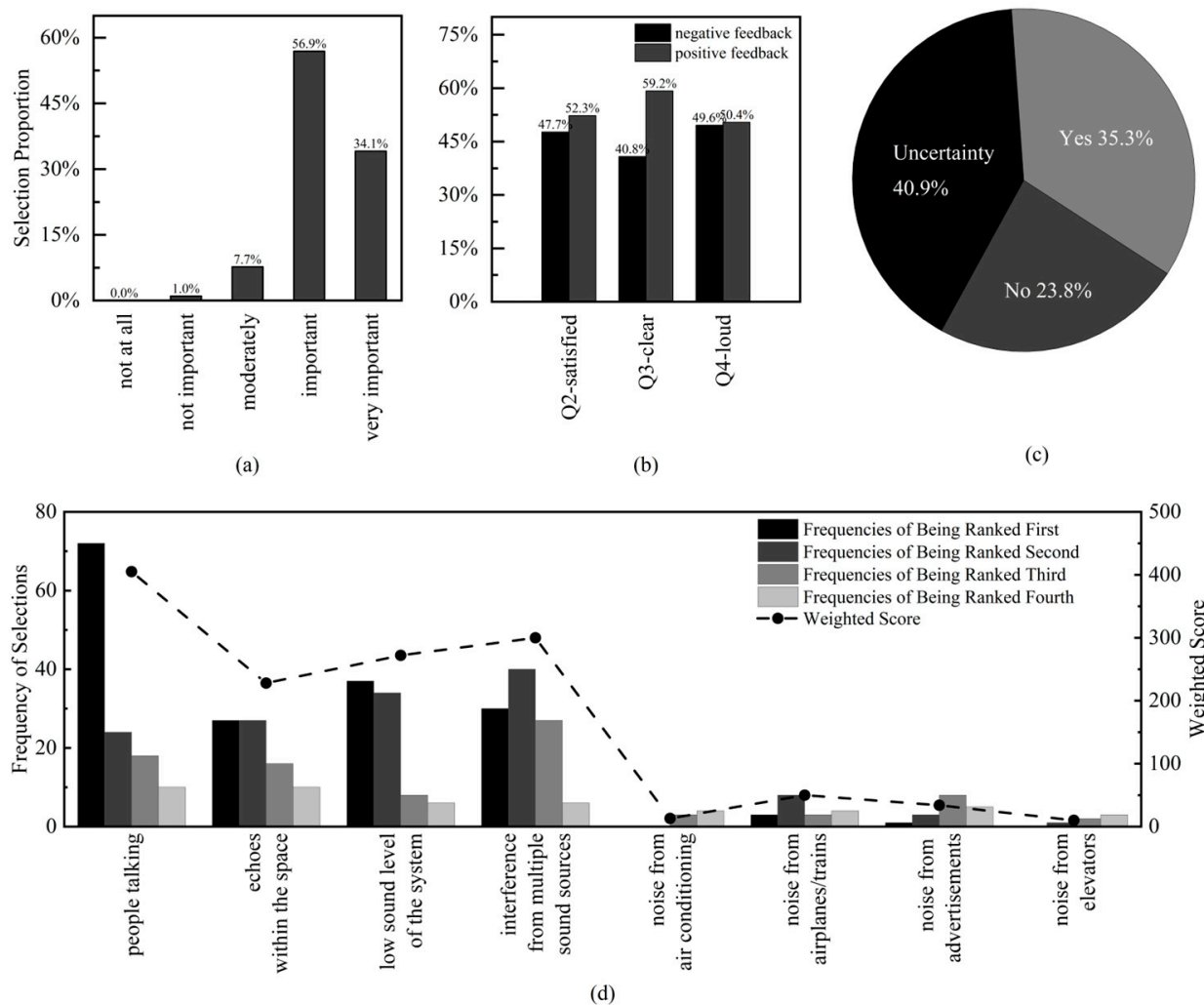
**Figure 6.** Statistics of the questions in the first part of the survey. (**a**) Respondents' perception of the importance of the sound reinforcement system (Q1); (**b**) Respondents' satisfaction with the clarity of the sound reinforcement system, as well as their evaluation of its clarity and loudness (Q2–Q4); (**c**) Respondents' perception of echo and delayed sounds (Q5); (**d**) Respondents' selection and ranking of factors affecting clarity (Q6).

