OPEN ACCESS Freely available online
S. Kim and W. Howie, "Influence of the Listening Environment
on Recognition of Immersive Reproduction of Orchestral Music Sound Scenes"
J. Audio Eng. Soc., vol. 69, no. 11, pp. 834–848, (2021 November).
DOI: https://doi.org/10.17743/jaes.2021.0035

Influence of the Listening Environment on Recognition of Immersive Reproduction of Orchestral Music Sound Scenes

SUNGYOUNG KIM, *AES Member,* **AND WILL HOWIE,** *AES Member*
(sxkiee@rit.edu) (wghowie@gmail.com)

Electrical, Computer and Telecommunication Engineering Technology, Rochester Institute of Technology, Rochester, NY

This study investigates how a listening environment (the combination of a room's acoustics and reproduction loudspeaker) influences a listener's perception of reproduced sound fields. Three distinct listening environments with different reverberation times and clarity indices were compared for their perceptual characteristics. Binaural recordings were made of orchestral music, mixed for 22.2 and 2-channel audio reproduction, within each of the three listening rooms. In a subjective listening test, 48 listeners evaluate these binaural recordings in terms of overall preference and five auditory attributes: perceived width, perceived depth, spatial clarity, impression of being enveloped, and spectral fidelity. Factor analyses of these five attribute ratings show that listener perception of the reproduced sound fields focused on two salient factors, spatial and spectral fidelity, yet the attributes' weightings in those two factors differs depending on a listener's previous experience with audio production and 3D immersive audio listening. For the experienced group, the impression of being enveloped was the most salient attribute, with spectral fidelity being the most important for the non-experienced group.

0 INTRODUCTION

0.1 Room, an Active Aural Entity

Throughout human history, the acoustics of constructed environments have influenced human experience. The acoustic conditions of a given space actively influence how a listener perceives sonic information within it; the physical characteristics of a space reconfigure a sound's sonic profile. A solo violin, for example, might produce a wildly different sonic impression depending on whether it was heard in a car or a concert hall. Blesser and Slater [1] call this reconfigured set of sonic profiles "aural architecture." A pattern of reflections is key to understanding the aural architecture of an enclosure. A single reflection not only mimics the origin of the sound source but also reveals unique properties of the reflective surface. Each listening room, therefore, has its own unique identity of aural architecture that assists people in recognition of the surrounding geolocational information [1, pp. 37–41 & 336–347] and influences our emotional status [1, pp. 335]. Recent research in archaeoacoustics has demonstrated that sociocultural information can be extracted from architectural acoustics when the perspectives of human listeners are studied [2] and especially when culturally specific sound sources are considered [3].

Although all spaces have aural significance for human experience, perhaps the most obvious examples in recent history are music venues such as concert halls. As modern musical traditions developed, both musicians and audiences have observed that acoustical conditions alter their appreciation of performed music. Composers of Western art music have sought to explore the interaction of players and rooms within their music, while instrument makers consider the room response within their designs [4][5].

What about a reproduced sound field in a room? A tonmeister is trained to capture a sound field with an optimal balance between a direct sound source and the spatial characteristics of the recording space [6]; the captured sound field can be described by a distinct acoustic profile through which a listener can sense its aural architecture. When reproducing this distinct sound field within another room, an inevitable conflict occurs between the acoustic profiles of the recording and reproduction environments. A recent study by Klein et al. [7] shows that a listener's memory of a room's acoustics influences their experience hearing reproduced sound fields. The authors call this cognitive room conflict "room divergence." The listener's previous experience could form an expectation-oriented bias [8] and modify their listening experience of any reproduced sound field within a given space.

Kaplanis et al. conducted a systematic study [9] that investigated listeners' perception and preference of reverberation in rooms for two-channel reproduction. They presume that room acoustics and associated reverberations cause the reproduced sound scene to be distorted, "as the recorded signal is superimposed with the spatiotemporal response of the reproduction room." To determine the perceptual influences of room reverberations, they captured nine distinct spatial room impulse responses for two-channel reproduction from four listening rooms using a 3D vector intensity probe. Subsequently the captured room impulse responses were processed to playback via a 43-channel loudspeaker array in an anechoic chamber. This "auralization" technique allowed listeners to virtually experience nine target room acoustics and compare them in real time. The study found that perception of "Reverberance" is highly correlated with "Width & Envelopment," negatively correlated with "Proximity," and decorrelated with (perceptual dominance of) "Bass." In addition, the results show that listeners prefer high proximity and less reverberant room acoustics for two-channel reproduction.

Moreover, Toole has exhaustively researched room interactions between loudspeakers and loudspeaker-reproduced sound fields [10]. Through a series of controlled listening experiments and associated psychoacoustic models, Toole discovered how listeners' affective responses and perceptual judgments are influenced by peripherals in the listening chain. Toole specifically showed how the combination of loudspeakers' radiation patterns and wall reflections within the listening room affect both timbral and spatial attributes of reproduced music.

Toole coined the term "Circle of Confusion" to explain the "never-ending cycle of subjectivity" [10, p. 19] within audio production, listening, and evaluation. Essentially, a lack of technical standards within audio control room design has lead to a situation in which "recordings vary, sometimes quite widely, in their sound quality, spectral balances, and imaging." A recording is made while monitoring with a certain set of loudspeakers, while one or many entirely different set(s) of loudspeakers are used in the reproduction and subjective evaluation of said recording. This is further complicated by interactions between monitor speakers and room acoustics that may modify the reproduced sound field of the original recording, thus rendering the subsequent evaluation process erroneous. For five-channel stereophonic audio recording and reproduction [11], Toole's "Circle of Confusion" is equally applicable: there is a strong "interaction of multiple loudspeakers and the listening room when those loudspeakers are reproducing a multichannel recording" [10, p. 100][12].

A reproduced sound field, therefore, should be understood as a complex aural product of a three-way interaction across the acoustics of the recording venue, acoustics of the reproduction room, and sound reproduction medium. Among those three aspects, the acoustics of the reproduction room and sound reproduction medium are determined by an end-user's circumstances, resulting in numerous variations.

