OPEN ACCESS Freely available online

CC BY

S. Sun, S. Chen, and Y. Shen, "The Effects of Lighting on Perceptual Sound Quality in Subjective Listening Tests," *J. Audio Eng. Soc.*, vol. 73, no. 5, pp. 293–302 (2025 May). https://doi.org/10.17743/jaes.2022.0201.

The Effects of Lighting on Perceptual Sound Quality in Subjective Listening Tests

SHUYUAN SUN,* *AES Student Member*, **SIMIAO CHEN, AND YONG SHEN,** *AES Member*
(sunshuyuanly@gmail.com) (schen@smail.nju.edu.cn) (yshen@nju.edu.cn)

Key Laboratory of Modern Acoustics (Ministry of Education), and Institute of Acoustics, Nanjing University, 22 Hankou Road, Nanjing, 210093, People's Republic of China

In subjective listening tests, listening conditions can significantly affect the measured results. Although there are some established standards, quantitative specifications for lighting level are rarely disclosed. This raises a question: does the lighting environment affect the perceived sound quality? In some audio-related industries, it is even widely believed that a darker environment is more conducive to improving the subjective sound quality. However, this conclusion clearly lacks convincing experimental data support and a scientific theoretical basis. In this study, a series of listening tests on preference sound quality were conducted under different lighting environments. The analyses on overall rating, program material, and the dispersion of raw data were presented. The results provide evidence that there are no significant effects of lighting on perceived overall sound quality. However, the unexpected significant effects related to program material were observed in some experiments. Additionally, the dispersion of ratings could be affected by the lighting level. These effects may contribute to the common misconception that a dim listening environment can improve perceived sound quality. More information and guidance toward conducting accurate and valid listening tests can be provided by comprehensive understanding of these effects.

0 INTRODUCTION

In subjective listening tests of perceptual sound quality, listening conditions can have a significant effect on the measured results [1]. Factors such as the size ratio and volume of the listening space [2], reflected sound [3–7], background noise [8], playback sound pressure level or loudness [9, 10], and reverberation time [11, 12] can all influence the outcomes. In addition, the results can be affected by loudspeaker position and placement [13, 14]. These valuable studies have led to the development of certain standards regarding the listening room [15, 16].

However, there are few clear definitions and quantitative specifications for lighting level requirements in these standards. Although some studies on subjective perception in audio and video interactions have been published [17–22], most publications mainly focus on changing auditory perception through visual stimulation. Furthermore, the lighting requirements for most standard acoustic listening tests still refer to the standard [23] for subjective evaluation of video quality, and there are few studies on the effects of lighting in subjective listening tests. In some audio-related industries, it is even widely believed that a darker listening environment is more conducive to improving the subjective perceived sound quality of loudspeakers. Undoubtedly, this conclusion clearly lacks convincing experimental data support and a scientific theoretical basis. The scarcity of research on this issue not only hinders the development of more scientific and comprehensive standards but also imposes constraints on conducting more reliable and accurate subjective listening experiments. Notably, Bech [1, 24] provided a preliminary but extremely instructive discussion on the impact of lighting on subjective sound quality evaluation.

In this study, a series of listening tests on preference sound quality were conducted. Listeners were required to comparatively evaluate five loudspeakers under different lighting environments. All experiments were double-blind listening tests conducted in a standard listening room. The purpose of the study is to examine the effects of lighting in subjective listening tests. Following this introduction, the present paper describes the listening experiments in SEC. 1; the main results and comparative analyses in SECS. 2 and 3, respectively; and conclusions in SEC. 4. Further works are advocated in SEC. 4.

*To whom correspondence should be addressed, email: sunshuyuanly@gmail.com

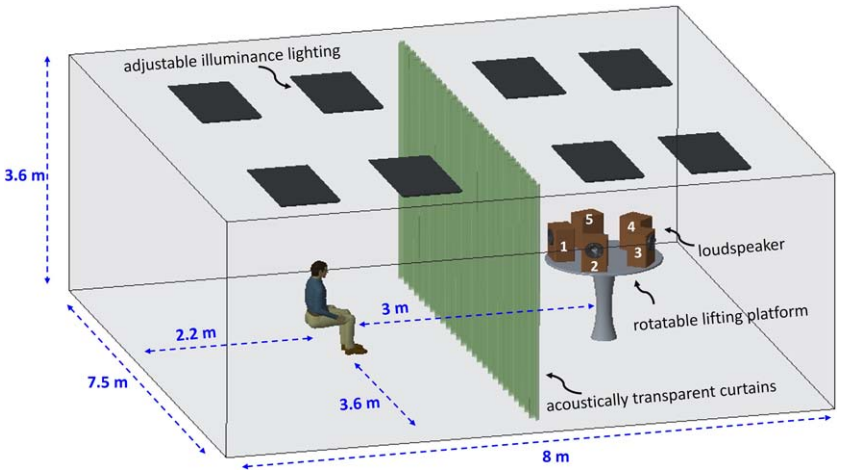


Fig. 1. An overview of the experimental setup presented in this article.

1 LISTENING EXPERIMENTS

1.1 Program Material and Listening Panel

According to a previous study [25], the following four programs were selected for the listening tests. These programs have sufficient bandwidth and spectral density to reveal the difference among the evaluated loudspeakers:

- JT: James Taylor, "That's Why I'm Here";
- JW: Jennifer Warnes, "Bird on a Wire";
- LF: Little Feat, "Hangin' on to the Good Times"; and
- TC: Tracy Chapman, "Fast Car."

The test subjects consisted of ten listeners, including six males and four females, ranging in age from 20 to 36 years. All reported no hearing damage. They had backgrounds in acoustics, either as professionals in the audio-related industry or students specializing in acoustics, and all had extensive experience in listening experiments.

Table 1. Description of the evaluated loudspeakers.