### 3.2. Effects of Echo Interference on Speech Intelligibility

The variation in speech intelligibility across echo delay conditions is shown in Figure 7. A generalized linear mixed model (GLMM) revealed a significant main effect of echo delay on speech intelligibility ($p < 0.001$). Pairwise comparisons (Table 4) indicated that the delays could be classified into three groups: no echo (0 ms), short-delay echo (100 ms, 200 ms), and long-delay echo (300 ms, 400 ms, 500 ms, 800 ms, 1000 ms). Significant differences were found between these groups but almost no significant differences within them, except for the 400 ms condition. Long-delay echoes exceeding 300 ms clearly resulted in more severe effects. Although the lowest intelligibility occurred at 1000 ms, delays beyond 300 ms did not cause further significant deterioration. Quantitatively, the average drop in intelligibility from the no-echo to the short-delay group was about 5–6%, while the drop from the short-delay to the long-delay group was about 5–8%. These results suggest that speech intelligibility decreases as delay increases, and echoes with delays of 300 ms or longer substantially reduce speech intelligibility, but further increases in delay yield no additional significant loss.
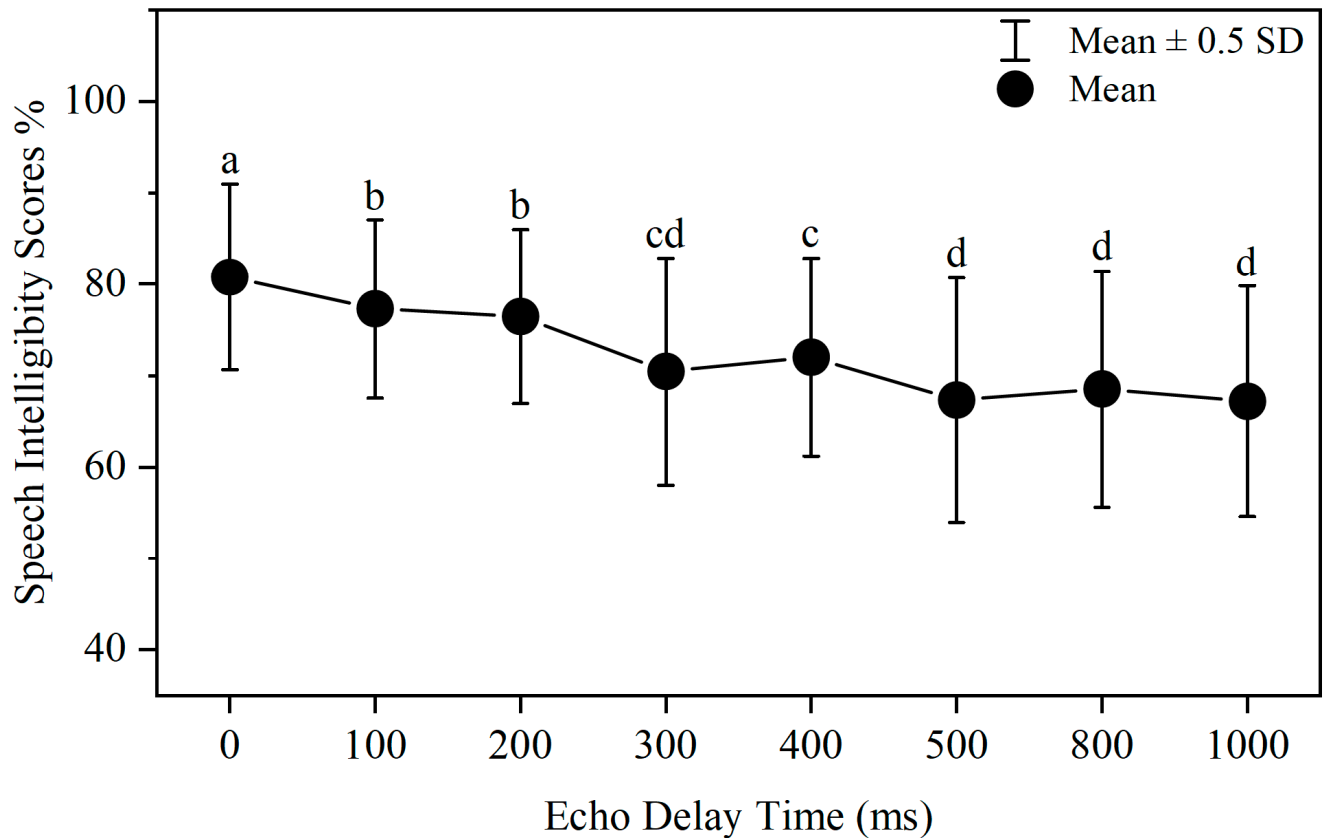
**Figure 7.** Variation in speech intelligibility across echo delay conditions. Values not sharing the same letter are significantly different ($p < 0.05$), while those sharing the same letter show no significant difference.

**Table 4.** Pairwise comparisons of speech intelligibility across echo delay conditions.

| | 0 ms | 100 ms | 200 ms | 300 ms | 400 ms | 500 ms | 800 ms | 1000 ms |
|---|---|---|---|---|---|---|---|---|
| 0 ms | \ | 5.1% ** | 5.8% ** | 12.9% ** | 10.7% ** | 14.5% ** | 13.5% ** | 14.8% ** |
| 100 ms | 5.1% ** | \ | 0.6% | 7.8% ** | 5.6% ** | 9.4% ** | 8.4% ** | 9.6% ** |
| 200 ms | 5.8% ** | 0.6% | \ | 7.1% ** | 4.9% ** | 8.7% ** | 7.8% ** | 9.0% ** |
| 300 ms | 12.9% ** | 7.8% ** | 7.1% ** | \ | −2.2% | 1.6% | 0.6% | 1.9% |
| 400 ms | 10.7% ** | 5.6% ** | 4.9% ** | −2.2% | \ | 3.8% ** | 2.8% * | 4.0% ** |
| 500 ms | 14.5% ** | 9.4% ** | 8.7% ** | 1.6% | 3.8% ** | \ | −1.0% | 0.2% |
| 800 ms | 13.5% ** | 8.4% ** | 7.8% ** | 0.6% | 2.8% * | −1.0% | \ | 1.2% |
| 1000 ms | 14.8% ** | 9.6% ** | 9.0% ** | 1.9% | 4.0% ** | 0.2% | 1.2% | \ |

