# Perception difference for approaching and receding sound sources of a listener in motion in architectural sequential spaces

Tingting Yang and Jian Kang

Citation: The Journal of the Acoustical Society of America **151**, 685 (2022); doi: 10.1121/10.0009231
View online: https://doi.org/10.1121/10.0009231
View Table of Contents: https://asa.scitation.org/toc/jas/151/2
Published by the Acoustical Society of America

JASA ARTICLE

# Perception difference for approaching and receding sound sources of a listener in motion in architectural sequential spaces

Tingting Yang and Jian Kang[a)]
*Institute for Environmental Design and Engineering, The Bartlett, University College London, London WC1H 0NN, United Kingdom*

**ABSTRACT:**
This study investigates the dynamic auditory perception in large sequential public spaces for listeners in motion with a stationary primary sound source. Virtual soundwalks, involving four music and voice sources and validated with *in situ* soundwalks, were conducted in an exhibition space. The perception differences between the approaching and receding sound sources were explored, and three major effects were found. The rising sound received a higher rating in each room with a greater perceived change in the loudness than the falling sound despite equal changes in both levels (approach effect). The difference was greater for the room connected to the source room. The loudness in the room connected to the source room receives a sharp drop (plummet effect) for the receding sound source, which was larger for music than for voice. The effect of the background sound impairing the perceptual priority of rising sound was profound in the receiving rooms. The loudness patterns could not be extended to other perceptual attributes, including reverberation. An increasing symmetry of the overall perception between the different sound source types was observed (convergence effect) either by the approaching or receding sound sources. The overall asymmetry of the directional aspects occurring with the noise and voice was not as distinguishable as with music.
 https://doi.org/10.1121/10.0009231

(Received 7 June 2021; revised 16 October 2021; accepted 10 December 2021; published online 3 February 2022)
[Editor: Shiu-Keung Tang] Pages: 685–698

## I. INTRODUCTION

The characteristics of the dynamic sounds that people hear during normal movement, i.e., clearer and louder or unclearer and quieter on their chosen path, are important for a variety of topics. For instance, well-defined traffic flow patterns are desirable in large-scale buildings with multiple spaces, e.g., theatres, museums, stores, and streets. In contrast to natural environments, the direction of a listener in motion within a building is crucial as the sound propagates through sequential architectural spaces on the same path. Several visual and non-perceptual factors, e.g., lighting, events, public address, and voice alarm systems,[1,2] influence traffic flow patterns for acoustic, aesthetic, and functional purposes. However, investigations on the dynamic auditory perception for multiple sound sources and background noise are seldom reported.

To start with the perceived difference between two sound signals, whether identical or not, Cremer and Müller proposed two methods for evaluating the subjective loudness.[3] One method is a reiterative approach wherein the signals are controlled by the subject. Beginning with a signal $A_1$, which is louder than signal $A_2$, the subject gradually lowers the loudness of signal $A_1$ until it becomes weaker than signal $A_2$. The second method is to set a reference signal wherein the subject listens to steadily stimulated signals $A_1, A_2, \ldots, A_n$ and is required to compare each signal to the reference signal. Thereafter, the loudness of signal $A_n$ is evaluated as either larger or smaller than the reference signal. Cremer and Müller suggested that the latter method is more reliable because the results obtained by the former method could be different if the order is reversed, i.e., the quieter signal $A_2$ approaches the louder signal $A_1$. To some extent, this raises the perceptual disparity between the rising and falling sound levels as a key question in the study of acoustics.

To understand dynamic sounds in the context of the frequency spectrum, waveform amplitude, or both, a series of technical measures have been proposed to estimate the loudness of fluctuating sound. These measures, including the energy-equivalent level of a steady sound ($L_{eq}$) and the 95th percentile of the loudness distribution $N_5$,[4] assume that all temporal portions of a sound contribute equally to overall loudness.[5] However, this conjecture was demonstrated to be incorrect in previous studies in which the listeners' judgments of the global loudness of a level-fluctuating noise with a duration of 1 s are influenced more by the first 100–300 ms of the sound than by its middle portion.[6–8] The temporal weighting of the loudness presents a pattern similar to the primacy effect in the short-term memory.[9,10] The beginning of the temporal sound has higher weights, indicating that the first portion contributes more to the perceived loudness of the sound than the middle portion. Also, the end portion has higher weights to a lesser extent than the middle portion as a recency effect.[6,7] The rising and falling sounds has been demonstrated to have different perceptions.[11] This evidence suggest that the changes in the "loudness" for each portion or rising and falling tones in the architectural

[a)]Electronic mail: j.kang@ucl.ac.uk, ORCID: 0000-0001-8995-5636.

sequential spaces, could be a complex and nonlinear phenomenon.

For the building environment, Bruce and Davies[12] suggested that prior experiences of similar spaces and the perceived loudness affect the expectation of the soundscape through the soundwalks in urban environments. Through numerous listening tests, Wang *et al*.[13] explored the noise acceptance by evaluating the recordings taken in transportation spaces with the acoustic sequence, i.e., when the listener was either stationary or walking in spaces where the sounds form sequences. The acoustic units were divided into strong, medium, and weak levels. Wang *et al.* concluded that all of the acoustic sequences exhibited "attenuation effects." The high-acceptance units offer an "enhancement effect," wherein the acceptance increases after a high-level acceptance. The low-acceptance units display a "boost effect," in which the acceptance increases after a low-level acceptance. These approaches have reported outcomes that potentially explain how people deal with the familiarity of a sound source or sequence. However, visual information plays an important role in calibrating the auditory space and, therefore, the lack of spatial information during the experiment raises concerns in practice. Several applications are being developed by using the virtual and augmented reality revolution to explore the visual content along with the acoustic spatiality. Although there are continuing discussions on the validity between the real environment (RE) and virtual environment (VE), the success of virtual soundwalks,[14] dynamic auditory perception,[15] and a convincing virtual acoustic environment, which is guided with a real-world sound field, could provide a reference for further studies in this regard.

As part of the spatial information, the auditory distance perception related to the rising and falling tones, enabling the location of the objects, is important in the spatial awareness. However, the research relative to the studies of the directional aspect of the sound localization remains scarce.[16] The primary cues are the sound level, reverberation, and frequency. Also, the effect of the background noise and multiple sound sources could have remarkable relevance to the distance judgments.[17] However, the effects of noisy environments on the auditory distance perception are unknown except for the study of Mershon *et al*.,[18] which reported that the perceived auditory distance for the sound sources decreased as the background noise level increased under the condition of presenting a distance between 0.75 and 6.0 m in rooms with a high or low reverberation effect,. Moreover, the auditory distance information is useful in segregating the sound sources in complex acoustical conditions when the background noise or reverberations are present, helping focus attention and improve the identification of the sound source, including in "cocktail party" situations.[19–21]

The overarching aim of this study is to explore the dynamic auditory perception of a listener in motion, i.e., approaching and receding sound sources, with a stationary primary sound source in acoustically complex enclosures, i.e., large sequential public spaces where background noise or reverberations are present. Also, the effect of the sound source type is explored. In this study, the soundwalks were conducted in the VEs, which are validated with the soundwalks in the REs. The research contains the implication that none of the paths are sufficiently loud in dB to cause auditory discomfort, which could yield path avoidance behavior.

## II. METHODOLOGIES

### A. Site selection

Several representative sites were considered as pilots to appropriately select rooms based on the identified research questions. The condition of having a source room with only one stationary primary source and also having different paths to the source room via a succession of receiving rooms are ideal for the dynamic auditory perception experiments; however, they are challenges to meet in reality for large public spaces. As such, the site was peculiar in nature and selected for its conditions.

The selected case site was a portion of the exhibition spaces on the fourth floor of the Tate Modern, London, United Kingdom. Figure 1 illustrates the primary source and the interior conditions of the source and receiving rooms of the investigated spaces. The primary sound source "Babel, 2001" is an artistic sound installation. As a permanent exhibit by artist Cildo Merireles,[22] the concept of this work is to depict an imaginary "confusion" with incomprehensible information. It is compiled using hundreds of radios shaped as a cylindrical tower with a radius and height of approximately 6.0 and 8.0 m, respectively, which generate the unintelligible mixes of music and voices.