Number	Description
Loudspeaker 1	2-way vented box system, enhancing more bass than regular loudspeakers
Loudspeaker 2	Bookshelf speaker with a closed box design, featuring one satellite driver and a mid-bass driver
Loudspeaker 3	Portable speaker featuring a closed box design with one full-range unit
Loudspeaker 4	3-way vented box system incorporating a woofer, a midrange, and a diamond dome tweeter
Loudspeaker 5	2-way monitor speaker with bass reflex design, emphasizing reproduction precision, and sonic details rather than flattering sound

1.2 Experimental Setup

All experiments in this article were conducted in a listening room, which was rectangular with a volume of 216 m^3 (8 × 7.5 × 3.6 m). The measured reverberation time RT_{60} falls from 0.5 to 0.2 s between 100 Hz and 4 kHz. The listening distance between loudspeaker and listener was 3 m. The listening location was 2.2 m away from the rear wall and 3.6 m away from the side wall. To ensure that the acoustic center of the loudspeaker was nearly on the same horizontal plane as the listeners' ears, the evaluated loudspeakers were placed on a rotatable lifting platform, which was positioned along the same axis as the listener.

Listeners were required to comparatively evaluate five loudspeakers (Loudspeakers 1, 2, 3, 4 and 5) in the listening tests. The selected loudspeakers varied in quality, ranging from lower to higher levels. Table 1 gives a summarized description of the loudspeakers. The ceiling of the listening room was equipped with carefully controlled lighting equipment. The optical illuminance of this lighting system was adjustable to achieve different lighting conditions. The experimental setup presented in this article is illustrated in Fig. 1.

1.3 Experimental Design

A total of five listening tests were conducted. Each test had the same listening conditions and testing methods, except for the varying lighting conditions. The lighting conditions are described in Table 2. It is important to note that the selection of these lighting levels was within the subjectively acceptable range for the listeners and did not cause significant discomfort. Regarding subjective perception under different lighting levels, to some extent, the results of brightness assessment in furnished small rooms or offices [26–28] can be referenced: most participants felt that lighting above 300 lux was too bright, while lighting below 10 lux was rated as too dim. It should be noted that the order in which the listeners participated in the different listening tests was random, meaning that the lighting conditions were

Table 2. Lighting conditions in different listening tests.

Test Serial Number	Lighting Level
1	Illuminance = 380 lux
2	Illuminance = 250 lux
3	Illuminance = 150 lux
4	Illuminance = 60 lux
5	Illuminance = 6 lux

Note. The chromaticity in all experiments was D_{65}, as recommended in [23].

not presented in a sequence from bright to dark or from dark to bright. Additionally, since not all listeners completed all the listening tests in one session on the same day, the order in which the lighting levels were presented varied for each listener.

Each listening test involved comparison category rating on different loudspeakers via monophonic system and focused on preference sound quality without evaluating spatial information. All listeners completed eight trials (4 programs × 2 observations) in each listening test. For each trial, listeners were required to comparatively evaluate five loudspeakers on an 11-point preference scale [29–33]. Listeners could control the rotation platform through a laptop to rotate in a clockwise direction, thereby switching between different loudspeakers for evaluation in each trial. The height of platform was calibrated for each evaluated loudspeaker prior to the experiments, and it was adaptively adjusted during the listening tests. The time interval for switching between any two loudspeakers was within 2 s. For example, the maximum time of 2 s was needed for switching from Loudspeaker 1 to Loudspeaker 5, as shown in Fig. 1. This interval met the assumption that human short-term memory lasted on the range of 3 to 5 s [34]. And there were no time limits imposed on the listener to complete the evaluation.

The preference scale included semantic descriptors at every second interval labeled as really dislike (1), dislike (3), neither like nor dislike (5), like (7), and really like (9). The preference scale was with a resolution of one decimal place. Generally, a strong preference between evaluated objects was indicated by a separation in ratings of $\geq$2 points, a moderate preference $\geq$1 points, and an almost same preference $\leq$0.5 rating. Listeners were free to switch between each loudspeaker, pause the music at any time, and simultaneously complete the evaluation through the software application on a laptop, which was modified from a MATLAB interface for MUSHRA listening tests [35], as shown in Fig. 2.

Noting that in order to ensure the randomness of the loudspeakers and program materials presented during the experiments, the labels "Speakers A" to "Speaker E" and "Program 1/4" to "Program 4/4" displayed in the graphical user interface (GUI) for different listening trials corresponded to different physical loudspeakers or programs. For instance, in one trial, the speakers labeled as A, B, C, D, and E in the GUI actually corresponded to Loudspeakers

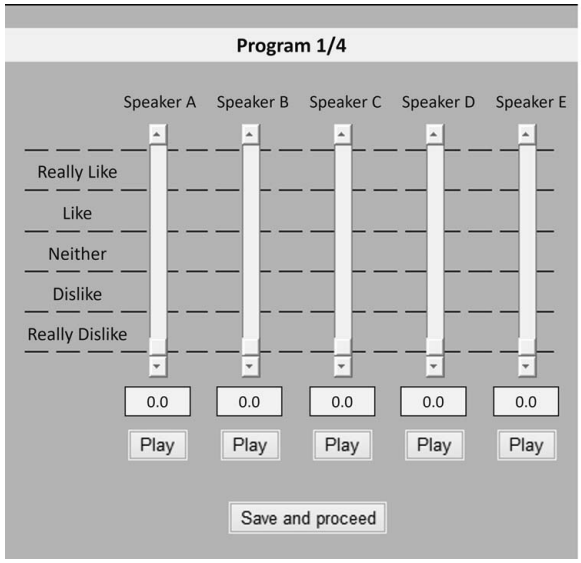


Fig. 2. The GUI used in the listening tests.

1, 2, 3, 4, and 5, respectively; whereas in another trial, the same labels corresponded to Loudspeakers 3, 4, 1, 5, and 2.

Listeners could also adjust the illuminance of the laptop at any time to ensure that the screen brightness remained at a comfortable level under different lighting conditions. Loudspeakers were calibrated based on ITU-R 1770 [36], with an average playback level of 80 dB (A-weighted). The listening duration did not exceed 30 min per day for each listener to avoid the negative influence of listener fatigue [37, 38]. All listening tests were double blind, meaning that acoustically transparent curtains were used to conceal the loudspeakers, and the order of loudspeaker presentation in each trial was randomized.

2 RESULTS

The main purpose of most listening tests is to explore the relationship between several independent variables and a single dependent variable. Typically, the relationship between the dependent and independent variables is analyzed separately for each dependent variable. Thus, the univariate analysis method of analysis of variance (ANOVA) is most commonly applied.