Unmarked values indicate non-significant results; * $p \leq 0.05$, ** $p \leq 0.01$. Cells shaded in grey indicate non-significant results.

Echo strength also had a significant effect on speech intelligibility ($p < 0.001$), with the −10 dB and +5 dB conditions yielding higher scores, whereas the −5 dB and 0 dB conditions produced lower scores. This difference became increasingly pronounced with longer echo delays, reaching a maximum at 800 ms, where the intelligibility gap across conditions was up to 17% ($p < 0.001$). Even though the interaction between echo strength and echo delay was significant, the results across conditions consistently showed a three-category pattern (Figure 8): no echo (0 ms), short-delay echo (100 ms and 200 ms), and long-delay echo (300 ms, 400 ms, 500 ms, 800 ms, and 1000 ms).
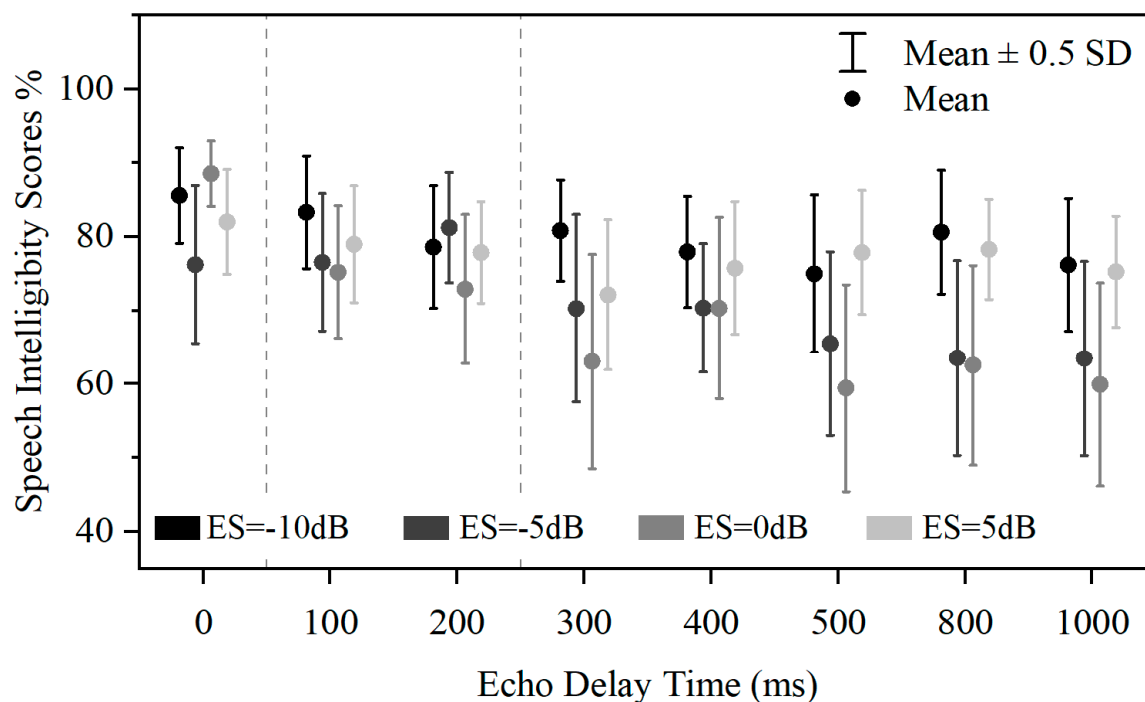
**Figure 8.** Speech intelligibility scores under different echo strength conditions (ES represents echo strength in the figure).