According to the floor plan, a visitor can pass through the exhibition either in the sequence from rooms 4 to 1 or rooms 1 to 4, i.e., starting or ending with the source room. The source room (room 1, 13.0 m × 9.0 m × 9.8 m) was a large box-shaped space with indigo blue lighting, which was sequentially connected with the following three white smaller exhibition units: first receiving room (room 2, 6.3 m × 8.0 m × 4.9 m), second receiving room (room 3, 6.3 m × 6.3 m × 4.9 m), and third receiving room (room 4, 6.3 m × 6.3 m × 4.9 m). The receiving rooms were normal exhibition spaces with no prescribed sound source. The acoustic environment of the site was subject to almost uniform wall and floor finishes and the social norm of keeping the voice level down, which was taken as the baseline condition.

### B. *In situ* measurement

To recreate the acoustic attributes of the RE in a VE, *in situ* measurements were conducted in unoccupied and occupied conditions in January 2019. The acoustic parameters were recorded using a binaural recording and analysis kit (HEAD acoustics SQobold, Herzogenrath, Germany) in each investigated room.

To clarify the effect of the primary source in the source and receiving rooms, the measurements were first conducted under the unoccupied condition, i.e., at the beginning of the


FIG. 1. (Color online) (Left) The site configuration of two experiment paths (approaching/receding sound sources) with the start/end and rating points in the VE. The corresponding photos/renderings of the source (left) and receiving room (right) are depicted for the (top) RE and (bottom) VE.

visiting hours, between 10:00 and 10:10 a.m., without any crowd transit in the investigated spaces. The value of $L_{Aeq\text{-}1min}$ in the source room (room 1) was 66.8 dB, and the values in the receiving rooms (rooms 2–4) were 60.5, 55.4, and 53.6 dB, respectively ($L_{Aeq}$, the sound level in decibels with an $A$-weighting system on the frequency domain, which generally reflects the human response, used extensively for the general purpose noise measurements). Therefore, the entire level attenuation in the spaces was 12.2 dB, and the level difference between two rooms was 6.3, 5.1, and 1.8 dB. To describe the temporal variability and low-frequency contents, $L_{Ceq\text{-}1min} - L_{Aeq\text{-}1min}$ was calculated and found to be the highest in the third receiving room (room 4), thereby indicating that its acoustic environment was dominated by low frequency ($L_{Ceq}$, the sound level in decibels having the same total sound energy as the fluctuating level measured with a $C$-weighting system on the frequency domain, which correlates better with the human response to high noise levels). The results of the unoccupied condition were also verified using a sound level meter (NTi Audio XL2, Schaaen, Switzerland).

The measurements with the background sounds in a succession of rooms were simultaneously conducted with the subjective experiments under the occupied condition during the entire visiting period with the visitors inside the spaces. The $L_{Aeq\text{-}1min}$ for each room were the mean values of the objective results of the corresponding subjective evaluations conducted simultaneously, further detailed in Sec. II E. The value of the source room (room 1) was 68.3 dB and the values of the receiving rooms (rooms 2–4) were 62.6, 58.8, and 59.7 dB, respectively. Hence, the entire level attenuation was 8.6 dB, and the level differences between the rooms with crowds were 5.7, 3.8, and −0.9 dB, respectively.

### C. Virtual reproduction

The VE was constructed using the Unreal game engine (Epic Games, Cary, NC) in accordance with the acoustic and contextual attributes obtained in the RE. For the contextual attributes, the room shape, scale, and interior finishes (e.g., the material of the floor, wall, and ceiling) corresponded for the RE and VE and were, therefore, replicated. On the other hand, the consistency of the content detail of the exhibits is pursued to an acceptable extent under the assumption that the auditory perception of the participant during the experiment is not affected by the differences in the contextual attributes of the exhibits between the RE and VE. However, "Babel, 2001" was well presented visually in the VE for its importance and specification. As shown in Fig. 1, the radios were arranged by the size at different levels using uniform materials, and the lighting on them was replicated to the authors' best efforts with the specifics.

There is considerable value in validating the acoustic spatiality, consisting of the primary source, "Babel, 2001," which attracts worldwide visitors, rather than modelling what would be possible to build in the large spaces, such as factories. More importantly, the underlying goal of this research, discussed in the Introduction, is also to explore the effects of the multiple sound sources and background noise, seeking a wider application of the dynamic auditory perception in the VE. As a result, the baseline condition for the validation between the RE and VE was selected as the occupied condition.

To reproduce the primary sound source, "Babel, 2001," in the VE, the recordings taken in the source room at the site were not used as the clips for the laboratory listening test for two reasons. First, the content of "Babel, 2001" was constantly changing, therefore, the representation of the recorded sample was not considered sufficient. Another reason is that considering the effect imposed by the reverberation, the recordings could be too generalized and, additionally, the background noise could vary for each position on different days. The reproduction method should be more generic for the validation. Figure 2(a) shows the spectrograms of the site recordings, obtained under the unoccupied condition in the source room (room 1). The character

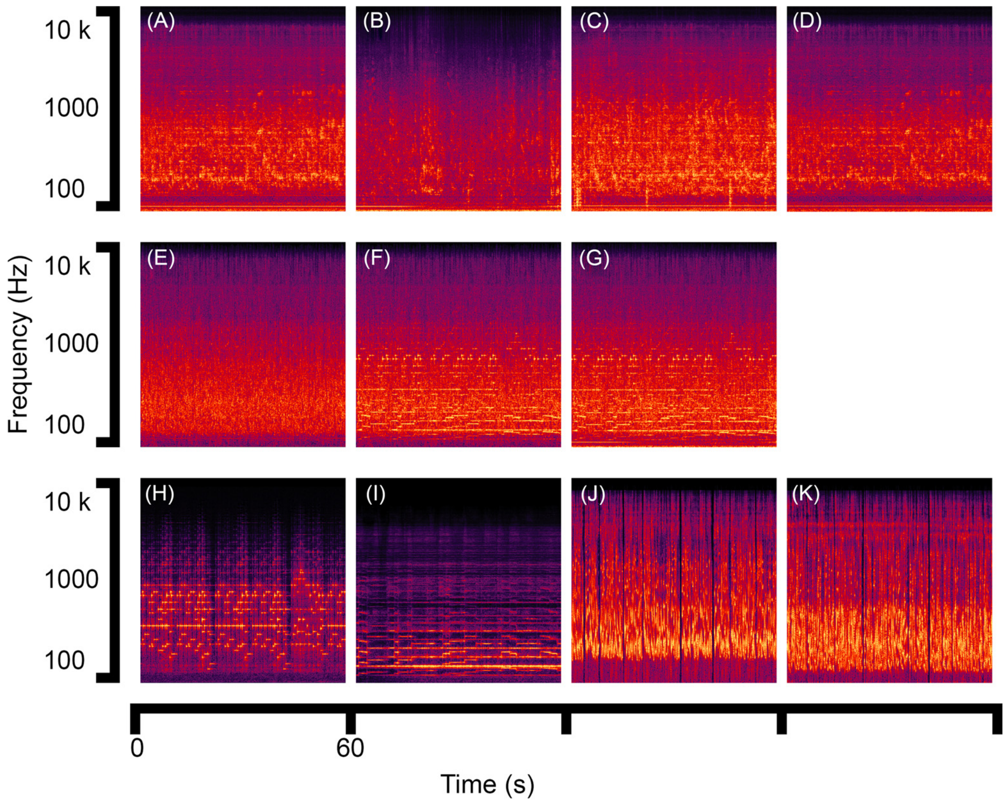


FIG. 2. (Color online) The spectrograms, including the site recordings (first row), samples for the validation between the RE and VE (second row), and samples for the virtual experiment (third row), are shown for the (a) unoccupied condition in the source room (RE); (b) background sound of the occupied condition (RE); (c) occupied condition in the source room (RE); (d) mixes of the unoccupied condition in the source room and background sound of the occupied condition; (e) mix of ten pairs of voices (VE); (f) mixes of ten pairs of voices, piano, and cello (VE); (g) mixes of ten pairs of voice, piano, cello, and background sound (VE); (h) piano (VE); (i) cello (VE); (j) female (VE); and (k) male (VE).

of the primary source, as described on the Tate Modern webpage,[23] is an analog audio streaming "a cacophony of low, continuous sound," barely audible to the human ear. It originates from the early 16th century Babel (see the Tower of Babel[24]) where, according to the biblical story, God made all of the builders speak different languages. The installation is reported to be a collection of 800 radios,[25] and every moment of "Babel, 2001" is unique. Therefore, an environment with multiple voice and music sounds in the source room (room 1), i.e., the cocktail party effect, is relevant to the context.