In the present paper, the normality of the raw data was firstly tested using the Kolmogorov-Smirnov method, and homogeneity of variances was tested using Levene's test. Also, deviations from normal distribution were visually inspected. Although the requirements of applying ANOVA were not completely fulfilled, some small violations of the assumptions could be neglected for the robustness of ANOVA.

Afterward, a four-way ANOVA was performed. The fixed factors were Program (four levels), Observation (two levels), Lighting (five levels), and Loudspeaker (five levels), with preference rating as the dependent variable. Additionally, all high-order interaction effects related to Loudspeaker were also considered.

Table 3. ANOVA analysis results.

Source	df	F	Significance
Lighting	4	1.533	0.19
Observation	1	0.134	0.714
Program	3	4.187	0.006
Loudspeaker	4	557.149	$\leq$ 0.001
Loudspeaker × Lighting	16	1.259	0.215
Loudspeaker × Program	12	3.121	$\leq$ 0.001
Loudspeaker × Observation	4	1.447	0.216
Loudspeaker × Lighting × Program	60	0.384	1.000
Loudspeaker × Lighting × Observation	20	1.360	0.132
Loudspeaker × Program × Observation	15	0.465	0.958
Loudspeaker × Lighting × Program × Observation	60	0.308	1.000

Note. Dependent variable: preference rating.

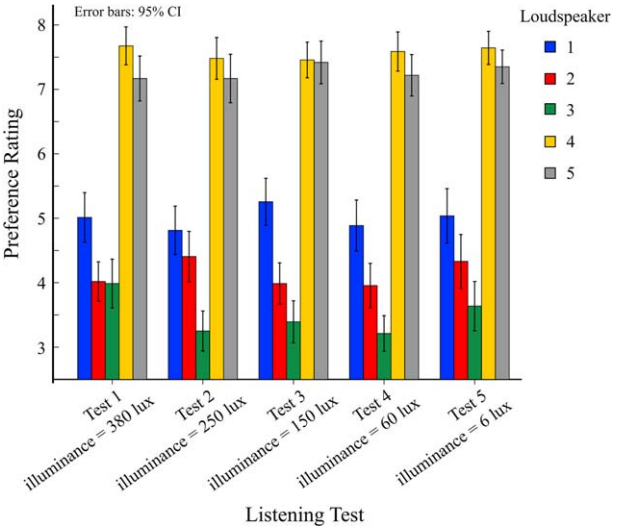


Fig. 3. Mean loudspeaker preference ratings along with 95% confidence intervals.

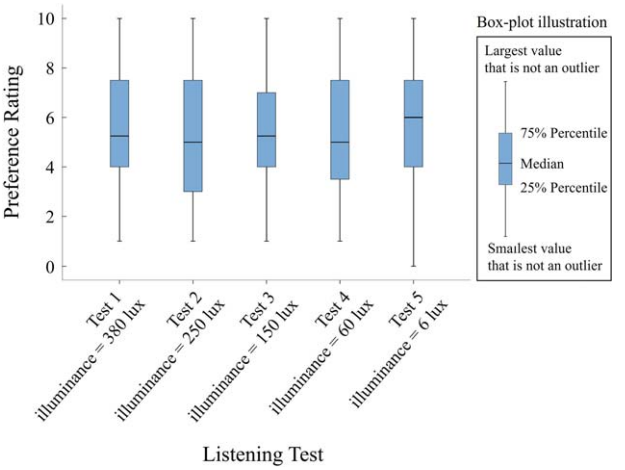


Fig. 4. The box plot of the subjective test results.

2.1 Main Effects in Each Test

As shown in Table 3, the statistically significant factors are summarized. The analysis results revealed that Lighting had no significant effect on preference rating, while Loudspeaker was identified as the main effect, meaning that the evaluated loudspeakers were significantly distinguished in each listening test. Meanwhile, no other significant effect or interaction effects due to the category Observation were observed, indicating that the experimental results were reliable and reproducible.

Surprisingly, the ANOVA also showed a significant effect and interaction effects related to Program on preference ratings. Although some previous studies [25, 32] have provided experimental verification for the usage of these program materials, the selected stimuli still significantly affected the evaluation results in some of the tests with different lighting conditions. Further analysis will be provided in SEC. 3.

2.2 Mean Opinion Scores of Each Test

As plotted in Fig. 3, the mean opinion scores of preference ratings along with 95% confidence intervals are presented. It can be observed that the loudspeaker mean ratings were essentially the same under different lighting conditions. No obvious discrepancies in the mean loudspeaker preference ratings among the tests were detected. Furthermore, the rank order of the preferred loudspeaker remained consistent across different tests. These experimental results provide evidence that there are no significant effects of lighting level on the overall preference rating of perceived sound quality.

However, it is worth noting that the 95% confidence intervals of the evaluation results for some speakers varied across different lighting environments. For instance, there was almost no overlap between the confidence intervals of the evaluation results for Loudspeaker 3 in Tests 1 and 2, as shown in Fig. 3. The box plot of the evaluation results provided a concise overview of the distribution of the received raw data in the listening tests, as illustrated in Fig. 4. The box plot represents the range occupied by 50% of the measured subjective ratings along with the indication of the median. More details of different elements are explained in Fig. 4. The box plot showed that there was less variation in the data distribution of the subjective results in Test 3 compared to the other tests. This indicates that the dispersion of loudspeaker preference ratings might be affected by different lighting environments in subjective listening tests. Further analysis will be provided in SEC. 3.