### 3.3. The Relation Between Speech Intelligibility and STI

The relationship between speech intelligibility and STI in extra-large spaces, together with the best-fitting third-order polynomial (Equation (1), $R^2 = 0.73$), are shown in Figure 9. Speech intelligibility increased rapidly with rising STI and reached its maximum level at around 0.4.
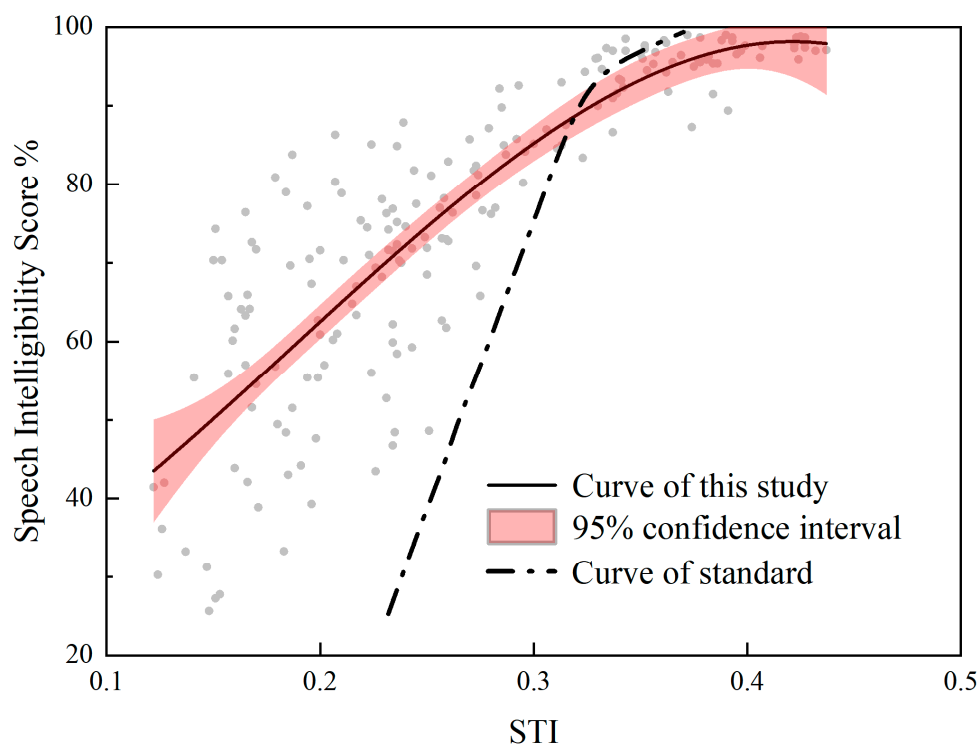


**Figure 9.** Relation between speech intelligibility scores and STI of this study, and curve of GB standard [25].

A comparison with the curve in the relevant standard [25], which was also obtained using the sentence test, is presented. The slopes of the two curves showed a clear difference. When the STI was below a certain threshold (approximately 0.32), the standard curve predicted lower speech intelligibility than that observed in extra-large spaces. Conversely, at higher STI levels, the standard curve predicted higher speech intelligibility than that observed in extra-large spaces. As a result, under most conditions, the GB curve failed to provide an accurate prediction of speech intelligibility in extra-large spaces.

$$\text{Speech Intelligibility} = 22.478 + 97.769\text{STI} + 843.402\text{STI}^2 - 1506.837\text{STI}^3 \qquad (1)$$

## 4. Discussion

### 4.1. The Impact of Mandarin Characteristics on Echo Interference

Based on the preceding analysis, echoes were classified into three groups: no echo (0 ms), short-delay echo (100 ms and 200 ms), and long-delay echo (300 ms, 400 ms, 500 ms, 800 ms, and 1000 ms). The degree of impact on speech intelligibility increased across these groups. It is noteworthy that when the echo delay reached 300 ms, speech intelligibility deteriorated significantly, and this effect did not worsen further with longer delays. This may be attributed to the combined effects of informational masking [26] and Mandarin expression: at a delay of around 300 ms, the direct sound and the echo differ by approximately one character, which causes substantial interference with speech recognition. When the delay exceeds 300 ms, the overlap between the direct sound and the echo decreases; however, the large discrepancy in speech content between the two signals may introduce additional confusion at the level of cognitive processing. However, in order to maximize the range of echo delays within the limited experimental duration, we did not include conditions such as 600 ms or 900 ms, which could also potentially result in character overlap. If speech intelligibility scores at these delays were found to be significantly lower than those of their neighboring conditions (e.g., 600 ms lower than 500 ms and 700 ms, or 900 ms lower than 800 ms and 1000 ms), it could provide stronger evidence that the impact of echoes on intelligibility is influenced by the characteristics of Mandarin expression. If such influencing factors are confirmed to exist, it would be worthwhile to investigate whether similar effects occur in other linguistic contexts.

### 4.2. STI Rating for Extra-Large Spaces

The speech intelligibility–STI curves derived from the two extra-large spaces showed clear differences from the current standard curve (Figure 9), indicating that long-delay echoes affect the prediction of speech intelligibility. This finding suggests that the prediction standards for speech intelligibility in extra-large spaces may need to be revised. Steeneken [27] divided the STI into five levels: Bad, Poor, Fair, Good, and Excellent, based on the relation between STI and speech intelligibility. The corresponding intelligibility thresholds were approximately 58.2%, 82.0%, 92.5%, and 97.1%. Therefore, according to the fitted relation curve between speech intelligibility and STI in extra-large spaces (Figure 8), the STI values corresponding to these thresholds were, respectively, 0.18, 0.28, 0.34, and 0.39. However, the sample size in this study was limited, and establishing the relationship between speech intelligibility and STI in extra-large spaces, as well as determining STI thresholds, requires a larger dataset for reference.