To begin the process of recreating an analog source in digital form, featuring a mixed content of human voices and music, the criteria of using the minimum amount of music and voice sources that could efficiently reproduce the stationary primary source in the VE were established. A sufficient number of voice sounds were required to simulate the situation of two or more people speaking at the same time. The selection of the samples is not specified except for the subject matter based on the broadcasts with female or male voices. Figure 2(e) shows the spectrograms of the sound signals for ten pairs of female or male voices for 60 s. As the paired number of speakers increased to 10, i.e., 20 people simultaneously having conversations in the source room (room 1), there was no distinguishable pattern in the time-frequency plane. Furthermore, this number of sound sources greatly exceeded the number used to simulate the cocktail party effect, which is usually between three and eight sources. The specifications for the music sample were developed on the high-/low-frequency basis. A piano [Fig. 2(h)] and a cello sample [Fig. 2(i)] were chosen as they are not emotional or famous enough to be distinguished. Figure 2(f) shows the spectrograms of the mixes of the sound signals for ten pairs of female or male voices, piano, and cello. In a naturally occurring auditory scene, music provides examples in which the levels of the different instruments are not carefully calibrated, thus, each instrument can often overwhelm the others. Therefore, the two selected instrument sounds varied in their frequencies and could create a new type of fused sound to be perceived by a listener.

In terms of the background signal to simulate the crowd transit, a 3-min binaural recording was used. This recording was captured in a normal exhibition space under the equivalent geometric and interior conditions of room 4 with no audio resource, located in another exhibition space on the same floor as the selected site, and looped across the source and receiving rooms. Figure 2(b) shows the spectrograms of the site recordings. The mixes of the unoccupied condition in the source room (room 1) and the occupied condition in the other room [Fig. 2(d)] were similar to those obtained in the occupied condition in the source room [Fig. 2(c)], demonstrating that the building conditions of the recorded room were mostly identical to those of the site. The content contained the various sounds of footsteps and conversations in the Tate Modern, such as the clicking of high heels and children's voices.

Figure 2(g) shows the spectrograms of the mix of samples representing "Babel, 2001" in the VE, which is generally similar to the site recording [Fig. 2(c)]. Two distinct attenuation distance curves were applied to appropriately define the acoustic properties of the frequency content. For the acoustic attributes, four audio volumes and reverberations were created in accordance with the individual boundary conditions of each room. These four audio volumes were separately set to approximately 6.0-, 4.0-, and 0.0-dB reductions in the sequence from rooms 1 to 4, ensuring the level difference between adjacent rooms in the sequence measured in the RE to be correspondingly reflected in the VE. In addition, only the volume of the third receiving room (room 4) was filtered by 800 Hz to satisfy the low-frequency propagation, concerning the effect of the source distance. The simulated sound level coming out of the headphones of the primary source, "Babel, 2001," in the source room (room 1) was approximately 67.0 dB(A), calibrated by a sound level meter (Nti Audio XL2, Schaaen, Switzerland), which was corresponded to the measured sound level in the RE. Meanwhile, the level of background noise was separately considered to be approximately 55.0 dB(A) and set to be equivalent across the rooms.

### D. Perceptual attribute

The subjective data obtained in the RE and VE under comparison were collected via the same questionnaire based on the consistent assessments for each individual room. This involved 4 sheets of paper with the room number arranged in sequence for the participant to complete 1 experiment involving 12 perceptual attributes in the areas of room acoustics and soundscape, along with the soundwalks in the occupied condition. The eight perceptual attributes of the room acoustics were loudness, "clarity," "reverberation," "spaciousness," "listener envelopment," "intimacy," "warmth," and "stage support," which are frequently used in the subjective evaluation of the performance spaces for the music and voice content.[26–28] The four perceptual attributes of the soundscape were "annoyance," "directionality," "acoustic comfort," and "overall impression," which further describe the affective quality and acoustic spatiality of the built environments.[29–32] The five-point Likert scale was used as suggested in ISO/DIS 12913-2,[33] and the verbal labeling was provided under each scale.

### E. *In situ* evaluation

The *in situ* subjective surveys were conducted in January 2019. The administration at the Tate Modern supported the field experiment under the condition that there was no disturbance in the art gallery. Therefore, all of the participants were randomly approached and asked for consent in the concourse rather than inside the art gallery during their departure. They were visitors to the exhibition who had completed their visitations to the investigated spaces. Consequently, it was observed that after understanding the experimental content in the listening area, the participant did not focus on the exhibits during the experiment, which they were assumed to have done previously.

The procedures of the soundwalk were carefully explained to the participants with detailed instructions provided. They were required to voluntarily walk either toward or away from the primary source on the same path. The movement was categorized into two groups—the approaching-sound-source group and the receding-sound-source group, which were used to rate the marked positions at the center of each room. Moreover, the participants were accompanied by the researcher until the experiment ended, and the researcher recorded their instant exposures to the acoustic environments simultaneously for approximately 1 min at the designated spots in the rooms using the binaural recording and analysis kit (HEAD acoustics SQobold, Herzogenrath, Germany), for which the objective results were discussed in Sec. II B.

### F. Listening test

The listening tests for the VE were conducted in a design laboratory (4.0 m × 10.0 m × 4.0 m) in January 2021. The background noises in the quiet laboratory were measured as below 35.0 dB(A) with no distinguished background noise, inferring the participant especially using active-noise cancelling headphones (Bose QuietComfort 35, Framingham, MA). The calibration of the sound volume was undertaken before being presented to the subjects in the test. The VE, as shown in Fig. 1, was shared on a monitor, which was streaming with the laptop. Such an environment can be recreated with a game engine implemented in immersive virtual reality. The reason for presenting the recreated environment to the test subjects via a monitor was to avoid any unknown effects of the virtual reality tools because a more general criterion was targeted for the auditory perception to achieve the research goal. All of the subjects were invited, and a randomized double-blind experiment was performed to categorize them into two groups (i.e., approaching-sound-source group or receding-sound-source group). Each group contained 20 participants.

The subject was informed that the listening tests comprised several rounds with varying visual and audio stimuli in a random playback sequence. They were sitting in front of the monitor while taking the listening test. Before the experiment, no self-reported hearing problems were declared by the participants. A brief guidance on the procedures, including the task assignments and device control with the keyboard and mouse, was provided prior to conducting the virtual experiment. Once started, the subject was no longer accompanied by the researcher physically, and they were under the partial control of the VE. They were restricted to a prescribed path, which was basically designed as a line moving across the spaces at a default walking speed. It should be noted that the subjects were allowed to observe the virtual space by rotating the mouse when they were in motion. Meanwhile, they could use the keyboard to call out/off the questionnaire. They provided the ratings using a mouse when they stopped at the center of each space, as illustrated in Fig. 1. A completion interface, directing the subjects into the next test, appeared as they completed the questionnaire at the end of the sequence either in the source room (room 1) for the approaching-sound-source group or the third receiving room (room 4) for the receding-sound-source group. There was a mandatory halt between the two tests at the start location for 10 s in silence to avoid any distraction imposed by prior experiences before being allowed to voluntarily start the new test. The total time duration required for each subject to complete the tests was typically less than 1 h.

## G. Subjects

Seventy-two visitors, between 18 and 60 years old, were involved in the *in situ* surveys, and a total of 216 valid questionnaires were collected. Forty subjects aged between 18 and 22 years old, were voluntarily recruited for the listening test. As discussed in Sec. II E, prior to the soundwalk, all of the participants in the RE had exposure to the acoustic environment rendered by the primary source. Comparatively, all of the participants in the VE did not have any prior knowledge and field experience of the site, which was verified after the completion of the experiment. In addition, they were restricted to the preset route, which made it almost impossible to pay further attention to the details of the exhibits. The differences between the subjects were to ensure the objectivity of both of the results obtained in the RE and VE, which could be potentially affected by the exhibits.

To avoid the possibility of age-dependent limitations, this study only used the results from those participants between 18 and 22 years of age rather than the entire sample of the broader group. The number of subjects in the RE and VE satisfied the sample criterion of a normal distribution (i.e., greater than 30). However, although the ages of the subjects in the RE and VE were within the same range, according to the one sample $t$-test ($p$) between the subjects aged between 18 and 22 years old and those between 18 and 60 years old in the RE, the former group containing 28% of the participants from the latter group, the age-dependent effects in the broader group were significantly limited. The results revealed that there were significant differences for the spaciousness in room 3 ($p = 0.040$) and listener envelopment in room 4 ($p = 0.024$). In addition, it was observed that almost every perceptual attribute, especially the loudness, reached the value of $p = 1.000$. Therefore, the results of this study could be reasonably extended to a wider age range.