3 COMPARATIVE ANALYSIS AND DISCUSSION

In the previous section, the analysis results indicated that, although there were no significant effects of lighting environments on overall perceived sound quality, the 95% confidence intervals of rating results were still appreciably affected by different lighting environments in subjective listening tests. Moreover, unexpected significant interaction effect between Loudspeaker and Program were observed in some experiments where the lighting conditions seemed

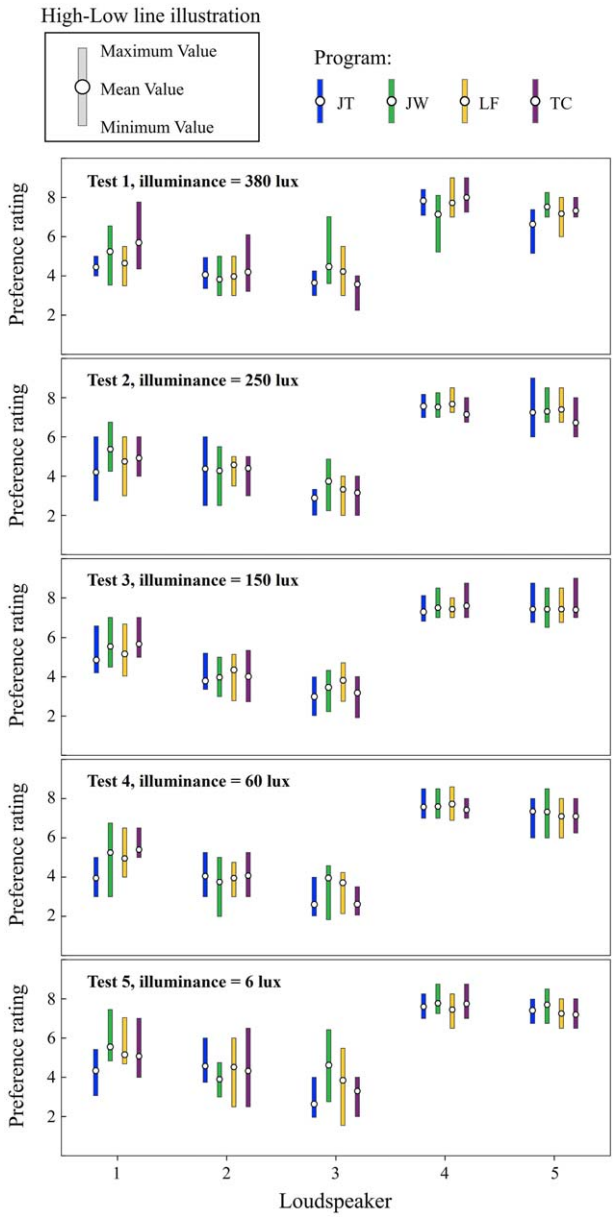


Fig. 5. High-Low line results of the evaluated loudspeakers under different program materials.

too dark or too bright. This section provides a comparative analysis and discussion of these issues.

3.1 Effects of Program Material

The ANOVA showed that in the listening tests conducted under different lighting environments, the interaction effect between Loudspeaker and Program on preference rating varied across experiments. As illustrated in Fig. 5, the mean opinion scores of preference ratings under different program materials along with maximum and minimum values in each listening test are presented. The mean values under different programs are represented by the solid dots. The *x*-axis corresponds to the different loudspeakers. The top and bottom of the High-Low lines denote the maximum and minimum values, respectively, while different colors are for the different programs. It should be noted that, to mitigate the influence of individual extreme values, the maximum and minimum values in Fig. 5 were calculated as the averages of the highest and lowest three values from each listening test.

It could be clearly found that there were fewer differences among the four programs in terms of their mean scores and the distribution of the ratings in Test 3, whereas this phenomenon was not observed in the other tests. The interaction effect between Loudspeaker and Program were more susceptible to the influence of lighting environments in the listening tests where the lighting levels were too bright or too dark. For example, in Test 1, under a lighting environment with an illuminance of 380 lux, the maximum evaluation results for Loudspeaker 3 under Program JW (7.2 points) were noticeably higher than those under other materials. Similar results for Loudspeaker 1 under Program TC (7.8 points) were obtained in Test 1, which was conducted under the brightest lighting level.

Meanwhile, in Test 5, under a lighting environment with an illuminance of 6 lux, the mean preference rating for Loudspeaker 1 under Program JT (4.1 points) were the lowest, even falling below the minimum values of some other program materials (minimum rating of Program JW = 5.0 points; minimum rating of Program LF = 4.8 points). Additionally, the lowest program preference ratings for Loudspeakers 4 (5.1 points) and 5 (5.0 points) in Test 1 were significantly lower than the average values in Test 3 (mean rating of Loudspeaker 4 = 7.5 points; mean rating of Loudspeaker 5 = 7.3 points). Similarly, the highest program preference ratings for Loudspeaker 3 in Test 5 (6.5 points) were also much higher than the average values in Test 3 (3.3 points).

Moreover, lower-quality loudspeakers appeared to be more susceptible to the interaction effect between Program and Loudspeaker under dim lighting conditions in the experiments. Fig. 5 shows that the mean program ratings were almost the same for the two loudspeakers with highest preference sound quality, Loudspeakers 4 and 5 in Tests 3, 4, and 5 (the differences in mean ratings between different program materials did not exceed 0.5 points). However, this phenomenon was not observed for the lowest-quality loudspeaker, Loudspeaker 3. Meanwhile, the evaluation results for the lowest-quality loudspeaker, Loudspeaker 3 in Test 5, were also more dispersed compared to the results in Test 3. Clearly, there was a significant interaction effect between Program and Loudspeaker on preference ratings of perceived sound quality in the experiments. Different assessment results could still be obtained under varying lighting environments, even when the evaluated loudspeakers and program materials were exactly identical in subjective listening tests.

Additionally, the variation of mean preference ratings under different programs in each listening test were also analyzed. As illustrated in Fig. 6, the *x*-axis represents the different listening tests, while the *y*-axis denotes the mean preference rating. The evaluation results of each program material in different listening tests are represented by bar charts in distinct colors in each row of subgraphs, while