*4.3. Comparison Between Listening Satisfaction and Speech Intelligibility*

In addition to speech intelligibility data, participants' subjective ratings of listening difficulty under different echo conditions were also collected using a four-point scale (0–3). Ratings of 0 and 1 were considered to be positive evaluations of listening difficulty, which better reflected participants' satisfaction with the acoustic environment. A comparison (Figure 10) revealed that, across all conditions, the proportion of participants considering the listening environment as acceptable was consistently lower than the proportion of correct responses in speech intelligibility. Therefore, in designing a more favorable acoustic environment for extra-large spaces, it is essential to consider not only speech intelligibility but also users' subjective perceptions.
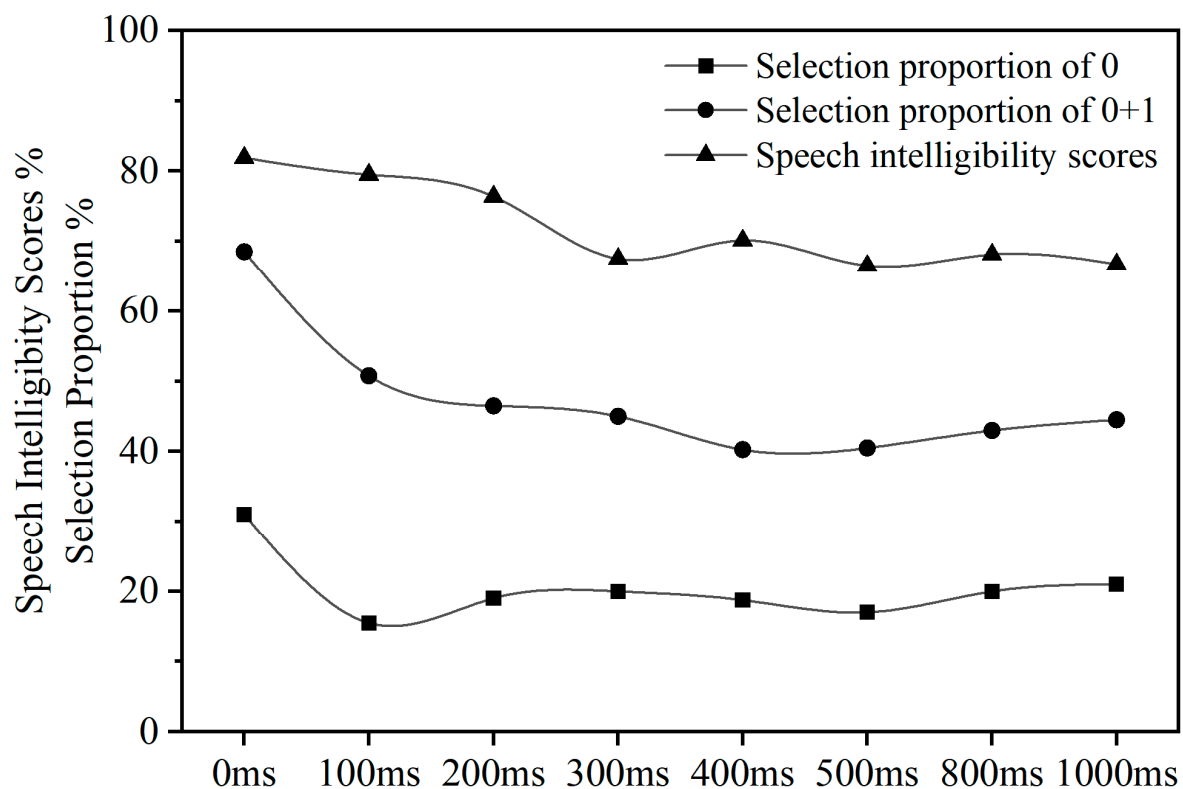


**Figure 10.** Proportion of participants giving positive ratings (0 or 1) under different echo conditions and corresponding speech intelligibility scores.

*4.4. Limitations and Future Research*

In this study, to control the experiment duration, echo delays were not subdivided in greater detail, and some conditions that might have supported the specific characteristics of Mandarin in echoic environments, such as 600 ms and 900 ms, were not included. This should be addressed and further examined in future research. In addition, as the classification of listening satisfaction adopted in this study was relatively preliminary, further research is needed to more accurately capture individuals' auditory experiences in echoic extra-large spaces.

We look forward to future studies focusing on long-delay echoes in extra-large spaces. The number of participants in the listening experiment was 20, and all were Mandarin speakers within a similar age range, which could be further expanded and generalized. Similarly, the STI–speech intelligibility curve obtained from the two building cases deviates from the standard. Establishing a prediction curve suitable for extra-large spaces therefore requires multiple studies and additional data.