To avoid or at least reduce influence associated with such variations, there has been an effort to create industry standards for listening room acoustics and monitoring characteristics, including [13] and [14]. With those standards, researchers and developers can conduct controllable and repeatable experiments (acoustic measurements and psychoacoustic observations) and provide meaningful guidelines for listeners in non-standard situations. It is worthwhile to mention again that the room acoustics and loudspeaker characteristics are intertwined, influencing a listener's perceptual and cognitive responses. In this paper, therefore, the authors treat the interaction of those two aspects as a single variable, to be referred to as the "listening environment."

0.2 The New Era of Immersive Audio

We are currently experiencing an extensive, fast growth of "immersive" audio technologies and research.[1] This is also true for the broader area of virtual reality, which includes the paradigms known as augmented reality, mixed reality, and extended reality often collectively called "XR technologies." How applicable will existing guidelines for sound reproduction be for these new types of listening experiences? Slater [16] asserts that a visual immersive experience is directly related to the peripheral system capacity, i.e., the fidelity that preserves "equivalent real-world sensory modalities." Therefore it follows that reproduction channels may serve as the peripheral system capacity required for auditory immersion, which has been a motivation for the development of multi-layer loudspeaker reproduction to properly incorporate sonic information from above (and sometimes below) the listeners [17, chap. 7].

Oode et al. [18] found that for sound reproduction systems that include height channels, the sensation of listener envelopment (LEV) increases as the number of loudspeakers increases. 22.2 Multichannel Sound (22.2), an example of one of several standardized immersive audio formats, utilizes nine loudspeakers in the upper layer, 10 loudspeakers in the middle layer, and three loudspeakers in the bottom (floor) layer. Originally proposed by the Science and Technology Research Laboratories of the Japan Broadcasting Corporation (NHK) in 2003, the system is designed to provide audiences with a highly precise, natural, and fully "immersive" experience [19]. A key idea behind many such large-scale 3D audio reproduction systems is that as the number of sound reproduction locations increases, we approach a point where one could capture and reproduce all of the details that make up a space's unique sonic "fingerprint" as independently as possible. This may imply that

[1] A discussion on the definition of auditory immersion and immersive experience is beyond the scope of this topic. Interested readers are advised to read [15], which is an excellent review and proposal of the concept of "immersion."

Table 1. Room dimensions (m), reverberation times (T30), and clarity indices (C80) for the three listening rooms. The values are calculated from the impulse response between the front center loudspeaker and listening position.

Room	Width	Length	Height	Volume	T30	C80
1	6.0 m	7.8 m	3.2 m	149.8 m^2	150 ms	39.02 dB
2	5.9 m	4.2 m	2.8 m	69.4 m^2	370 ms	14.22 dB
3	6.8 m	6.8 m	4.5 m	208.1 m^2	310 ms	20.15 dB

room acoustics become less influential for a listening environment that possesses a greater number of loudspeakers, eventually becoming a nuisance variable.

The current study is based on the following research question:

> By manipulating immersive audio content and the reproduction system, can we reduce undesired room-loudspeaker interactions and corresponding variations in listener perception of stimuli across various sonic attributes? More simply, when compared with stereo reproduction, does immersive reproduction deliver a listening experience that reduces the influence of the listening environment on the listening experience?

To investigate this question, three listening environments, along with a reference anechoic chamber, were compared. 22.2 was the sound reproduction format used in each space. The proceeding section introduces the listening rooms' acoustic profiles as well as reproduction systems. This is followed by a description of the main experiment and a subjective test of those rooms and describes the creation of stimuli, the participants, and test results. Finally a summary is provided of the study's implications for understanding the relationship between physical and perceptual factors within a listener's recognition of 3D sound fields, based on the coupling of listening room acoustics and 3D sound reproduction systems.

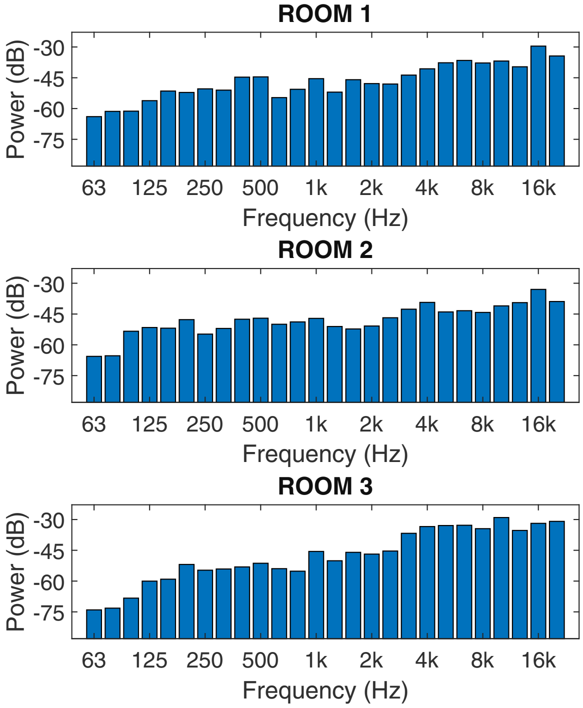

Fig. 1. Power spectra at the listening position of the three listening environments. These spectra are results from an interaction of both the room acoustics and 22 loudspeakers.

1 PHYSICAL CHARACTERISTICS OF LISTENING ENVIRONMENTS

We chose three listening rooms, labeled as ***1***, ***2***, and ***3***, that show distinct acoustic profiles and represent various listening scenarios ranging from a professional mixing situation to causal domestic listening.

The acoustic characteristics of the three rooms are summarized in Table 1. Room ***1***'s dimension ratios, reverb time, and background noise level fulfill ITU-R BS.1116 [13] requirements for multichannel audio critical listening environments. This room is equipped with a total of 28 Musikelectronic Geithain ME-25 loudspeakers. Room ***2*** does not comply with ITU-R BS.1116 recommendations: it was chosen as an example of an ordinary, everyday listening environment. The loudspeakers are GENELEC 8020B. Room ***2*** is a converted conference room with added acoustical treatment to remove strong wall reflections. This room is the smallest of the three yet has the longest reverberation time. Room ***3***, the largest room of the three, has a slightly longer reverberation time than ITU-R BS.1116 recommends: the space was originally designed as a recording studio. Background noise and floor area, however, are compliant with ITU-R BS.1116. Room ***3*** is equipped with 22 KS Digital C5 Powered Studio Monitors.