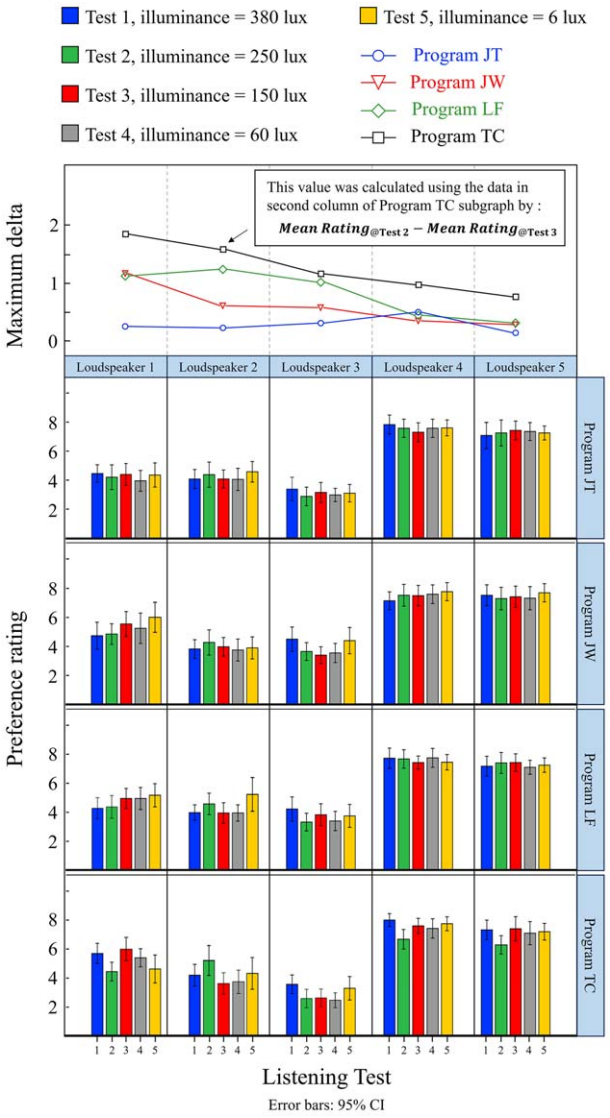


Fig. 6. Mean preference ratings along with 95% confidence intervals, under different programs in each listening test.

the evaluation results of each loudspeaker are presented in each column of subgraphs.

To facilitate clarity and direct comparison, the maximum delta of mean preference ratings under different programs is depicted by a solid line in the top chart. This value was calculated as the difference between the maximum and minimum mean evaluation scores of the given program material across different experiments. An example is also provided in the top chart of Fig. 6. Generally, the mean preference rating under Program JT showed the smallest variation across the various experiments, while the fluctuations in the evaluation results under Program TC were the most noticeable. The subjective evaluation consistency of Programs JW and LF across the experiments lay between that of Programs JT and TC.

Some patterns could be found from the acoustic characteristics of the program materials. Fig. 7 shows the long-term spectral analysis of each program. Program TC has the largest fluctuation in curve, while Programs JW and

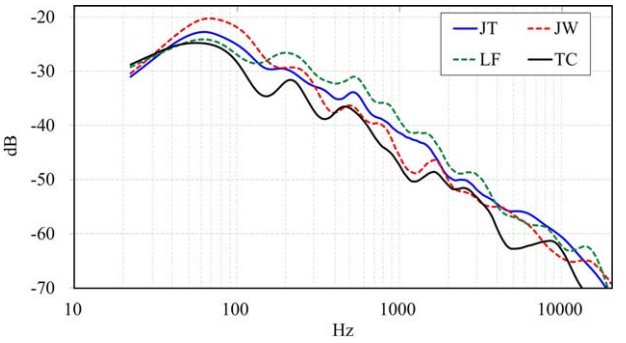


Fig. 7. The long-term spectrum (one-third octave smoothing) of each program material.

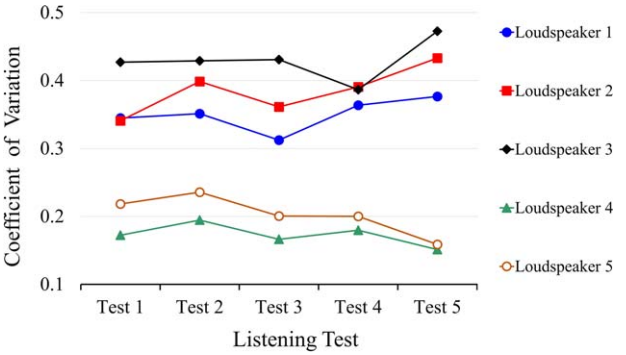


Fig. 8. CV values of listening tests.

LF place more emphasis on the reproduction of low frequencies and mid frequencies, respectively. In general, the curve fluctuation of Program JT is relatively small, and the proportions of high, medium, and low frequencies are relatively balanced. Therefore, for the selection criteria of program materials, it is recommended to first consider the smoothness of the power spectrum, followed by the spectrum balance. This is similar to the experimental results in [32].

3.2 Effects on Distribution of Measured Data

According to the analysis in SEC. 2, the dispersion of preference ratings could be affected by different lighting environments in subjective listening tests. For each listening test, the dispersion of individual loudspeaker preference rating was calculated using the coefficient of variation (CV). The CV value is defined as the ratio of the standard deviation to the mean, and it is commonly used as a normalized measurement of dispersion in a probability distribution or frequency distribution.

As illustrated in Fig. 8, the trends of Loudspeakers 4 and 5, the two highest perceived sound quality loudspeakers, were almost the same, although the CV values for different loudspeakers were discrepant. Higher CV values were obtained from the tests conducted in two bright lighting environments with an illuminance of 380 and 250 lux, while Test 5, conducted in the darkest lighting condition with an illuminance of 6 lux, had the lowest CV values, indicat-

ing that the evaluation results of tested speakers were more concentrated.

In contrast, for the other three loudspeakers with lower perceived sound quality, there were significant differences in CV values and trends among the tested loudspeakers. However, it could still be confirmed that Test 5, conducted in the dimmest environment, showed the largest dispersion of ratings for all of these three tested loudspeakers.

3.3 Relation Between Lighting Level and Perceptual Sound Quality

In the presented results, there was insufficient experimental evidence showing significant effects of lighting level on overall perceptual sound quality in subjective listening tests. However, the analyses revealed that in the listening tests conducted under different lighting environments, the influence of Program on preference ratings varied across experiments. Additionally, the dispersion of loudspeaker preference ratings could be affected by different lighting conditions. These effects may partly explain why people generally mistakenly believe that a dim listening environment can improve subjective sound quality.

Firstly, in the listening tests conducted under inappropriate lighting conditions, Program could be a significant factor, and the interaction effect between Program and Loudspeaker could be more pronounced. For certain program materials, listeners were more likely to give higher or lower overall preference ratings in some tests where the lighting level was too dark or too bright. The evaluation results were more susceptible to the influence of listening contents. Moreover, lower quality loudspeakers seemed to be more affected by the interaction effect between Program and Loudspeaker under dim lighting conditions. This subjective difference in perceived sound quality could reach up to 86% in these experiments: as shown in Fig. 5, the highest program rating of Loudspeaker 3 in Test 5 was around 6.5 points, while the mean opinion score of overall preference ratings was about 3.5 points. This indicates selecting appropriate program materials in listening tests conducted under improper lighting conditions is more challenging.