Building upon the current findings, future studies could incorporate advanced acoustic and spatial analysis methods to further clarify the mechanisms by which specific echo delays and strengths influence speech intelligibility. Furthermore, research could focus on strategies to improve speech intelligibility under long-delay echo conditions. Developing speech signals suitable for long-delay echo conditions and optimizing loudspeaker arrangements are considered important strategies for adjusting subjective auditory perception in other special acoustic environments [28–30], and thus represent research directions worthy of further attention.

## 5. Conclusions

In this study, a questionnaire survey and laboratory experiments were conducted to investigate the influence of long-delay echo interference on speech intelligibility in extra-large spaces.

The survey results revealed that echo interference was a major factor in public dissatisfaction with announcement clarity. The experimental results showed that echoes could be classified into three groups: no echo (0 ms), short-delay (100–200 ms), and long-delay ($\geq$300 ms), with intelligibility decreasing across groups. Echoes of 300 ms or longer substantially reduced intelligibility, while further increases in delay had little additional effect. Echo strength also influenced intelligibility, with greater reductions observed when the reflected sound was equal to or 5 dB lower than the direct sound.

Based on these results, the speech intelligibility–STI relationship in extra-large spaces with long-delay echoes clearly deviates from the standard curve. This indicates that long-delay echoes significantly impair intelligibility and prediction accuracy, suggesting that the standard curve needs revision for such environments.

**Author Contributions:** Conceptualization, H.M., C.W. and S.D.; methodology, H.M., C.W. and S.D.; software, S.D.; validation, C.W. and S.D.; formal analysis, W.W. and S.D.; investigation, W.W. and S.D.; resources, H.M., C.W., W.H. and B.H.; data curation, W.W. and S.D.; writing—original draft preparation, W.W., S.D., C.W. and H.M.; writing—review and editing, W.W. and H.M.; visualization, W.W. and S.D.; supervision, H.M.; project administration, H.M.; funding acquisition, H.M. and C.W. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to the data obtained will not be publicly disclosed or shared with any other institutions or individuals. Secondly, the experimental conditions are designed to be non-invasive and safe, posing no risk of harm to participants. Finally, all par-ticipants signed informed consent prior to taking part in the study, having been thoroughly briefed on the complete content of the experiment. Under these circumstances, School of Architecture, Tianjin University grants an exemption from requiring formal ethics approval.

**Data Availability Statement:** Methodological details and processed datasets are available from the corresponding author upon request.

**Conflicts of Interest:** Authors Wenlin Hu and Bin He were employed by the company National Engineering Research Center for Digital Construction and Evaluation of Urban Rail Transit, China Railway Design Corporation and National-Local Joint Engineering Laboratory for Rail Transit Survey and Design, China Railway Design Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Appendix A. Data Collected from the Questionnaire Survey

This part serves as a supplementary section to Section 3.1, presenting the detailed survey results in Tables A1 and A2.

**Table A1.** Survey data from valid questionnaires, excluding Q6 and Q8.

| Question | Option | Number of Respondents |
|---|---|---|
| Q1 | Not at all | 0 |
| | Not important | 2 |
| | Moderately | 15 |
| | Important | 110 |
| | Very important | 66 |
| Q2 | Not at all | 5 |
| | Not satisfy | 18 |
| | Moderately | 70 |
| | Satisfy | 78 |
| | Very satisfy | 22 |
| Q3 | Not at all | 5 |
| | Not clear | 25 |
| | Moderately | 49 |
| | Clear | 96 |
| | Very clear | 18 |
| Q4 | Not at all | 2 |
| | Not loud | 30 |
| | Moderately | 64 |
| | Loud | 81 |
| | Very loud | 16 |
| Q5 | No | 46 |
| | Uncertainly | 79 |
| | Yes | 68 |
| Q7 | Male | 105 |
| | Female | 88 |
| Q9 | <20 | 21 |
| | 20–40 | 89 |
| | 40–60 | 81 |
| | >60 | 2 |
| Q10 | <15 min | 61 |
| | 15–60 min | 110 |
| | >60 min | 22 |

**Table A2.** Questionnaire statistics for Q6.

| Option | Frequencies of Being Ranked | | | |
|---|---|---|---|---|
| | First | Second | Third | Fourth |
| people talking | 72 | 24 | 18 | 10 |
| echoes within the space | 27 | 27 | 16 | 10 |
| low sound level of the system | 37 | 34 | 8 | 6 |
| interference from multiple sound sources | 30 | 40 | 27 | 6 |
| noise from air conditioning | 0 | 0 | 3 | 4 |
| noise from airplanes/trains | 3 | 8 | 3 | 4 |
| noise from advertisements | 1 | 3 | 8 | 5 |
| noise from elevators | 0 | 1 | 2 | 3 |

## Appendix B. Details of Acoustic Simulation

This section provides supplementary information about the acoustic simulation, including model establishment, parameter configuration, surface material assignment, and model verification.

To ensure computational efficiency without compromising accuracy, the geometric complexity of the models was reasonably controlled, and excessively detailed or small surfaces were simplified when appropriate.

In this study, the number of rays was set to 1,000,000, and the impulse response length was set to 10,000 ms.