Distinct differences were found within the rooms' reverberation times (T30), as shown in the second right-most column of Table 1. Ashok et al. [20] previously simulated these three rooms using CATT-Acoustics software and found that clarity (C80), among many other measured acoustic characteristics, was the parameter that differs the most across the three listening rooms. C80 compares the ratio between energy within the first 80 ms of the impulse response with energy from the remaining part of the impulse response and has been found to relate well to perceived clarity within a space [21]. Results of the current measurements (the right-most column of Table 1) correspond well with the previous simulation results, showing that the three rooms differ within the clarity metric. Specifically Room ***2*** does not properly diffuse nor absorb acoustic energies, resulting in an overabundance of reflected sound after the first 80 ms of the impulse response. While the room is equipped with several bass traps to reduce unwanted low-frequency modes, it seems to fail to effectively suppress strong early reflections.

Fig. 1 displays the power spectra of the three listening environments as an interaction of both the room acoustics and all loudspeakers, excluding the two subwoofers, at the listening position, measured with an omni-directional microphone (DPA 4006). Each listening environment shows unique spectral characteristics. For example, rooms ***1*** and

3 have relatively weak low-frequency responses (< 100 Hz), while Room ***2*** shows relatively weak high-frequency responses (5 to 8 kHz). In addition, the standard deviations of one-third octave power spectra were calculated: 13.43 dB for Room ***1***, 11.30 dB for Room ***2***, and 20.07 dB for Room ***3***.

Having obtained the acoustic information for each listening environment, a series of subjective listening tests were undertaken. Again, "listening environment" in this study refers to a combination of the listening room and its loudspeakers; the acoustic information displayed in Table 1 and Fig. 1 are also results from the interaction of room and loudspeaker. The authors understand the importance of controlling the loudspeaker to observe sole room-induced influences, as Kaplanis et al. did [9]. However, even in professional recording and mixing studios, type and brand of loudspeakers can differ across a wide range of options, following the engineer's preference or budgetary restriction. Practically, it would be impossible to standardize a certain type of loudspeaker for a given type of audio reproduction. Rather we chose to focus on perceptual effects and their structure for immersive audio reproduction, which may be induced from the listening environment, characterized by room acoustics and loudspeakers.

2 SUBJECTIVE EVALUATION

2.1 Stimuli Preparation

Two immersive recordings of orchestral music were prepared for this subjective evaluation. The first piece is "Mars" from *The Planets* by Gustav Holst, performed by the National Youth Orchestra of Canada. The recording was made in the Music Multimedia Room (MMR) at McGill University in Montreal, Canada. Extensive details for the recording procedure are documented in [22]. The second piece is "Kaido-Tosei" by Kiyoshi Nobutoki, performed by the Tokyo University of the Arts Orchestra and Choir. The recording was made at Sougakudou, Tokyo University of the Arts in Tokyo, Japan.[2] These orchestral performances were recorded and mixed optimally for both 22.2 and two-channel reproduction.

For each of the three listening rooms, binaural recordings were made of the 22.2 and two-channel orchestral music reproductions, using the same Head And Torso Simulator, Brüel & Kjær 4100D, located at the listening positions.[3] Thus the optimal listening perspective from each room was captured for each reproduction condition. For the subjective evaluation, a reference condition was also included. When listeners are asked to quantify the perceived magnitude of an auditory precept, many contextual factors (non-experiment variables) modulate their responses, resulting in low intra and inter-subject consistency. Zielinski et al. [8] published a comprehensive review of such non-experimental variables encountered in audio quality listening tests. One way to reduce potential effects from those variables (including the range equalizing bias) is to incorporate a reference or a direct anchor [8, pp. 445] to which all stimuli are compared. For the current experiment, the author chose an anechoic room located at the Research Institute of Electrical Communication (RIEC), Tohoku University [23], as a reference to which listeners can compare perceptual attributes.

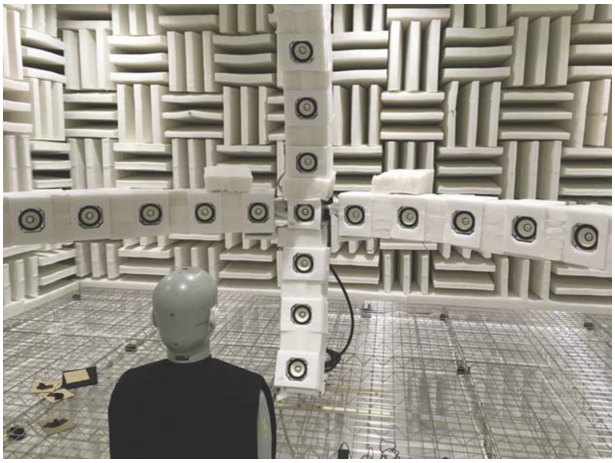

Fig. 2. A picture of the RIEC anechoic room at Tohoku University, the reference room used in this study. The Head and Torso Simulator (the Brüel & Kjær 4100D) is located at the center of the loudspeaker array for the binaural recording of the 22.2 and two-channel reproduced orchestral music pieces.

The RIEC anechoic room is equipped with a horizontal loudspeaker array of 36 loudspeakers (every 10°) and another array in the median plane (at 10° resolution from –30° to +90°) as shown in Fig. 2. The horizontal array height is motorized for precise adjustment from the floor to the ceiling. The same binaural microphone used to capture stimuli in the other listening environments was placed at the center of the loudspeaker array and captured the reproductions of 22.2 and two-channel orchestral music. This resulted in a total of 16 listening test stimuli (four listening environments, two reproduction formats, and two musical selections.)

The loudness of all resultant binaural stimuli was matched through a combination of (1) aligning their RMS levels and (2) manual adjustments by professional recording engineers.

2.2 Participants

Four listener groups were formed at the institutions where the four listening rooms are located: McGill University (10 participants), Tokyo University of the Arts (TUA; 10 participants), Rochester Institute of Technology (RIT; 21 participants), and Tohoku University (ToU; seven participants). The listeners in the McGill and TUA groups all had previous experience hearing sound scenes reproduced through a 22.2-channel system, while the RIT and ToU listeners had no such experience. All participants had normal listening skills and no difficulty for everyday auditory

[2] The recording was released by Naxos label and related information can be found in http://naxos.jp/news/nycc-27300 (Japanese).