Secondly, the dispersion of loudspeaker preference ratings could be affected by different lighting conditions. Undoubtedly, minimizing the dispersion of experimental data is crucial for achieving accurate and consistent results in subjective listening tests. However, the presented results showed that for loudspeakers with lower perceived sound quality, a larger distribution of ratings occurred in the listening tests where the lighting level was too dark. Specifically in Test 5, the dimmest environment resulted in the greatest dispersion of experimental data in this study. In contrast, for loudspeakers with higher perceived sound quality, a larger distribution of ratings was obtained in the tests conducted under excessively bright lighting conditions, while darker listening environments tended to reducing the data dispersion. These phenomena suggest that more listening trials may be needed when conducting subjective listening tests on perceived sound quality in excessively improper lighting environments.

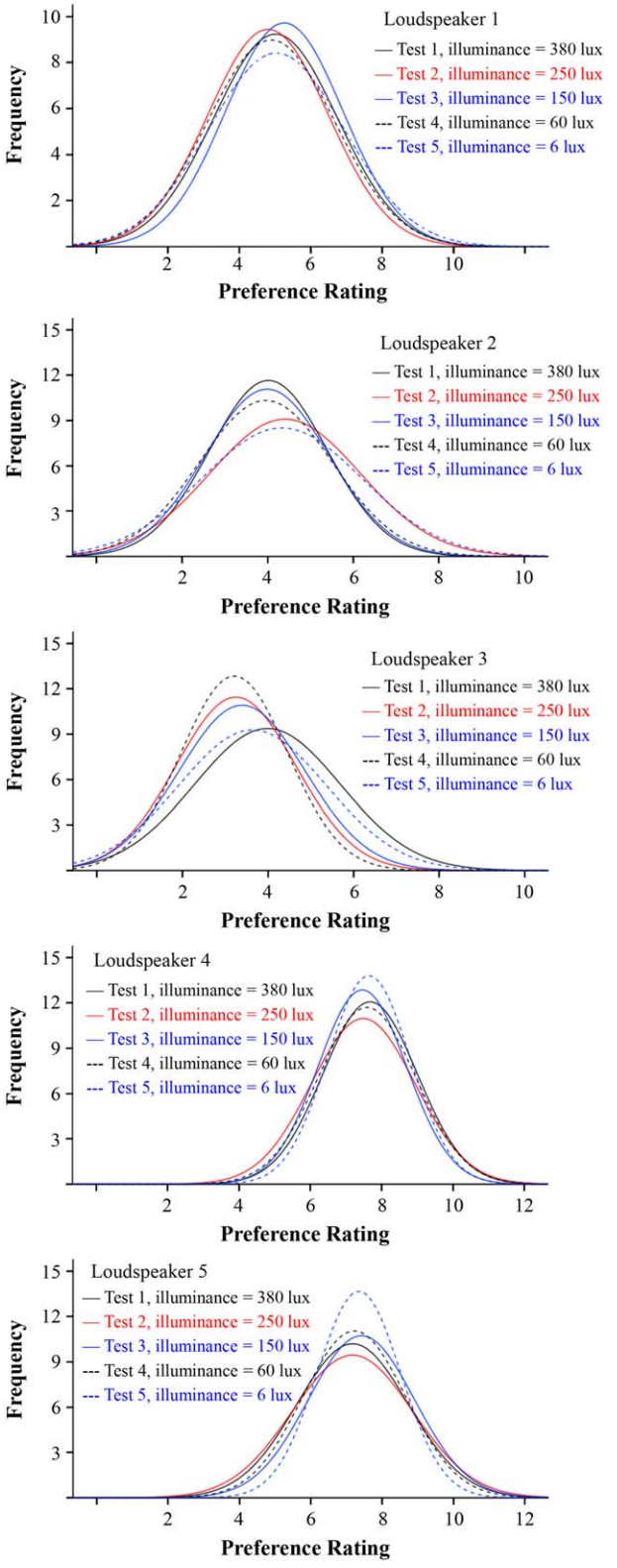


Fig. 9. Fitted distribution curves of preference ratings in each listening test.

Penultimately, in dim environments, listeners were more prone to give higher scores than the mean overall preference ratings of the evaluated loudspeakers in some listen-

ing trials, especially for loudspeakers with lower perceived sound quality in these experiments. As plotted in Fig. 9, it is evident that in Test 5 with the dimmest lighting condition, Loudspeakers 2 and 3 were more likely to be rated higher. However, higher-quality speakers were minimally affected by this effect. Meanwhile, another phenomenon also revealed that high-quality speakers are less influenced by lighting conditions compared to low-quality speakers. As presented in Fig. 6, the values of Loudspeakers 4 and 5 were relatively lower than those of Loudspeakers 1, 2, and 3, in terms of the delta between the maximum and minimum mean evaluation scores of the program material (Programs JW, LF, and TC) across different experiments.

Finally, for the phenomena discussed above, one possible explanation is that subjective perceived loudness [17, 18] and auditory thresholds [19, 39] could be affected by different visual stimuli. The impact of visual stimuli is somewhat correlated with the frequency characteristics of the audio source. Different program materials and loudspeakers exhibited varying acoustic frequency characteristics, leading to diverse subjective evaluation results under different lighting environments. In SEC. 3.1 of this paper, the selection criteria discussion on the importance of the spectrum smoothness and spectrum balance of program materials also indirectly supported this hypothesis.

In general, based on the experimental data provided in this paper, the subjective listening tests conducted under lighting conditions with illuminance levels between 60 and 150 lux showed relatively low dispersion in the measured data, as well as minimized interaction effect between Program and Loudspeaker for all evaluated loudspeakers. This suggests that such lighting conditions may be a proposed choice for conducting acoustical subjective listening tests. This range of values is lower than the recommendation in ITU-R BT.500-15. The environmental illuminance on the screen should be 200 lux for subjective assessments in home environment, as specified in ITU-R BT.500-15. And for subjective assessments in a laboratory environment, the standard does not provide specific numerical recommendations but only offers general guidelines: low room illumination. However, there are insufficient data for deeper analysis. Further investigation by varying the lighting conditions in more controlled and comprehensive methods could provide more insight in the relationship between lighting level and perceptual sound quality.