Surface Material Assignment: Both absorption and scattering coefficients were assigned according to the actual acoustic and surface characteristics of each area in the two buildings. The absorption coefficients were created manually based on the measured or literature data, while the scattering coefficients were determined according to the roughness of each surface. The detailed absorption and scattering coefficients used for all surfaces are provided in Table A3.

Model Verification and Error Correction: After completing the parameter and material settings, the reliability of the models was validated by comparing simulated and measured acoustic parameters. The differences between simulation and measurement results for the main parameters—particularly T30, D50, and C80—at 500 Hz and 1000 Hz were all within one Just Noticeable Difference (JND). According to ISO 3382-1 [31], the JND values are 5% for T30 and D50, and 1 dB for C80.

**Table A3.** Absorption Coefficients and Scattering Coefficients of Surface Materials in the Models.

| Model | Surface | Frequency (Hz) | | | | | | Scattering Coefficient |
|---|---|---|---|---|---|---|---|---|
| | | 125 Hz | 250 Hz | 500 Hz | 1000 Hz | 2000 Hz | 4000 Hz | |
| Tuanbo Lake Velodrome | Ceiling | 0.32 | 0.38 | 0.48 | 0.60 | 0.60 | 0.58 | 0.3 |
| | Wall | 0.1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.3 | 0.25 |
| | Stand | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.05 |
| | Audience seats | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.5 |
| | Side windows | 0.18 | 0.06 | 0.04 | 0.05 | 0.02 | 0.02 | 0.05 |
| | Lounge area | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.25 |
| | Track surface | 0.15 | 0.18 | 0.25 | 0.25 | 0.2 | 0.15 | 0.05 |
| Terminal 2 of Tianjin Binhai International Airport | Ceiling | 0.36 | 0.36 | 0.58 | 0.68 | 0.58 | 0.38 | 0.3 |
| | Interior walls | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.25 |
| | Seating area | 0.27 | 0.28 | 0.36 | 0.37 | 0.35 | 0.34 | 0.5 |
| | Office area walls | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.05 |
| | Commercial area walls | 0.2 | 0.2 | 0.25 | 0.25 | 0.25 | 0.25 | 0.05 |
| | Floor | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.05 |
| | Glass curtain wall | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |

**Table A4.** Verification of Simulation Accuracy Against Field Measurements.

| Model | Measurement Point | T30 (s) | | D50 (%) | | C80 (dB) | |
|---|---|---|---|---|---|---|---|
| | | Simulated | Measured | Simulated | Measured | Simulated | Measured |
| Tuanbo Lake Velodrome | R1 | 3.095 | 3.055 | 62.5 | 60.4 | 3.6 | 3.995 |
| | R2 | 3.055 | 3.135 | 59.0 | 60.7 | 3.3 | 2.96 |
| Terminal 2 of Tianjin Binhai International Airport | R3 | 1.14 | 1.105 | 87.0 | 87.5 | 10.0 | 11.45 |
| | R4 | 1.765 | 1.69 | 73.0 | 73.65 | 6.8 | 6.3 |
| | R5 | 1.525 | 1.585 | 88.0 | 86.1 | 9.5 | 9.245 |