[3] The first author brought the Rochester Institute of Technology's Brüel & Kjær 4100D binaural microphone to the four listening rooms and recorded with it.

communications. Their ages ranged from 19 to 47 years. Each subject performed the complete test twice. Stimuli were reproduced via Sennheiser HD-650 headphones through a Grace Design m902 headphone preamplifier. A custom graphic user interface (GUI) and data collection system was created using Cycling ’74’s Max software.

It is worth acknowledging that the authors could not use each listener’s individualized head-related transfer function (HRTF) for this study because it would have been highly impractical for each listener to visit each of the four listening rooms under investigation in different countries. Therefore, as previously discussed, subjects evaluated auditory images projected through the ears of the Brüel & Kjær 4100D binaural microphone and Sennheiser HD-650 headphones. While there are on-going efforts to adapt individualized HRTFs in 3D audio home listening experiences, this method of using non-individualized HRTFs for binaural rendering also reflects the reality typical of most “current” commercial virtual reality/augmented reality.

2.3 Perceptual Attributes

Participants evaluated eight binaural stimuli based on five salient perceptual attributes: perceived width, perceived depth, spatial clarity, impression of being enveloped, and spectral fidelity of the sound field across. These attributes were chosen based on previous literature on spatial audio perceptual evaluation. Below are the definitions of the attributes:

- Spectral Fidelity (or Balance): “The balance in tone color (spectrum) variation of the sound image” [24].
- Apparent Source Width (ASW): “The apparent horizontal spatial extent of the sound image” [25].
- (Spatial) Clarity: “Integration of various sound components of the sound image. The clearer the sound, the more details you can perceive in it” [26].
- Envelopment: “The sense of being enveloped by the sound field (both horizontally and vertically)” [27].
- Depth: “The overall impression of the depth of the sound image. Takes into consideration both overall depth of scene, and the relative depth of the individual sound sources” [28].

The authors did not include “localization” or “locatedness” of sound source(s) in this specific study for two reasons: (1) previous studies have already thoroughly covered the topic of room-related influence on auditory localization [29][30], and (2) the current study does include the attribute “(Spatial) Clarity,” a multi-dimensional attribute that broadly includes positional precision of sound sources.

For the attributes preference and spectral fidelity, listeners directly compared eight stimuli for each musical selection, using continuous sliders ranging in value from 0 to 100. As such, participants were asked to convert relative perceptual magnitudes to a number range of 0 to 100, with 50 being the baseline. Below 50, therefore, would indicate not-preferred or ill-balanced spectrum, while above would indicate preferred or well-balanced spectrum. When participants found it difficult to quantify those two attributes, they were allowed to compare the eight stimuli with a “reference” condition, activated using a separate button within the GUI. The “reference” condition (a hidden reference) was the anechoic space with 22.2 reproduction. Subjects were not told that the reference corresponded to a specific stimulus but rather that it represented a rating of “50” on the perceptual scale and could therefore use the reference to estimate the relative strengths and weaknesses of the eight stimuli within the GUI. This reference signal, therefore, provided all participants with the same baseline.

Similar to the method described above, a GUI using ranged sliders and a reference signal could have been used to rate the stimuli in terms of the various spatial attributes under investigation, as is often seen in perceptual audio evaluation [31–33]. Quantifying specific spatial attributes, however, is challenging when one is asked to generate a number to describe them. For example, if a listener rated a perceived width of 50, what is the meaning of this specific number? Researchers, therefore, have endeavored to develop a new method that indirectly elicits and quantifies the spatial attributes of a sound field [34–37]. The general assumption of these methods is that within the cognitive mechanism there is a dual-coding system [38] that allows perceived characteristics to be recognized with language and another to be recognized with mental imagery.

Spatial attributes in auditory evaluation can be represented and quantified effectively and with less variance when using non-verbal methods, drawing and pointing, resulting in fewer cognitive loads and chances for misinterpretation. Two recent immersive audio studies by Martin et al. [39][40] made use of such a methodology, asking subjects to *draw* the perceived extent of a sound image rather than describe it. The first author used this method in a previous subjective evaluation [41] and found that this drawing method is faster, results in more consistent quantification, and is relatively language-independent, since this method is less reliant on exact translations of the attributes under investigation to the subject’s native language. In the current study, therefore, four spatial attributes (Width, Depth, Envelopment, and Clarity) were quantified via an indirect drawing method that asked listeners to determine the shape of a perceived sound image using the GUI.

Fig. 3 shows the custom GUI for the indirect drawing method. By controlling number boxes on the right side [within a range between 0 (minimum) and 100 (maximum)], a participant can control the corresponding shape (a dark-purple oval) of the orchestra image. For instance, the number boxes of “Auditory Source Width” and “Depth” respectively adjust left-right extent (width) and front-back extent (depth) of the oval and display the new shape in the GUI. Subjects were asked to match their perceived auditory image size with the oval shape for those two attributes. “Envelopment” adjusts the strength of an outer brown-colored circle, indicating how much the sound field envelops the subject. As for “Clarity,” the number box increases the thickness of two black lines inside the purple oval. Darker color of the black lines indicates that details of sound components are more clearly reproduced and perceived, while

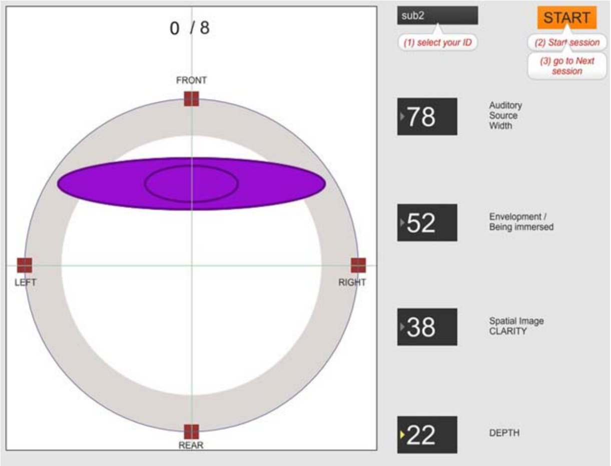

Fig. 3. A custom graphic user interface to convert perceived magnitudes to corresponding ratings of four spatial attributes. Four number boxes on right side respectively control graphical representations of four attributes (Width, Depth, Envelopment, and Clarity).

pale color indicates the opposite case. Instead of comparing a given stimulus against another auditory stimulus, participants use an equivalent visual imagery as a reference to quantify those four spatial attributes, evaluating four spatial attributes for one stimulus at a time.