4 CONCLUSION

In this study, a series of listening tests were conducted under various lighting conditions. Each test involved comparison category rating on different loudspeakers via monophonic system and focused on solely preference sound quality, which precluded the perception and evaluation of stereo imaging and spatial information. The results of each listening test were compared. The main purpose was to examine the effects of lighting on perceptual sound quality.

The presented results revealed that there was insufficient experimental evidence showing the significant effects of lighting level on overall perceived sound quality in subjective listening tests. However, Program could be a significant factor, and the interaction effect between Program and Loudspeaker appeared more pronounced in the listening tests conducted under overly dark or bright lighting conditions, indicating greater difficulty in selecting appropriate program materials. Furthermore, the dispersion of loudspeaker preference ratings could be increased by improper lighting conditions, implying that more listening trials may be needed when conducting subjective listening tests in such lighting environments. It is advisable to consider and select lighting conditions in subjective listening tests more comprehensively to obtain reliable and authentic evaluations.

Although the effects of lighting on perceptual sound quality have been initially explored, there are still some limitations in the research of this paper and a lot of future studies are still needed. All the listeners in the conducted tests had backgrounds in acoustics and all had extensive experience in listening experiments. Thus, the number of test subjects and observations considered was relatively small. Further studies would benefit from a larger number of participants with more diverse backgrounds, as well as a greater number of observations for each test.

Additionally, the selection of the listening stimuli and the lighting conditions was based on the available literature and existing research results. It remains challenging to draw a fully scientific and comprehensive conclusion regarding the extent to which lighting conditions significantly affect perceived quality across different program materials. More variety should be considered for generalizability, especially lighting levels beyond the range described in this paper. The issues of background noise, as well as ratio between screen brightness of control device and background lighting level, also require further investigation. The authors sincerely hope that this article can provide some initial inspiration and insight for professionals in this field and encourage further investigation and deeper exploration by researchers and scientists.

5 ACKNOWLEDGMENT

The authors would like to thank the listeners who participated in this study. This work was supported by National Natural Science Foundation of China (Grant No.12374446).

6 REFERENCES

[1] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application* (Wiley, Chichester, UK, 2006).

[2] R. Walker, "Optimum Dimension Ratios for Studios, Control Rooms and Listening Rooms," BBC RD 1993/8 (1993 Jan.).

[3] F. E. Toole, "Loudspeakers and Listening Rooms, Part II: The Perception of Reflections," *J. Acoust. Soc. Am.*, vol. 119, no. 5(Suppl), p. 3251 (2006 May). https://doi.org/10.1121/1.4786061.

[4] S. E. Olive and F. E. Toole, "The Detection of Reflections in Typical Rooms," *J. Audio Eng. Soc.*, vol. 37, no. 7/8, pp. 539–553 (1989 Jul.).

[5] S. Bech, "Timbral Aspects of Reproduced Sound in Small Rooms. I," *J. Acoust. Soc. Am.*, vol. 97, no. 3, pp. 1717–1726 (1995 Mar.). https://doi.org/10.1121/1.413047.

[6] S. Bech, "Timbral Aspects of Reproduced Sound in Small Rooms. II," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3539–3549 (1996 Jun.). https://doi.org/10.1121/1.414952.

[7] Y. Ando, "Subjective Preference in Relation to Objective Parameters of Music Sound Fields With a Single Echo," *J. Acoust. Soc. Am.*, vol. 62, no. 6, pp. 1436–1441 (1977 Dec.). https://doi.org/10.1121/1.381661.

[8] G. A. Soulodre, N. Popplewell, and J. S. Bradley, "Combined Effects of Early Reflections and Background Noise on Speech Intelligibility," *J. Sound Vibr.*, vol. 135, no. 1, pp. 123–133 (1989 Nov.). https://doi.org/10.1016/0022-460X(89)90759-1.

[9] G. A. Soulodre, "Evaluation of Objective Loudness Meters," presented at the *116th Convention of the Audio Engineering Society* (2004 May), paper 6161.

[10] A. Gabrielsson, B. Hagerman, T. Bech-Kristensen, and G. Lundberg, "Perceived Sound Quality of Reproductions With Different Frequency Responses and Sound Levels," *J. Acoust. Soc. Am.*, vol. 88, no. 3, pp. 1359–1366 (1990 Sep.). https://doi.org/10.1121/1.399713.

[11] A. J. Watkins, "Perceptual Compensation for Effects of Reverberation in Speech Identification," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 249–262 (2005 Jul.). https://doi.org/10.1121/1.1923369.

[12] F. E. Toole, *Loudspeakers and Rooms for Sound Reproduction –A Scientific Review*," *J. Audio Eng. Soc.* vol. 54, no. 6, pp. 451–476 (2006 Jun.).

[13] S. Bech, "Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position," *J. Audio Eng. Soc.*, vol. 42, no. 12, pp. 999–1007 (1994 Dec.).

[14] S. E. Olive P. L. Schuck, S. L. Sally, and M. E. Bonneville, "The Effects of Loudspeaker Placement on Listener Preference Ratings," *J. Audio Eng. Soc.*, vol. 42, no. 9, pp. 651–669 (1994 Sep.).

[15] IEC, "Sound System Equipment - Part 13: Listening Tests on Loudspeakers," Tech. Rep. 60268-13 (1998 Mar.).

[16] ITU-R, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," *Recommendation ITU-R BS.1116-1* (1997 Oct.).

[17] H. Fastl, "Audio-Visual Interactions in Loudness Evaluation," in *Proceedings of the 18th International Congress on Acoustics (ICA)*, pp. 1161-1166 (Kyoto, Japan) (2004 Apr.).

[18] C. Patsouras, T. Filippou, and H. Fastl, "Influences of Color on the Loudness Judgement," presented at the *3rd EEA European Congress on Acoustics Forum Acusticum* (Sevilla, Spain) (2002 Jan.).

[19] Z. Zhao and L. Xie, "Measurement on Just Noticeable Difference of Auditory Perception With Visual Stimuli," *Chin. J. Acoust.*, vol. 5, pp. 624–631 (2013 Sep.).