# References

1. Shimokura, R.; Soeta, Y. Evaluation of Speech Intelligibility of Sound Fields in Underground Stations. *Acoust. Sci. Technol.* **2011**, *32*, 73–75. [CrossRef]
2. Fujikawa, T.; Aoki, S. An Escape Guiding System Utilizing the Precedence Effect for Evacuation Signal. *J. Acoust. Soc. Am.* **2013**, *133*, 3362. [CrossRef]
3. Kootwijk, P.A.A. The Speech Intelligibility of the Public Address Systems at 14 Dutch Railway Stations. *J. Sound Vib.* **1996**, *193*, 433–434. [CrossRef]
4. Liu, H.; Ma, H.; Wang, C.; Kang, J. Prediction Model of Crowd Noise in Large Waiting Halls. *J. Acoust. Soc. Am.* **2022**, *152*, 2001. [CrossRef]
5. Kotus, J.; Szwoch, G. Speech Intelligibility Improvement for Public Address Systems in Noisy Environments Based on Automatic Gain Selection in Octave Bands. *Appl. Acoust.* **2025**, *235*, 110683. [CrossRef]
6. Huang, W.; Peng, J.; Xie, T. Study on Chinese Speech Intelligibility Under Different Low-Frequency Characteristics of Reverberation Time Using a Hybrid Method. *Arch. Acoust.* **2023**, *48*, 151–157. [CrossRef]
7. Li, X.; Zhao, Y. Exploring Factors Influencing Speech Intelligibility in Airport Terminal Pier-Style Departure Lounges. *Buildings* **2025**, *15*, 426. [CrossRef]
8. Pan, L.; Lu, W. The Discussions of the Code for Acoustical Design of Gymnasiums. *Audio Eng.* **2016**, *30*, 11–14. [CrossRef]
9. Haas, H. The Influence of a Single Echo on the Audibility of Speech. *J. Audio Eng. Soc.* **1972**, *20*, 146–159.
10. Wang, C.; Ma, H.; Wu, Y.; Kang, J. Characteristics and Prediction of Sound Level in Extra-Large Spaces. *Appl. Acoust.* **2018**, *134*, 1–7. [CrossRef]
11. Lochner, J.P.A.; Burger, J.F. The Influence of Reflections on Auditorium Acoustics. *J. Sound Vib.* **1964**, *1*, 426–448,IN15,449–454. [CrossRef]
12. Warzybok, A.; Rennies, J.; Brand, T.; Doclo, S.; Kollmeier, B. Effects of Spatial and Temporal Integration of a Single Early Reflection on Speech Intelligibility. *J. Acoust. Soc. Am.* **2013**, *133*, 269–282. [CrossRef] [PubMed]
13. Sakamoto, S.; Cui, Z.; Miyashita, T.; Morimoto, M.; Suzuki, Y.; Sato, H. Effects of Inter-Word Pauses on Speech Intelligibility under Long-Path Echo Conditions. *Appl. Acoust.* **2018**, *140*, 263–274. [CrossRef]
14. Sun, G. Main Points of Public-Address System Designing on Gymnasium and Stadium. *Audio Eng.* **2012**, *36*, 1–2+20. [CrossRef]
15. Cui, Z.; Sakamoto, S.; Morimoto, M.; Suzuki, Y.; Sato, H. Effect of Word Familiarity on Word Intelligibility of Four Continuous Words under Long-Path Echo Conditions. *Appl. Acoust.* **2017**, *124*, 30–37. [CrossRef]
16. Houtgast, T.; Steeneken, H.J.M. The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility. *J. Acoust. Soc. Am.* **1973**, *54*, 557. [CrossRef]
17. Houtgast, T.; Steeneken, H.J.M. A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria. *J. Acoust. Soc. Am.* **1985**, *77*, 1069–1077. [CrossRef]
18. Steeneken, H.J.; Houtgast, T. A Physical Method for Measuring Speech-Transmission Quality. *J. Acoust. Soc. Am.* **1980**, *67*, 318–326. [CrossRef] [PubMed]
19. Kang, J. Comparison of Speech Intelligibility between English and Chinese. *J. Acoust. Soc. Am.* **1998**, *103*, 1213–1216. [CrossRef]
20. Kang, J. *Acoustics of Long Spaces: Theory and Design Guidance*; Thomas Telford: London, UK, 2002; ISBN 978-0-7277-3013-8.
21. Zhu, P.; Mo, F.; Kang, J. Relationship Between Chinese Speech Intelligibility and Speech Transmission Index Under Reproduced General Room Conditions. *Acta Acust. United Acust.* **2014**, *100*, 880–887. [CrossRef]
22. Liu, H.; Hui, M.; Jian, K.; Chao, W. The Speech Intelligibility and Applicability of the Speech Transmission Index in Large Spaces. *Appl. Acoust.* **2020**, *167*, 107400. [CrossRef]
23. Hu, H.; Xi, X.; Wong, L.L.N.; Hochmuth, S.; Warzybok, A.; Kollmeier, B. Construction and Evaluation of the Mandarin Chinese Matrix (CMNmatrix) Sentence Test for the Assessment of Speech Recognition in Noise. *Int. J. Audiol.* **2018**, *57*, 838–850. [CrossRef]
24. Li, H.; Chen, J. Acoustic Design of the Dome. *Build. Entertain. Stad.* **2023**, *2023*, 33–39. [CrossRef]
25. *GB/T 12060.16-2017*; Sound System Equipment. Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index. Standards Press of China: Beijing, China, 2018.
26. Gutschalk, A.; Micheyl, C.; Oxenham, A.J. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. *PLoS Biol.* **2008**, *6*, e138. [CrossRef]
27. Steeneken, H.J.M.; Houtgast, T. Validation of the Revised STIr Method. *Speech Commun.* **2002**, *38*, 413–425. [CrossRef]
28. Wen, M.; Ma, H.; Wang, C. Older Adults' Perception of Urgency: Effects of Simple Temporal Patterns in Auditory Signals. *J. Acoust. Soc. Am.* **2025**, *158*, 2319–2330. [CrossRef]
29. Winkler, A.; Warkentin, L.; Denk, F.; Husstedt, H.; Sankowksy-Rothe, T.; Blau, M.; Holube, I. Reference Speech-Recognition Curves for a German Monosyllabic Test in Noise: Effects of Loudspeaker Configuration and Room Acoustics. *Int. J. Audiol.* **2025**, *64*, 695–704. [CrossRef] [PubMed]

30. Hodoshima, N. Effects of Urgent Speech and Congruent/Incongruent Text on Speech Intelligibility for Older Adults in the Presence of Noise and Reverberation. *Speech Commun.* **2021**, *134*, 12–19. [CrossRef]

31. *ISO 3382-1:2009*; Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces. International Organization for Standardization: Geneva, Switzerland, 2009.