3 SUBJECTIVE TEST RESULTS

Preliminary data analyses showed that the musical selection of the stimulus had no significant effect on either preference or perceptual attributes. The listeners' previous audio production experience and multichannel immersive music listening experience, however, may be an important factor for the quantification of perceived impressions. A recent study by Howie et al. found audio production experience to be a strong predictor of listener performance in 3D audio evaluation [42]. However, that same study found that previous experience hearing 3D audio had no effect on listener performance. In the current listening test, the McGill and TUA groups consisted of listeners with both such types of previous experience (experience group, 20 participants), while RIT and ToU listeners had little or no such previous experience (no-experience group, 28 participants). Therefore the authors focused on the analysis of a potential effect from this experienced listener group while merging data from the two different musical pieces.

3.1 Listeners' Preference

The preference ratings (dependent variable) were analyzed through a mixed two-way Analysis of Variance (ANOVA), in order to examine the effects of listening environment (ROOM: within-subject independent variable) and listener group—experienced vs. inexperienced (GROUP: between-subject independent variable). Our previous study [43] demonstrated a significant difference in listeners' hedonic and perceptual ratings between 22.2 and two-channel systems. Therefore we did not include those two reproduction systems in the current ANOVA model and conducted two separate ANOVA analyses for each system. As seen in Table 2 there is a statistically significant two-way interaction between GROUP and ROOM on preference ratings for both the 22.2-channel system as well as the two-channel [$F(3, 282) = 8.302$, $p < 0.001$].

In this study it was assumed that more loudspeakers in the reproduction system would reduce undesired room-loudspeaker interactions and such systems (with more loudspeakers) may allow listeners to experience very similar sonic phenomena, regardless of the listening environments. Since the ANOVA results seem to reject this hypothesis, we focused on another metric that differentiates between 22.2 and two-channel reproduction. The right-most column of Table 2 shows generalized eta square (ges, η^2) values, indicating each independent variable's *effect size* for the corresponding row. The ROOM effect size of the 22.2-channel system is 0.034, which is regarded as a "small" effect, while the counterpart is 0.255, a "large" effect. These values indicate that the ROOM variable had a stronger effect on listeners' preference ratings for two-channel reproduced music.

Fig. 4 displays the mean and confidence intervals of the preference ratings for the 22.2-channel (top panel) and two-channel (bottom panel) reproduction conditions. Note that the preference ratings should not be associated with the "superiority" of any of the three listening environments used in the experiment. The main purpose of this study is to investigate the influence of the listening environment on listener perception, not to find an answer to the "which-would-be-better?" question. Therefore we report the subjective ratings without the names of three rooms, instead using aliases (1, 2, and 3 for the three rooms and R for the reference anechoic room). The ROOM factor specifically decreases its effect for the "non-experienced" listener group. This may indicate that many "un-critical" listeners prefer an immersive sound system regardless of differences in listening environment.

3.2 Perceptual Attributes

Fig. 5 shows the mean ratings and confidence intervals for both preference and spectral fidelity for the 22.2-channel system. Two ratings are significantly correlated: Pearson correlation $r = 0.4326$ ($p < 0.001$) for the 22.2-channel system and $r = 0.4476$ ($p < 0.001$) for the two-channel system. This figure illustrates a similar pattern in the listeners' responses to both attributes (preference and spectral fidelity), which implies that listener preference was influenced by spectral changes associated with the listening environment. From the power spectra of the three listening environments (Fig. 1), readers may induce that listeners preferred the listening environment that produced the smallest spectral variation—Room ***2***. So in this study listeners appear to have associated the idea of "timbral fidelity" with a listening environment that has the least influence on the tonal balance of the recording, i.e., a room with a relatively flat power spectra.

Table 2. The two-way mixed ANOVA results of preference ratings. The results show that interaction between the ROOM and GROUP factor is significant ($p < 0.001$) regardless of the number of reproduction format. However the effect size of ROOM factor [the ges (η^2) values] is different for the 22.2-channel (0.034: small effect size) from the two-channel (0.255: big effect size).

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	39.986	<0.001	*	0.100
ROOM	3	282	4.473	<0.001	*	0.034
GROUP:ROOM	3	282	8.302	<0.001	*	0.061
			22.2-channel reproduced music			

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	27.169	<0.001	*	0.071
ROOM	3	282	43.612	<0.001	*	0.255
GROUP:ROOM	3	282	4.631	<0.001	*	0.035
			Two-channel reproduced music			

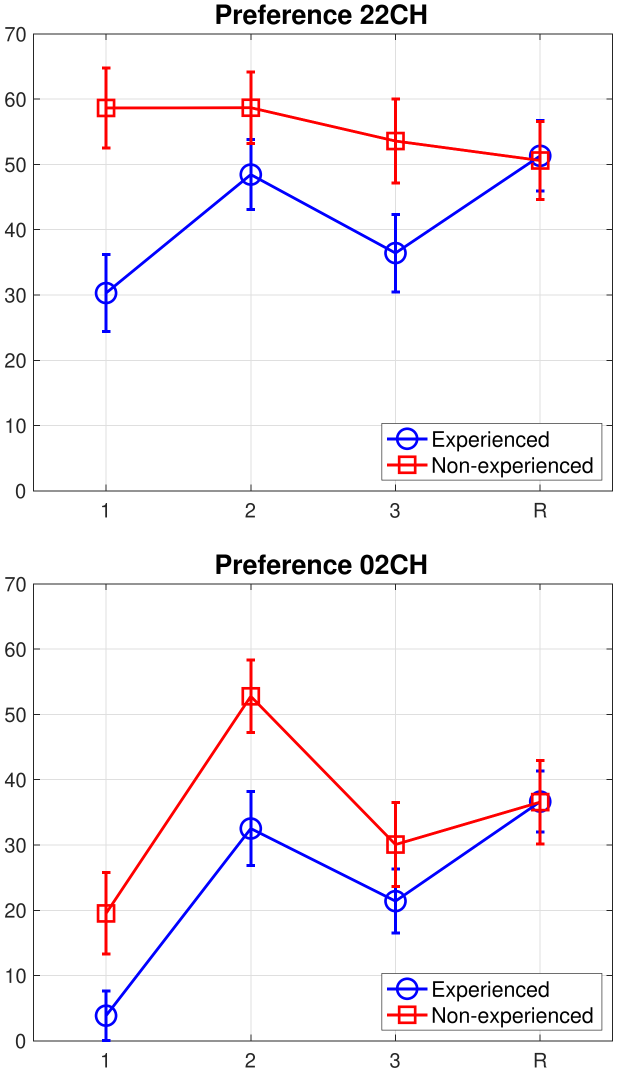

Fig. 4. Mean and confidence interval of preference ratings of four listening rooms (1, 2, 3, and R) for two reproduction systems: the 22.2-channel (top panel) and two-channel (bottom panel).