[20] J. G. Beerends and F. E. De Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *J. Audio Eng. Soc.*, vol. 47, no. 5, pp. 355–362 (1999 May). https://doi.org/10.1121/1.426776.

[21] S. Agrewal, S. Bech, K. B. Bærentsen, K. De Moor, and S. Forchammer, "Method for Subjective Assessment of Immersion in Audiovisual Experiences," *J. Audio Eng. Soc.*, vol. 69, no. 9, pp. 656–671 (2021 Sep.). https://doi.org/10.17743/jaes.2021.0013.

[22] S. Agrawal, S. Bech, K. De Moor, and S. Forchhammer, "Influence of Changes in Audio Spatialization on Immersion in Audiovisual Experiences," *J. Audio Eng. Soc.*, vol. 70, no. 10, pp. 810–823 (2022 Oct.). http://dx.doi.org/10.17743/jaes.2022.0034.

[23] ITU-R, "Methodologies for the Subjective Assessment of the Quality of Television Images," *Recommendation ITU-R BT.500-15* (2023 May).

[24] S. Bech, "Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data," in *Proceedings of the 8th International AES Conference: The Sound of Audio* (1990 May), paper 5427.

[25] S. E. Olive, "A Method for Training Listeners and Selecting Program Material for Listening Tests," presented at the *97th Convention of the Audio Engineering Society* (1994 Nov.), paper 3893.

[26] W. van Bommel, *Interior Lighting: Fundamentals, Technology, and Application* (Springer, Cham, Switzerland, 2019). http://dx.doi.org/10.1007/978-3-030-17195-7.

[27] L. Loe, K. P. Mansfield, and E. Rowlands, "Appearance of Lit Environment and Its Relevance in Lighting Design: Experimental Study," *Light. Res. Technol.*, vol. 26, no. 3, pp. 119–133 (1994 Sep.). http://dx.doi.org/10.1177/096032719402600301.

[28] R. M. Kirsch, *Lighting Quality and Energy Efficiency in Office Spaces*, Ph.D. thesis, Technische Universität Berlin, Berlin, Germany (2015 May).

[29] S. E. Olive, "Do College Students Prefer the Same Headphone Sound Quality as Trained Listeners?" in *Proceedings of the International Symposium on Electro-Acoustic Technologies (ISEAT)*, pp. 121–137 (Shenzhen, China) (2013 Nov.).

[30] S. E. Olive, "Differences in Performance and Preference of Trained Versus Untrained Listeners in Loudspeaker Tests: A Case Study," *J. Audio Eng. Soc.*, vol. 51, no. 9, pp. 806–825 (2003 Sep.).

[31] S. Temme, S. Olive, S. Tatarunis, T. Welti, and E. McMullin "The Correlation Between Distortion Audibility and Listener Preference in Headphones," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), paper 9118.

[32] S. Sun, Y. Shen, Z. Liu, and X. Feng, "The Effects of Recording and Playback Methods in Virtual Listening Tests," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 570–582 (2015 Jul.). https://doi.org/10.17743/jaes.2015.0058.

[33] S. Olive, T. Welti, and E. Mcmullin, "The Influence of Listeners' Experience, Age, and Culture on Headphone Sound Quality Preferences," presented at the *137th Con-*

vention of the Audio Engineering Society (2014 Oct.), paper 9177.

[34] K. Jensen and J. Hjortkjær, "An Improved Dissonance Measure Based on Auditory Memory," *J. Audio Eng. Soc.*, vol. 60, no. 5, pp. 350–354 (2012 May).

[35] E. Vincent, M. G. Jafari and M. D. Plumbley "Preliminary Guidelines for Subjective Evaluation of Audio Source Separation Algorithms," presented at the *UK ICA Research Network Workshop* (2006 Sep.), pp. 93–96.

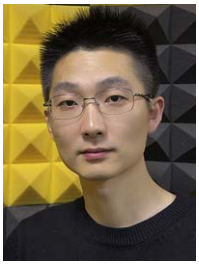
[36] ITU-R, "Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," *Recommendation ITU-R 1770.3* (2012 Aug.).

[37] R. Schatz, S. Egger, and K. Masuch, "The Impact of Test Duration on User Fatigue and Reliability of Subjective Quality Ratings," *J. Audio Eng. Soc.*, vol. 60, no. 1/2, pp. 63–73 (2012 Jan.).

[38] D. Schwarz, G. Lemaitre, M. Aramaki, and R. Kronland-Martinet, "Effects of Test Duration in Subjective Listening Tests," in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 515–519 (Utrecht, Netherlands) (2016 Sep.).

[39] J. J. O'hare, "Intersensory Effects of Visual Stimuli on the Minimum Audible Threshold," *J. Gen. Psychol.*, vol. 54, no. 2, pp. 167–170 (1956 Apr.). https://doi.org/10.1080/00221309.1956.9920271.

THE AUTHORS



Shuyuan Sun



Simiao Chen



Yong Shen

Shuyuan Sun obtained his B.Sc. and M.Sc. degrees in acoustics from Nanjing University in 2011 and 2015, respectively. He has worked at Apple Inc. and AAC Technologies Inc., primarily focusing on the development of audio-related solutions. He led a team involved in various research and development activities, including acoustic user experience research, auditory perception measurement, architecture design of audio system, acoustic transducer development, audio power amplifier design, audio algorithm development, and audio system testing and tuning. He holds over 25 authorized patents in China, the United States, and Japan, covering various fields such as consumer electronics, automotive, television, and wearable devices. He is currently pursuing a Ph.D. degree at the Institute of Acoustics at Nanjing University under the supervision of Prof. Shen.

•

Simiao Chen received her B.Sc. and Ph.D. degrees from Nanjing University in 2012 and 2025, respectively. Her research interest is mainly in the area of electroacoustics, spatial audio, and audio signal processing.

•

Prof. Yong Shen has been an AES member since 2004. He received B.Sc., M.Sc., and Ph.D. degrees in acoustics from Nanjing University where he continued his career in research and education. Yong works on R&D of electroacoustics and room acoustics. Yong has organized the International Symposium on Electro Acoustics Technology (ISEAT), a biannual symposium held in Shenzhen, since 2007.