## H. Data analysis

To assess the difference in the perceptual attributes, independent $t$-tests were used. In addition, to measure the similarities, the distance correlation analysis was conducted using the distance similarity measures by the Pearson correlations, solved using SPSS Statistics 26 (IBM United Kingdom Limited, Portsmouth, UK) and OriginPro 2021 (OriginLab Corporation, Northampton, MA). This method can be used to perform statistical tests such as computing similarities between pairs of automobiles based on certain characteristics, gaining a sense of which automobiles are similar to each other and which are different from each other.[29] In this study, this method was used to measure the similarities between the pairs of evaluations based on the perceptual attributes.

## I. Validation between RE and VE

To ensure an acceptable correspondence between the RE and VE, this study used three conditions in the VE, varying with the volume of the primary source or the reverberation effects of the rooms, as shown in Table I. The first VE condition was preset as the baseline condition, according to the virtual reproduction as discussed above. According to the independent $t$-tests ($p$) of the loudness between the RE and VE (first), for the approaching-sound-source group, the results of room 4 ($p = 0.007$) exhibited statistically significant differences; for the receding-sound-source group, there were statistically significant differences in room 1 ($p = 0.000$) and room 2 ($p = 0.008$). In addition, the mean rating of the loudness under the VE (first) was demonstrated to be much larger than that of the RE, especially for the receding-sound-source group, as shown in Fig. 3.

Consequently, the second condition reduced the volume of the primary source by 10.0 dB to pursue a better validity. The reduction of 10.0 dB is defined by the research of Sudarsono,[34] which shows that when participants are given the opportunity to adjust the sound level of a soundscape reproduction in the laboratory, they tend to adjust the sound level to −9.5 dB below the actual level. According to the independent $t$-tests ($p$) of the loudness between the RE and VE (second), no significant difference was found in all four of the investigated rooms. It was also observed in Fig. 3 that the mean ratings of the VE (second) were closer to those of the RE when compared to the result obtained by the VE (first). In addition, following the VE (second), the VE (third) increased the decay time of each room by 0.5 s to explore the potential reverberation effect. The results for the evaluation of the loudness were generally the same as those obtained by the VE (second), and there was only a

TABLE I. The experimental details of each experiment in the VE.

| Primary source | Background sound | VE condition | Experimental phase |
|---|---|---|---|
| "Babel, 2001" (validation) | Yes | First | VE-1 |
| | Yes | Second | |
| | Yes | Third | |
| | No | | |
| Piano | | | VE-2 |
| Cello | | Second | |
| Female | Yes/no | | |
| Male | | | |

https://doi.org/10.1121/10.0009231

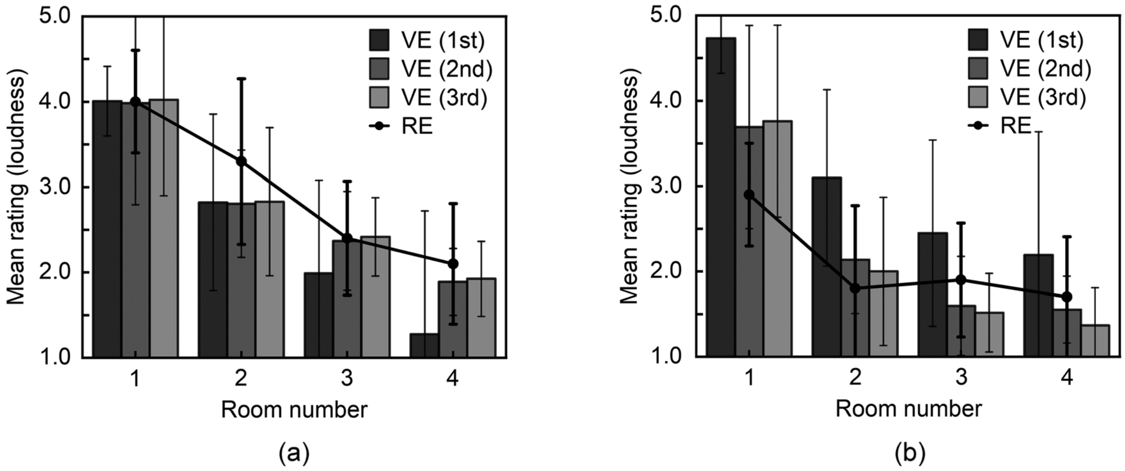


FIG. 3. The mean ratings and standard deviations of the loudness in the RE and three VEs are shown for the (a) approaching-sound-source group and (b) receding-sound-source group.

significant difference in the source room (room 1; $p = 0.050$) between the RE and VE (third).

The VE subjects were unaware of "Babel, 2001" as a piece of well-established art, whereas the RE subjects were those who supposedly had prior experiences of the exhibit. According to the subsequent feedback sheet provided after the completion of the experiment in the questionnaire, the VE subjects documented that they experienced confusion with the "chaotic" acoustic environment.

Notably, the mean rating of the loudness was equal to 4.0 in the source room (room 1) for the approaching-sound-source group, whereas the ratings for the receding-sound-source group were exceptionally high (4.7) in the VE (first) compared to those in the RE (2.9). This indicates that a distinction of the methodology between the RE and VE for the receding-sound-source group cannot be ignored. The VE subjects were automatically placed in the source room (i.e., the loudest room), whereas the RE subjects walked through the rooms before physically arriving at the source room because they were recruited in the concourse. Some of the VE subjects expressed their unpleasantness and fright upon suddenly listening to "Babel, 2001" for the first time. Such complaints were seldom received and documented by participants in the approaching-sound-source group. The values of the receiving rooms were generally smaller in the VE than those in the RE. The listener's sudden exposure to the primary source in the VE dominated the loudness perception.

Three distance correlation analyses were conducted with the approaching-sound-source group to measure the similarities between the RE and VE, as listed in Table II. Although the VE (first) was preset as the baseline in this study, the VE (second) was the "best performer" among the three for exhibiting greater similarities with the RE. Moreover, the loudness, reverberation, and stage support were well-developed throughout the three conditions; the clarity and warmth were considerably improved by decreasing the volume of the primary source in the VE (second). Enlarging the reverberation effects in each room did not lead to differences for the loudness, intimacy, and warmth and imposed detrimental effects on the clarity and directionality. There were statistically significant differences for certain perceptual attributes in some of the rooms. However, according to the independent $t$-tests ($p$) of each acoustic attributes in each room, no significant differences ($p < 0.05$) were observed for the evaluation of the loudness (room 1, $p = 0.964$; room 2, $p = 0.147$; room 3, $p = 0.939$; room 4, $p = 0.561$) and the reverberation (room 1, $p = 0.570$; room 2, $p = 0.723$; room 3, $p = 0.900$; room 4, $p = 0.078$) in each investigated room between the RE and VE (second).

### J. Virtual experiment

The experimental details of each test in the VE, including the validation and experiment, are listed in Table I. For each of the two sequenced experimental phases, VE-1 and VE-2, the participant completed five pairs of experiments, varying with the primary source (i.e., "Babel, 2001," piano, cello, female, and male) with or without background sound in either the first or second VE condition. Additionally, one validation for "Babel, 2001" with background sound in the third VE condition was conducted in VE-2. The validation

TABLE II. The proximities of the ratings by the approaching-sound-source group by the distance correlation similarity tests ($r$) between the RE and three VEs when using "Babel, 2001."

| Perceptual attribute | VE (first) | VE (second) | VE (third) |
|---|---|---|---|
| Loudness | 0.988 | 0.965 | 0.965 |
| Clarity | 0.196 | 0.749 | 0.000 |
| Reverberation | 0.773 | 0.840 | 0.882 |
| Spaciousness | 0.255 | 0.323 | 0.468 |
| Listener envelopment | −0.217 | −0.827 | −0.395 |
| Intimacy | −0.804 | −0.838 | −0.838 |
| Warmth | −0.738 | −0.233 | −0.233 |
| Stage support | 0.963 | 0.966 | 0.986 |
| Acoustic comfort | 0.607 | 0.475 | 0.498 |
| Directionality | 0.169 | −0.192 | −0.911 |
| Annoyance | 0.541 | 0.569 | −0.635 |
| Overall impression | −0.549 | 0.394 | −0.578 |