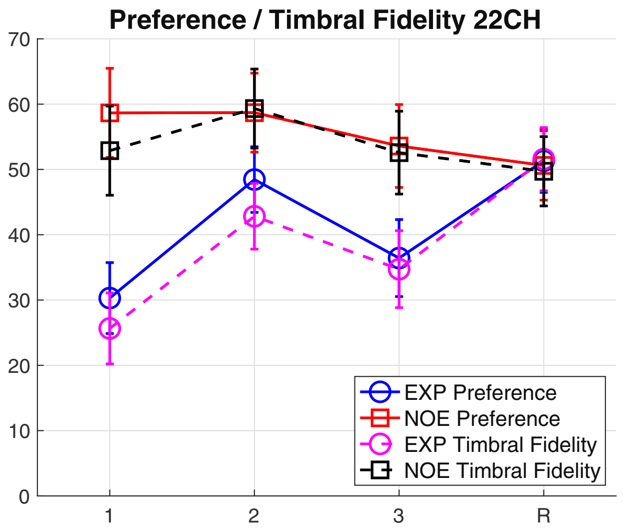

Fig. 5. Means and confidence intervals of preference ratings (solid lines) and spectral fidelity ratings (dashed lines) of four listening rooms. 20 experienced listeners' ratings are symbolized with circles while 28 non-experienced listeners are indicated with squares.

However at the same time it is not clear why Room ***1*** (with smaller spectral variation) was less preferred than Room ***3***. Here we may be seeing the influence of the source material. Unlike the test signals typically used to measure rooms, most musical signals do not have an even spectral balance. Different musical examples, therefore, may elicit "high" or "low" spectral fidelity in different reproduction environments depending on the genre/arrangement/mix aesthetic, etc. The source material for the current study, orchestral music, covers the full range of musical frequencies produced by acoustic instruments. If for example the arrangement or orchestration of a piece of music does contribute to a significant amount of low-frequency content in a recording, then power spectra differences below 160 Hz within two different rooms may not be very noticeable or important to listeners. The source material for the current study, orchestral music, covers the full range of musical frequencies produced by acoustic instruments. Perhaps, in this instance, the more robust low-frequency power spectra in Room ***1*** paradoxically could have resulted in a low-frequency reproduction within the recordings that was aesthetically too strong or exaggerated, which relatively impacted the perceived spectral fidelity ratings.

The two-way mixed ANOVA results of spectral fidelity ratings (Table 3) also reveal a similar pattern with the preference ratings, showing that an interaction between the ROOM and GROUP factor is significant ($p < 0.001$) regardless of the reproduction format. Specifically, the ROOM effect size [the ges (η^2) values] of the 22.2-channel system is 0.054, which is regarded as a "small" effect, while the two-channel is 0.3, a "large" effect. These values indicate that the ROOM variable had a stronger effect on listeners' spectral fidelity ratings for two-channel reproduced music. More loudspeakers seem to reduce differences in spectral fidelity.

The spatial attribute ratings collected from participants suggest some important implications as to how listeners recognize a given listening environment. Statistical analyses of the ratings show that the listening environment for two-channel reproduced music significantly modulated listeners' judgment of all spatial attributes except "Depth," the perpendicular extension of a sound image. Moreover, the mean values of all two-channel stimuli ratings were smaller than those associated with 22.2 reproduction. Since the spatial attribute ratings for all two-channel stimuli showed a similar tendency, we chose to exclude the two-channel data in the subsequent analyses and focused on the 22.2-channel ratings. Means and confidence intervals of four attribute ratings associated with 22.2-channel reproduced music are plotted in Fig. 6.

Results indicate that the 22.2-channel system allowed listeners to consistently judge spatial attributes in four different playback rooms (including the reference anechoic room). Two-way mixed ANOVA tests (similar to the ANOVA test for the preference ratings) were conducted for all spatial attributes: no interaction between the ROOM and GROUP factor was found for all attributes. The GROUP factor significantly modulates the ASW [$F(3, 282) = 22.963$, $p < 0.001$], LEV [$F(3, 282) = 4.644$, $p = 0.034$], and (Spatial) Clarity [$F(3, 282) = 15.134$, $p < 0.001$] ratings. The ROOM factor only significantly modulates the LEV rating [$F(3, 282) = 6.162$, $p < 0.001$], but the effect size [the ges (η^2) values] is quite small (0.031). The ANOVA results of all five attribute ratings are listed in Table 4 located in APPENDIXA.

Considered together, this data suggests that the perceived magnitudes of spatial attributes are not significantly influenced by the acoustics of the room (except a small effect on LEV). The "Depth" ratings are of particular interest considering the static nature of binaural sound capture, with actual listening typically involving head movement and rotation. Non-dynamic representations of binaural sound (without a head tracker) have been known to fail in generating an "externalized sound image" [44]. The two-channel reproduction and its binaural capture thus failed to deliver successful perpendicular extension of the sound field for all four rooms. However, the participants perceived extended depth for the 22.2-channel reproduced binaural stimuli without head tracking and corresponding dynamic binaural processing, even within the non-experienced listener group. It is possible that the ambient sound information in the recordings, when reproduced from the side, rear, and height channels, helped participants to better perceive an externalized sound image.

Also of interest is a comparison between ASW and LEV ratings. The two top panels (left: ASW and right: LEV) of Fig. 6 show no significant difference between the ASW and LEV ratings from the non-experienced group (square symbols). For this group, each listening environment (room plus loudspeaker) delivered the same magnitude of ASW and LEV. In contrast the experienced group (circle symbols) clearly differentiated the ASW from LEV. In other attribute ratings, the non-experienced group tended to rate more "generously" by giving higher scores; Fig. 4 shows that this group's preference ratings are much higher than the experienced group. This is similar to results from several previous studies by Olive and his co-authors investigating preferences among loudspeakers [45][46] or headphones [47], wherein it was observed that experienced and naïve listeners tend to use different parts of the rating scale, with less-experienced listeners generally giving higher ratings.