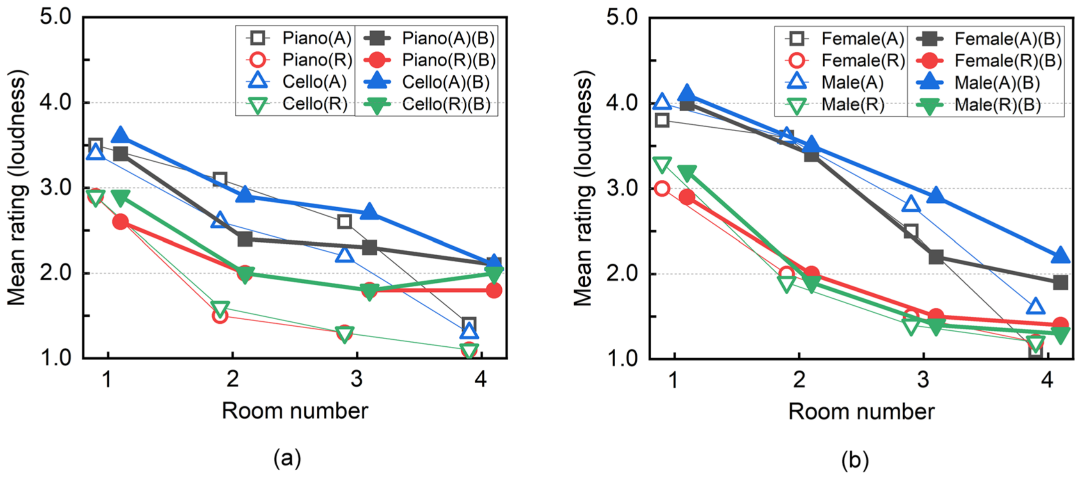


FIG. 4. (Color online) The mean ratings for the loudness of the approaching-sound-source group (*A*) and the receding-sound-source group (*R*), tested without or with background sound (*B*) are shown for (a) music and (b) voice.

and experiments were not separately conducted and were arranged in a random order, which was unique for each participant. For the VE (first and third) conditions, only the results of the validation (i.e., using "Babel, 2001" as the primary source) are presented in this study. Note that all of the results shown in Sec. III were obtained with the VE (second) condition in this study. The clips of the primary source used in the experiments had been already applied in reproducing "Babel, 2001." Figures 2(h)–2(k) show the spectrogram of each primary source.

## III. RESULTS

### A. Effects of the approaching and receding sound sources

Figure 4 shows the mean rating for the evaluation of the loudness of the approaching-sound-source and receding-sound-source groups by different source types tested without or with a background signal. The results revealed that people had very different auditory perceptions in the same actual space. Table III has further calculated the mean rating differences in the loudness between the approaching-sound-source and receding-sound-source groups without background sound. As shown in Table III by the columns of $A-R$, the mean ratings of the approaching-sound-source group were larger than those of the receding-sound-source group for all of the investigated spaces. This means that the sound with a gradual increase in the level across the spaces (i.e., approaching sound source) receives a higher importance in the perceived loudness (defined in this paper as the approach effect) no matter where the listener is located. The increased perceptual disparity was observed to be equivalent in the source room (room 1) for the music and voice with a similar mean rating difference of 0.6–0.8. However, this mean rating difference was found to be larger in the receiving rooms (rooms 2–4), according to the different source types, and greater at the high-level receiving room (the one near the source room); that is, in the first receiving room (room 2), the mean rating difference was equivalent for the piano, female, and male voices source (1.6) except for the cello (1.0). In the second receiving room (room 3), the mean rating difference was smaller for the cello than for the piano and smaller for the female than for the male, as shown in Table III. Also, as shown in Table III, according to the independent *t*-tests ($p$) of the mean rating tested without background sound between the approaching-sound-source and receding-sound-source groups, there were significant differences ($p < 0.01$) in the first receiving room (room 2) and the second receiving room (room 3) for all of the investigated sources. However, no significant differences ($p < 0.01$) were observed in the source room (room 1) and third receiving room (room 4). This also indicates that the greatest perceptual priority by the approaching sound source occurred in the receiving rooms near the source room, not in the source room.

TABLE III. The mean rating differences in the loudness between the approaching-sound-source group (*A*) and the receding-sound-source group (*R*), tested without background sound with the independent *t*-test (*p*).

| | Piano | Cello | Piano – cello | | Female | Male | Female – male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-R$ | $A-R$ | $A$ | $R$ | $A-R$ | $A-R$ | $A$ | $R$ |
| 1 | 0.6 | 0.5 | 0.0 | 0.0 | 0.8 | 0.7 | −0.2 | −0.3 |
| 2 | 1.6[b] | 1.0[a] | 0.5[a] | 0.0 | 1.6[c] | 1.6[b] | 0.0 | 0.0 |
| 3 | 1.3[b] | 0.9[b] | 0.4[a] | 0.0 | 1.0[b] | 1.4[c] | −0.3 | 0.0 |
| 4 | 0.3 | 0.2 | 0.0 | 0.0 | −0.1 | 0.4[a] | −0.5[b] | 0.0 |

[a]$p < 0.05$.
[b]$p < 0.01$.
[c]$p < 0.001$.

To explore how background noise affects the judgments in this context, Table IV further calculated the mean rating differences in the loudness between the approaching-sound-source and receding-sound-source groups with background sound. As shown in Table IV by the columns of $A-R$, the mean ratings were also larger for the approaching-sound-source group than for the receding-sound-source group, which means that the discussed approach effect was

https://doi.org/10.1121/10.0009231

TABLE IV. The mean rating differences in the loudness between the approaching-sound-source group (*A*) and the receding-sound-source group (*R*), tested with background sound (*B*) with the independent *t*-test (*p*).

| | Piano | Cello | Piano − cello | | Female | Male | Female − male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-R$ | $A-R$ | $A$ | $R$ | $A-R$ | $A-R$ | $A$ | $R$ |
| 1 | 0.8 | 0.7 | −0.2 | 0.0 | 1.1[a] | 0.9 | 0.0 | −0.3 |
| 2 | 0.3 | 0.9[a] | −0.5[a] | 0.0 | 1.4[b] | 1.6[c] | 0.0 | 0.0 |
| 3 | 0.5 | 1.0[a] | −0.4 | 0.0 | 0.7[a] | 1.5[c] | −0.7 | 0.0 |
| 4 | 0.4 | 0.1 | 0.0 | −0.2 | 0.6 | 0.9[b] | −0.3 | 0.0 |

[a]$p < 0.05$.
[b]$p < 0.01$.
[c]$p < 0.001$.

TABLE V. The mean rating differences in the reverberation between the approaching-sound-source group (*A*) and the receding-sound-source group (*R*), tested without background sound with the independent *t*-test (*p*).

| | Piano | Cello | Piano − cello | | Female | Male | Female − male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-R$ | $A-R$ | $A$ | $R$ | $A-R$ | $A-R$ | $A$ | $R$ |
| 1 | 0.7 | 0.7 | 0.0 | 0.0 | −0.2 | 0.2 | 0.0 | 0.0 |
| 2 | 1.3[b] | 0.0 | 0.4 | −1.0[a] | 0.4 | 0.8 | −0.4 | 0.0 |
| 3 | 0.3 | 0.5 | 0.3 | 0.5 | −0.2 | 0.4 | −0.3 | 0.3 |
| 4 | 0.5[a] | 0.1 | 0.5[b] | 0.0 | 0.1 | 0.1 | 0.0 | −0.2 |

[a]$p < 0.05$.
[b]$p < 0.01$.

maintained with the masking. Comparing Tables III and IV, the mean rating difference between the rising and falling levels slightly increased by approximately 0.2 in the source room (room 1) with background sound. Furthermore, in the first receiving room (room 2), the mean rating differences were stationary except for the piano, which decreased by 1.3. In the second receiving room (room 3), the mean rating differences were kept at the same level for the cello and male, whereas those of the piano and female decreased by 0.6 and 0.3, respectively. Therefore, the masking effect on the impairing perceptual priority of a rising sound did not occur in the source room but did occur in its connected receiving room. Also, according to the results of the independent *t*-tests (*p*) as shown in Table IV, no significant differences ($p < 0.01$) were observed in the first receiving room (room 2) and second receiving room (room 3) in the columns of the piano $A-R$ anymore, demonstrating that the masking effect was greater for the piano than for the cello.

In terms of the other perceptual attributes (e.g., reverberation), the mean ratings were also larger for the approaching-sound-source group than for the receding-sound-source group in most cases, except for a small value of −0.2 for the female in the source room (room 1) and the second receiving room (room 3), as shown in Fig. 5 and Table V. The perceptual difference in the source room (room 1) was identical to that of the loudness (0.7) for music, however, the perceptual difference for the voice was much lower (0.2). This also indicates a disparity in the perceiving reverberation and loudness for the voice, that is, in the case of the different loudness, the reverberation could be identical when a listener enters or leaves the source room. However, according to the results of the independent *t*-test (*p*), as shown in Table V and VI by all of the columns of $A-R$, in general, the asymmetry of the directional aspects in the reverberation was hard to distinguish because no statistically significant differences ($p < 0.01$) were observed.

To measure the approaching effect for a global view of all of the perceptual attributes, Table VII shows the distance correlation similarity tests (*r*) between the approaching-sound-source and the receding-sound-source groups with all of the investigated source types. The results showed that the greatest similarity was delivered by "Babel, 2001" in the source room (room 1, $r = 0.976$), whereas the lowest similarity was obtained by the piano in the first receiving room (room 2, $r = -0.245$). The similarities of the piano were lower than those of the other three sources, and those of the

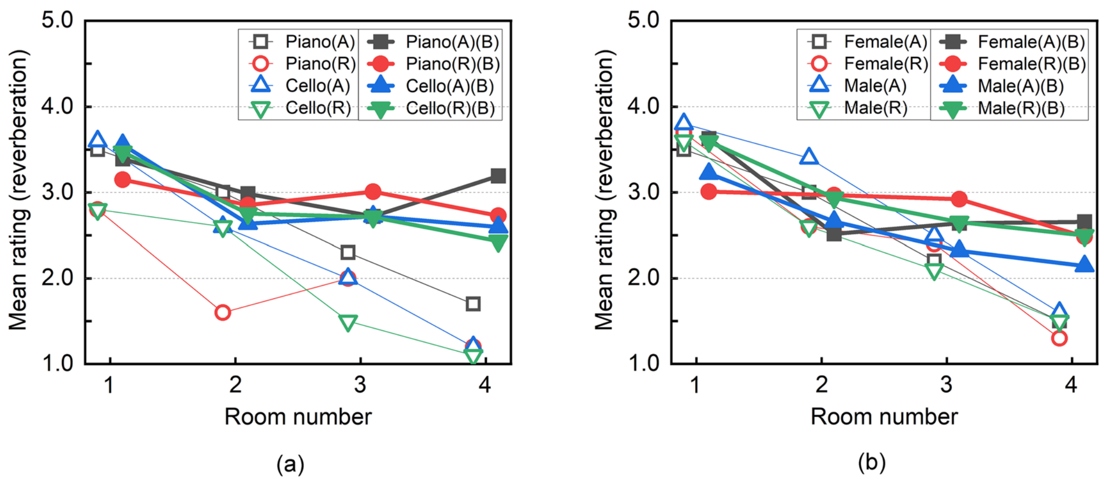


FIG. 5. (Color online) The mean ratings for the reverberation of the approaching-sound-source group (*A*) and the receding-sound-source (*R*), tested without and with background sound (*B*), are shown for (a) music and (b) voice.

TABLE VI. The mean rating differences in the reverberation between the approaching-sound-source group ($A$) and the receding-sound-source group ($R$), tested with background sound with the independent $t$-test ($p$).

| | Piano | Cello | Piano − cello | | Female | Male | Female − male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-R$ | $A-R$ | $A$ | $R$ | $A-R$ | $A-R$ | $A$ | $R$ |
| 1 | −0.2 | 0.1 | −0.3 | −0.2 | −0.6 | 0.4 | −0.6 | 0.4 |
| 2 | −0.1 | 0.1 | 0.1 | 0.3 | 0.5 | 0.3 | −0.2 | 0.1 |
| 3 | 0.3 | 0.0 | 0.1 | −0.1 | 0.3 | 0.3 | −0.2 | 0.1[a] |
| 4 | −0.5 | −0.2 | −0.1 | −0.9 | −0.2 | 0.4 | −0.1 | −0.1 |

[a] $p < 0.05$.

cello seemed to be similar to "Babel, 2001." This suggests that the perceived changes between the rising and falling levels for the noise could be less distinguishable, especially when compared to high-frequency music. The asymmetry of the directional aspects most occurred with music but not with broadband noise and voice, especially at the high levels. Furthermore, the greatest similarity with the background sound was also exhibited by "Babel, 2001" in the source room (room 1, $r = 0.992$), and the lowest similarity was also obtained by the piano in the second receiving room (room 3, $r = 0.508$).

## B. Effect of the sound source type

As shown in Fig. 4, for the approaching-sound-source group, the overall range of ratings of the loudness was equivalent for the piano and cello and slightly smaller for the female than for the male. This suggests that the rising intensity piano and cello have comparable perception, whereas the rising intensity male could have perceptual priority to the female. Additionally, as shown in Table III by the columns of piano − cello and female − male, the mean rating was identical between the piano and cello in the source room (room 1) while different by 0.5 in the first receiving room (room 2) and 0.4 in the second receiving room (room 3). On the other hand, the mean rating between the female and male was slightly different in the source room (room 1; −0.2), second receiving room (room 3; −0.3), and third receiving room (room 4; −0.5). These results suggest that the largest difference imposed by the sound source type for a rising level sound did not occur in the source room but did occur in the receiving rooms.

Comparatively, as shown in Fig. 4, for the receding-sound-source group, the overall range of the ratings of the loudness was equivalent for the piano and cello, and opposite to the results of the approaching-sound-source group, was slightly larger for the female than for the male. Therefore, it is found that when a listener approaches or recedes from the sound source, the perceptual priority of the female and male could be different, that is, for the female when receding from the sound source. On the other hand, as shown in Table III by the columns of piano − cello and female − male, the mean ratings were equivalent for the piano and cello and the female and male in most of the investigated rooms except for a difference of 0.3 between the female and male in the source room (room 1). This indicates that in the case of the receding sound source, the loudness difference across the various sound source types is needed to be more concerned in the source room (i.e., when the sound is loud or close to the listener). It is observed that the loudness in the room connected to the source room (room 2) received a sharp drop (defined in this paper as the plummet effect), that is, it was rated as 2.9 in room 1, which is particularly high compared to the ratings in rooms 2–4, which range from 1.1 to 1.5 as shown in Fig. 4. However, this plummet effect was weaker for the voice, that is, the value of room 2 was larger than that of the music for approximately 0.5, which made the difference between rooms 1 and 2 smaller. Note that the plummet effect was not observed in the approaching sound source and was on a sound source type basis, for which the music was larger than the voice.

Table VIII calculated the mean rating differences in the loudness between the conditions without and with background sound for either the approaching-sound-source group or receding-sound-source group. As shown in the columns of $A - A(B)$ and $R - R(B)$, the ratings of the rising piano decreased and those of the cello increased as a result of the masking. The magnitudes of the music were greater in the receiving rooms but not in the source room. Also, although no significant differences ($p < 0.05$) were observed in rooms 1–3, according to the results of the independent $t$-tests ($p$) as shown in Table VIII, and the mean rating differences were very small, the masking effect could be larger for the female than for the male. As shown in Fig. 4, the discussed plummet effect for a falling sound was kept under the masking, although the difference between the source room (room 1) and first receiving room (room 2) becomes smaller because of the increase in the first receiving room (room 2).

In terms of the reverberation, as shown in Table V by the columns of piano − cello and female − male, for the

TABLE VII. The correlation distance similarity tests ($r$) between the approaching-sound-source and receding-sound-source groups with the investigated source types.

| | Without background sound | | | | | With background sound | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Room | Babel | Piano | Cello | Female | Male | Babel | Piano | Cello | Female | Male |
| 1 | 0.976 | 0.065 | 0.870 | 0.839 | 0.853 | 0.992 | 0.645 | 0.955 | 0.804 | 0.918 |
| 2 | 0.800 | −0.245 | 0.820 | 0.574 | 0.543 | 0.809 | 0.691 | 0.819 | 0.775 | 0.641 |
| 3 | 0.568 | 0.671 | 0.591 | 0.859 | 0.810 | 0.849 | 0.508 | 0.901 | 0.682 | 0.768 |
| 4 | 0.885 | 0.486 | 0.936 | 0.728 | 0.836 | 0.846 | 0.742 | 0.800 | 0.754 | 0.603 |

TABLE VIII. The mean rating differences in the loudness between the conditions without and with background sound ($B$) for either the approaching-sound-source group ($A$) or the receding-sound-source group ($R$) with the independent $t$-test ($p$).

| | Piano | | Cello | | Female | | Male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ |
| 1 | 0.1 | 0.3 | −0.2 | 0.0 | −0.2 | 0.1 | −0.1 | 0.1 |
| 2 | 0.7[a] | −0.5 | −0.3 | −0.4 | 0.2 | 0.0 | 0.0 | 0.0 |
| 3 | 0.3 | −0.4 | −0.5 | −0.4 | 0.2 | 0.0 | 0.0 | 0.0 |
| 4 | −0.7[a] | −0.7 | −0.8[b] | −0.9[b] | −0.8[b] | −0.2 | −0.6[a] | −0.1 |

[a] $p < 0.05$.
[b] $p < 0.01$.

approaching-sound-source and receding-sound-source groups, in the source room (room 1), there was no mean rating difference for the piano and cello and the female and male, whereas the mean rating differences in the receiving rooms (rooms 2–4) could be observed. Table IX shows the mean rating differences in the reverberation between the conditions without and with background sound. It is worthwhile to note that the masking effect of the background sound imposed on the reverberation was limited in the source room (room 1), according to the results of the independent $t$-tests ($p$) because no significant differences ($p < 0.05$) were observed, although the mean ratings increase to a greater or lesser degree. Meanwhile, it is also limited for the voice in the receiving rooms (rooms 1–3).