However, the same group's ASW ratings are significantly lower than the other. ASW and LEV are both associated with distinct spatial characteristics of a reproduced sound field. ASW relates to the perception of an auditory image's horizontal dimension, while LEV relates to a listener's impression of being enveloped or surrounded by sound. Hence "extent" is a key word within the definition of ASW, while "sense" is more pertinent for LEV. It is worthwhile to note that LEV is also closely related with auditory immersion. These findings suggest a cognitive difference between the two groups: the experienced group could cognitively separate the concept of the physical extension of a sound image from being enveloped (or immersed), while the non-experienced group mixed those two concepts.

These possible cognitive differences between the two listener groups can be better visualized using factor analysis. Two factors were extracted with eigenvalues equal to or greater than 1 after the Promax rotation. The corresponding loadings of five attributes are plotted in Fig. 7. The left panel is for the experienced group and right panel for the non-experienced group. For the experienced group, the first factor (42.7%) is associated with LEV as the sole leading attribute, and the second factor (16.8%) with a combination of four other attributes. In contrast, the non-experienced group factorizes all the space-related attributes (ASW, LEV, Clarity, and Depth) as the first factor (46.9%) and spectral fidelity as the second factor (18.8%).

Both groups seem to rely on LEV and spectral fidelity as base axes of their semantic spaces, yet the dominant factor and corresponding attribute is quite different between the two groups. The experienced group primarily recognizes the differences of listening environments based on LEV, while the other group does so with spectral fidelity. In the

Table 3. The two-way mixed ANOVA results of spectral fidelity ratings. The results show that interaction between the ROOM and GROUP factor is significant ($p < 0.001$) regardless of the reproduction format. However, similar to the preference ratings, the effect size of ROOM factor [the ges (η^2) values] is different for the 22.2-channel (0.054: small effect size) from the two-channel (0.3: big effect size).

Effect	DFn	DFd	F	*p*	*p* < .05	ges
GROUP	1	94	52.859	<0.001	*	0.115
ROOM	3	282	7.037	<0.001	*	0.054
GROUP:ROOM	3	282	7.961	<0.001	*	0.061
			22.2-channel reproduced music			
Effect	DFn	DFd	F	*p*	*p* < .05	ges
GROUP	1	94	13.787	<0.001	*	0.050
ROOM	3	282	63.031	<0.001	*	0.300
GROUP:ROOM	3	282	4.001	<0.001	*	0.026
			Two-channel reproduced music			

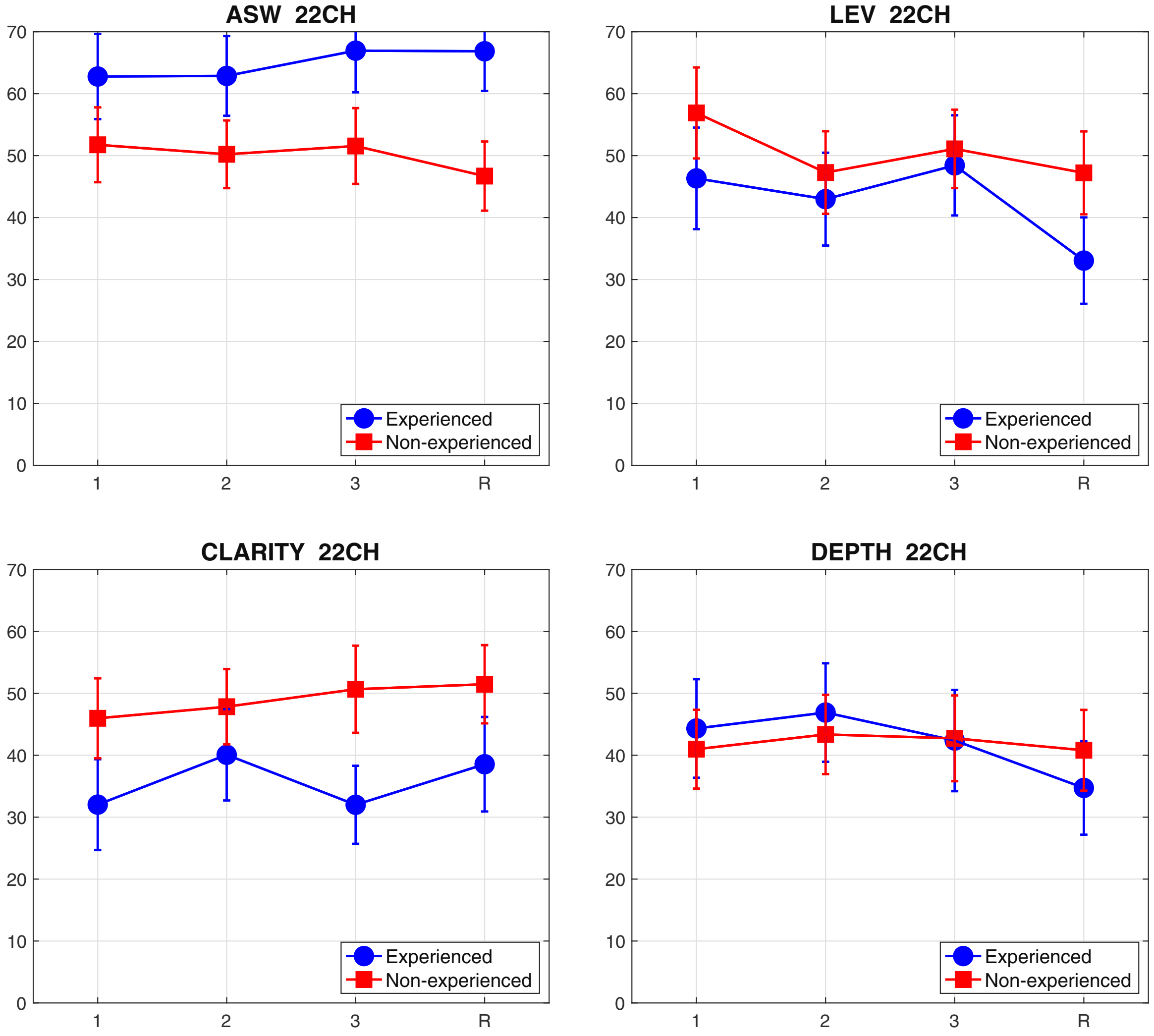

Fig. 6. Means and confidence intervals for four spatial attribute ratings associated with the 22.2-channel system. Twenty experienced listeners' ratings are plotted with circle symbols and 27 non-experienced listeners with square symbols. The top-left panel is for the attribute ASW, top-right for LEV, bottom-left for Clarity, and bottom-right for the Depth.