To measure the perceptual priority between the piano and cello, as well as the female and male, with all of the perceptual attributes, Table X shows the distance similarity correlation tests for the approaching-sound-source and receding-sound-source groups. The results revealed that either a rising or falling voice was similar for the female and male. However, the similarity between the piano and cello for the approaching-sound-source group was 0.682, whereas the similarity of the receding-sound-source group was −0.604 in the source room (room 1). This indicates that the rising music could be perceptually identical, whereas the perception of the falling music could be distinctly different between the piano and cello.

It is interesting to note that the similarity between the music decreased with the increasing source distance for the approaching-sound-source group, and this pattern was opposite to the pattern of the receding-sound-source group, of which the similarity gradually increased the increasing source distance. This suggests an increasing symmetry between the music (defined in this paper as the convergence effect); that is, when approaching the sound source, the perception difference between the piano and cello gradually grows from unsimilar to similar, surprisingly, and when receding from the sound source, the perception difference between the piano and cello also gradually grows from unsimilar to similar. This increasing symmetry was also observed under the masking effect. Furthermore, as shown in Table X, the masking effect considerably increased the similarities between the piano and cello for the receding-sound-source group, which indicates that the masking effect was larger for the falling music than the rising music.

## IV. DISCUSSION

### A. The distinction of spatial information for the approaching and receding sound sources

The results in this study show that it could be insufficient to confirm the validity of the listening tests when dealing with the dynamic auditory perception for the approaching and receding sound sources by disregarding the actual spatial information, especially in indoor space. The condition of a room as a source or receiving room is important. For instance, the perceptual difference between the approaching and receding sound sources was greater for the receiving rooms than for the source room. This indicates a challenge to apply the guidelines in a natural environment as the source room does not experience the greatest perceptual differences

TABLE IX. The mean rating differences for the reverberation between the conditions without and with background sound ($B$) for either the approaching-sound-source group ($A$) or the receding-sound-source ($R$) with the independent $t$-test ($p$).

| | Piano | | Cello | | Female | | Male | |
|---|---|---|---|---|---|---|---|---|
| Room | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ | $A-A(B)$ | $R-R(B)$ |
| 1 | 0.3 | −0.6 | 0.1 | −0.7 | 0.5 | 0.0 | 0.2 | 0.4 |
| 2 | 0.1 | −1.4[b] | −0.1 | −0.1 | −0.2 | −0.1 | 0.4 | −0.1 |
| 3 | −0.7 | −0.7[a] | −1.2[b] | −0.8[b] | −0.2 | −0.7 | −0.2 | −0.2 |
| 4 | −1.0[a] | −2.0[c] | −1.2[c] | −1.5[b] | −1.0 | −1.3[b] | −0.9[a] | −0.6[a] |

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

TABLE X. The correlation distance similarity tests ($r$) between the music (piano and cello) and voice (female and male) sources tested with or without background signals for the approaching-sound-source ($A$) and receding sound-source groups ($R$).

| | Without background sound | | | | With background sound | | | |
|---|---|---|---|---|---|---|---|---|
| | Music | | Voice | | Music | | Voice | |
| Room | $A$ | $R$ | $A$ | $R$ | $A$ | $R$ | $A$ | $R$ |
| 1 | 0.682 | −0.604 | 0.973 | 0.964 | 0.685 | 0.325 | 0.936 | 0.886 |
| 2 | 0.463 | −0.205 | 0.886 | 0.955 | 0.433 | 0.607 | 0.960 | 0.856 |
| 3 | 0.042 | 0.268 | 0.916 | 0.964 | 0.769 | 0.849 | 0.837 | 0.886 |
| 4 | 0.659 | 0.681 | 0.851 | 0.956 | 0.598 | 0.947 | 0.453 | 0.898 |

in this context. When a listener enters or leaves the source room, the perception of the source room is not of a great difference; however, the perception of the room connected to the source room is highly dependent on a directional basis. The plummet effect, although many stimulus parameters have yet to be investigated, suggests that the ratios of the room volumes between the source and receiving rooms could have an impact on the perceptual priority of a rising intensity sound. Additionally, the ratings of the receiving rooms (rooms 2–4) were observed to be equivalent for the falling tones, which is not observed by the rising tone, as shown in Fig. 4. It is worthwhile to note that because rooms 2–4 were almost identical for the spatial information, the effect of the "expectation" is demonstrated to be stronger for a falling level than for a rising level.

Another distinction to consider with the spatial information for the dynamic auditory perception is that it makes the experiment more realistic by increasing the time of the experiments. Most experiments with only listening dimensions were using the sound sequence within 100 ms–2 s or intervals (e.g., 30 ms). However, in this study, the time taken for each participant in the VE to complete one survey in each room was much longer (e.g., 1 min) akin to the RE. Therefore, the conclusions in this study, e.g., the range of the loudness for the voices was larger for the approaching-sound-source group than for the receding-sound-source group, and the rating difference between rooms 1 and 2 was larger for the receding-sound-source group than for the approaching-sound-source group, are the results of a longer-term temporal effect with the interaction with the actual spatial information.

The study has inspired a series of options for future work. For instance, this research was developed under the assumption that the background noise for each room was identical; however, future work could explore the situations where one room has a particularly high level of background noise. Another avenue for further research could be exploring the situations in which the process of rising levels in the approaching sound source in sequential spaces is interrupted or restarted by additional sound events. Finally, the asymmetry patterns demonstrated in this study are a fundamental phenomenon of a stationary single sound source and background noise, using samples of no specific content. Apart from the voice and music, which are common sounds in indoor building environments, it would also be interesting to explore specific environmental sound sources, such as birds or water.

### B. The asymmetry of directional aspects of noise—The revelation of "Babel, 2001"

Considering the primary source, "Babel, 2001" is referred to the noise comprising a mix of music and voice, the asymmetry in the perceived loudness of the directional aspects have been demonstrated in this study. However, in general, the occurrence of an asymmetry in the directional aspects with "Babel, 2001" is not as distinguishable as music, as shown in Table VII by comparing the columns of "Babel, 2001" to those of music. According to Ref. 11, the rising intensity harmonic sounds have perceptual priority. The asymmetry occurred with the harmonic sounds but not with the broadband noise. Naturally occurring continuous broadband noise is less common and can be attributed to multiple sound sources (e.g., crowd noise). However, additionally, we found that although, in general, they are comparatively more similar in this context, for some perceptual attributes, e.g., the loudness, the asymmetry also occurred with noise when background noise and reverberations are present but not in a quiet environment.

The dynamic localization of these sources could be more significant in an indoor environment compared to a natural environment because of the crucial issues in the organization of the traffic flows concerning people's perception toward an approaching or receding large crowd, which is commonly seen in the indoor large scale. This is also akin to the cocktail party effect, wherein a group of people gathered in one space are considered as multiple sound sources. Therefore, as a revelation of "Babel, 2001," the asymmetry of the directional aspects shows that once people approach the crowd, they could overestimate the size of the crowd by perceiving an increasing level, which could be the largest in the room connected to the source room, and underestimate those conditions of the crowd by perceiving a falling sound in the same actual change in level.

## V. CONCLUSION

This study reported on auditory perception in large sequential public spaces with a listener in motion with a stationary primary sound source. Thus, virtual experiments were performed with *in situ* surveys for the validation. The reproduction of 10.0 dB less than the actual sound level was demonstrated to be necessary to imitate the feeling of being at the actual location in the indoor spaces. The existence of significant differences was determined in the auditory perception within these spaces to understand how the approaching and receding sound sources are perceived on the same path, and three major effects were found. In indoor building environments, this difference could create an overestimation of the source, and the rising intensity can signal the movement toward the source. This bias was not necessarily stronger at higher levels (i.e., the source room), suggesting that the rising loudness is more critical either close to a sound source or loud in the receiving rooms. The results indicate the importance of the dynamic rising loudness and an asymmetry of the dynamic intensity change.

For the effect of the approaching and receding sound sources, it is concluded that

- the rising levels when approaching the sound source were rated higher in each room (approach effect) and changed more than the falling levels despite having the same actual change in the level. This indicates that a change in the direction is an important factor in the perception of the dynamic loudness. None of the findings are predicted by the traditional psychophysical laws derived by simulating a static listener, indicating that there are differences between the static and dynamic loudness perceptions. Furthermore, the difference between the rising and falling levels was greater for the receiving rooms than the source room and greatest for the room connected to the source room;
- the masking effect impairing the perceptual priority of the rising sound was profound in the receiving rooms but not in the source room itself;
- the results of the loudness could not be extended to the reverberation; and
- the overall asymmetry of directional aspects occurring with broadband noise and voice was not as distinguishable as with music, but for some perceptual attributes, e.g., the loudness, the perception disparity does exist.

For the effect of the sound source type, it is concluded that

- as was found with the music and voice sound, a gradual changing level was perceived to change in an equivalent manner in the loudness between the piano and cello. For the female and male, it is on a directional basis which was larger for the approaching sound source. The difference between the sources was greater for the room connected to the source room rather than the source room itself. The rating of the room connected to the source room received a sharp drop (plummet effect), which was only observed for the receding sound source. The magnitude was dependent on the source type, of which the music magnitudes were larger than the voice magnitudes;
- the masking effects on the loudness and reverberation were limited in this study, although they were larger for the receiving rooms than for the source room and for the female than for the male; and
- an increasing symmetry of the overall perception between the different source types was observed (convergence effect) either by the approaching or receding sound source.

The approaching sound source can be a critically important environmental event. Naturally occurring continuous broadband noise can be the result of multiple sound sources such as a crowd. Overall, the results suggest that the modal and technical measures of the perceptual attributes, which do not account for the directional aspects, are oversimplifications. The prediction of the perceptual attributes, e.g., the loudness, can be improved by considering the dependence of the perceptual importance on the spatial position and direction toward the sound source.

## ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant No. 740696) and the Natural Science Foundation of China (Grant No. 51778169). We would like to thank the authorities of Tate Modern, London and all of the participants of the field surveys and virtual experiments. Our great thanks are also extended to Associate Professor Dr. Ting Zhang, and Shanghai Dianji University is gratefully acknowledged.

[1]BS 5839-1:2017: *Fire Detection and Fire Alarm Systems for Buildings—Part 1: Code of Practice for Design, Installation, Commissioning and Maintenance of Systems in Non-Domestic Premises* (British Standards Institute, London, 2017).
[2]BS 5839-6:2019: *Fire Detection and Fire Alarm Systems for Buildings—Part 6: Code of Practice for the Design, Installation, Commissioning and Maintenance of Fire Detection and Fire Alarm Systems in Domestic Premises* (British Standards Institute, London, 2019).
[3]L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics*, translated by T. J. Schultz (Applied Science, New York, 1982), Vol. 1.
[4]E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Model*, 2nd ed. (Springer, Berlin, 1999).
[5]W. Ellermeier and S. Schrödl, "Temporal weights in loudness summation," in *Fechner Day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophysics*, edited by C. Bonnet (Université Louis Pasteur, Strasbourg, 2000), pp. 169–173.
[6]B. Pedersen and W. Ellermeier, "Temporal weights in the level discrimination of time-varying sounds," J. Acoust. Soc. Am. **123**, 963–972 (2008).
[7]K. Dittrich and D. Oberfeld, "A comparison of the temporal weighting of annoyance and loudness," J. Acoust. Soc. Am. **126**, 3168–3178 (2009).
[8]J. Rennies and J. L. Verhey, "Temporal weighting in loudness of broadband and narrowband signals," J. Acoust. Soc. Am. **126**, 951–954 (2009).
[9]A. D. Baddeley, "Short-term memory for word sequences as a function of acoustic semantic and formal similarity," Q. J. Exp. Psychol. **18**, 362–365 (1966).
[10]D. Oberfeld and T. Plank, "The temporal weighting of loudness: Effects of the level profile," Atten. Percept. Psychophys. **73**, 189–208 (2011).
[11]J. Neuhoff, "Perceptual bias for rising tones," Nature **395**, 123–124 (1998).
[12]N. S. Bruce and W. J. Davies, "The effects of expectation on the perception of soundscapes," Appl. Acoust. **85**, 1–11 (2014).
[13]B. Wang, J. Kang, and W. Zhao, "Noise acceptance of acoustic sequences for indoor soundscape in transport hubs," J. Acoust. Soc. Am. **147**, 206–217 (2020).
[14]T. Oberman, K. Jambrošić, M. Horvat, and B. B. O. Šćitaroci, "Using virtual soundwalk approach for assessing sound art soundscape interventions in public spaces," Appl. Sci. **10**(6), 2102 (2020).
[15]A. Rungta, N. Rewkowski, R. Klatzky, M. Lin, and D. Manocha, "Effects of virtual acoustics on dynamic auditory distance perception," J. Acoust. Soc. Am. **141**, EL427–EL432 (2017).
[16]A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss," Atten. Percept. Psychophys. **78**, 373–395 (2016).

[17]D. A. Guth, R. G. Long, R. S. W. Emerson, P. E. Ponchillia, and D. H. Ashmead, “Blind and sighted pedestrians’ road-crossing judgments at a single-lane roundabout,” Hum Factors. **55**, 632–642 (2013).
[18]D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, “Effects of room reflectance and background noise on perceived auditory distance,” Perception **18**, 403–416 (1989).
[19]A. W. Bronkhorst, “The cocktail-party problem revisited: Early processing and selection of multi-talker speech,” Atten. Percept. Psychophys. **77**, 1465–1487 (2015).
[20]S. Haykin and Z. Chen, “The cocktail party problem,” Neural Comput. **17**, 1875–1902 (2005).
[21]G. Kidd, Jr., T. L. Arbogast, C. R. Mason, and F. J. Gallun, “The advantage of knowing where to listen,” J. Acoust. Soc. Am. **118**, 3804–3815 (2005).
[22]G. Brett, *Cildo Meireles* (Tate, London, 2008).
[23]See https://www.tate.org.uk/art/artworks/meireles-babel-t14041 for the character of the primary source Babel, 2001 (Last viewed 10 December 2021).
[24]See https://en.wikipedia.org/wiki/Tower_of_Babel for the details of the Tower of Babel (Last viewed 10 December 2021).
[25]See https://www.tate.org.uk/whats-on/tate-modern/exhibition/cildo-meireles/cildo-meireles-explore-exhibition/cildo-meireles-4 for the character of the primary source Babel, 2001 (Last viewed 10 December 2021).
[26]L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York, 2002).
[27]M. Barron, *Auditorium Acoustics and Architectural Design* (Spon/Chapman and Hall, London, 1993).
[28]R. J. Hawkes and H. Douglas, “Subjective acoustic experience in concert auditoria,” Acta Acust. Acust. **24**, 235–250 (1971).
[29]J. Kang, *Urban Sound Environment* (CRC Press, London, 2006).
[30]F. Aletta and A. Astolfi, “Soundscapes of buildings and built environments,” Build. Acoust. **25**, 195–197 (2018).
[31]M. Zhang and J. Kang, “Towards the evaluation, description, and creation of soundscapes in urban open spaces,” Environ. Plan B Urban Anal. City Sci. **34**(1), 68–86 (2007).
[32]J. Kang and M. Zhang, “Semantic differential analysis of the soundscape in urban open public spaces,” Build Environ. **45**(1), 150–157 (2010).
[33]ISO/DIS 12913-2: *Acoustics-Soundscape—Part 2: Data Collection and Reporting Requirements* (International Standard Organization, Geneva, Switzerland, 2017).
[34]A. S. Sudarsono, “Soundscape composition and relationship between sound objects and soundscape dimensions of an urban area,” Doctoral dissertation, University of Salford, 2017.