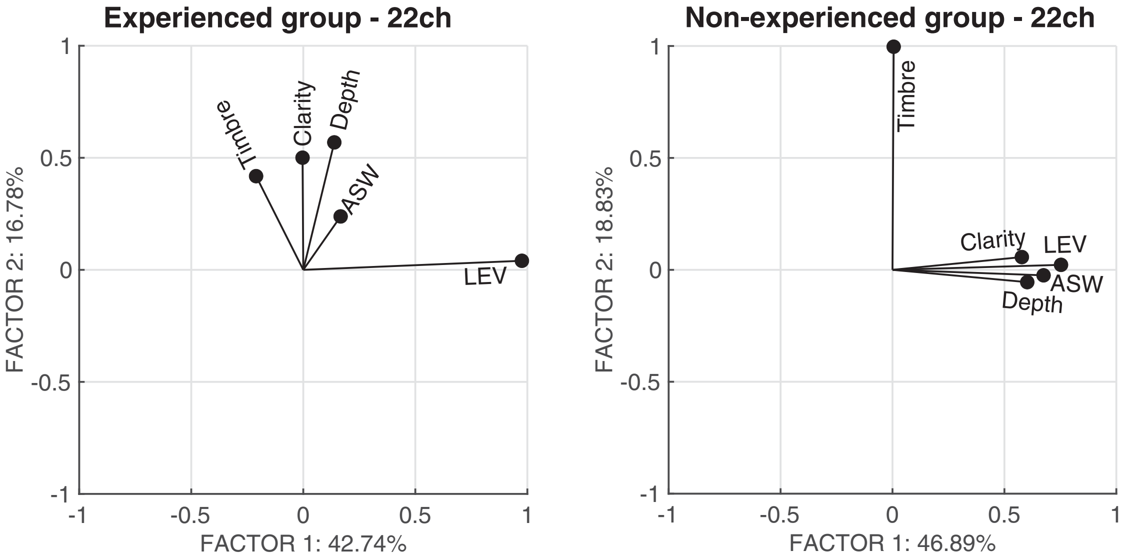

Fig. 7. A semantic space derived from a factor analysis visualizing the interrelation of five attributes. Factor 1 has a strong relation with LEV and Factor 2 has a relation with spectral fidelity for both the experienced and non-experienced groups.

two-factor semantic space, one axis is for a spatial characteristic and another for a timbral characteristic. This finding aligns with results from a previous study on multichannel audio [48] that showed a difference between naïve and experienced listeners in the weightings of "timbral and spatial fidelities."

In summary this study found that when an optimally recorded orchestral musical piece was reproduced through a 22.2-channel system, listeners perceived a similar spatial dimensionality for the reproduced sound field in any of the listening environments under investigation, even though these environments differed in terms of physical dimensions and reflecting surfaces. However the unique acoustical features of each reproduction room changed listeners' impressions of the spectral fidelity of the reproduced sound fields, regardless of the reproduction system (stereo or 22.2).

4 DISCUSSION

This study investigates several contextual effects in listener recognition processes of reproduced orchestral music, which include the listening environment and listeners' previous experience with audio production and 3D immersive audio listening. These environmental factors can have an influence on a subject's perception of a given stimuli, resulting in a particular form of systemic "bias" that affects a psychophysical test [49]. While Zielinski asserts [8] that it was not possible to find strong evidence or "direct data" within relevant literature supporting the existence of influences associated with contextual effects, other studies [50][51] support their significance. This contrast occurs because a given contextual effect interacts with other contextual effect(s), as the current study shows. In other words, a certain contextual effect may be latent until another condition starts inducing meaningful difference(s). For example, a previous study [52] examined the influence of musical selection (as a context) on listeners' preference of surround microphone techniques, showing that the preference is significantly modulated by musical selection (an aggregate of the composition in question, performance practice used by the musician, and performance itself). This implies that people listen with respect to the semantic content of musical performance paradigms within the reproduced sound field. Hence the current study incorporates orchestral music selections that are recognizably different in terms of musical style/content and instrumentation.

However in this study variations from two distinct musical selections did not induce significant differences in preference or perceptual attribute ratings. This does not imply that musical selection is meaningless within the evaluation of reproduced sound fields for 22.2. Rather it indicates that differences between the two orchestral music selections used in the current study were not large enough to modulate the listeners' responses. The two primary experimental factors in this study, the listening environment and previous audio experience among subjects, their interactions, and effects on listener perception, may be less significant or insignificant under different experimental conditions. It is therefore not possible to generalize the results of the current study to all potential sound reproduction scenarios, which was never the objective of this paper. Ideally further studies will be devised to collect more local evidence of influences associated with various contextual factors and listeners' preference and to form a compilation of case studies for the research community. In particular a relationship between preference ratings and other perceptual ratings is of interest yet is beyond the scope of the current study. This is partially due to a lack of physical parameters that can account for perceptual attribute magnitudes influenced by contextual factors.

The results discussed in SEC. 4.2 suggest that in this study listener perception of many spatial attributes under investigation were not significantly influenced by the listening environment (the combination of room and loudspeaker). This is especially important for multichannel-audio control rooms, where one wishes to avoid the room-in-room effect. Interestingly, however, the listening environments in this study did appear to have a small effect on perception of LEV. This is quite surprising, given that LEV is typically associated with diffuse lateral sound reflections arriving after the first 50–80 ms [21][53], the kind of acoustic energy that is not typically well represented in small rooms. Toole [10] and Griesinger [54] have both discussed envelopment in small rooms. Both assert that envelopment associated with reproduced sound within a small room must primarily be a result of diffuse reflected energy within the recording itself, reproduced from surrounding loudspeakers, particularly those at or near the side walls. Toole, however, suggests that the sense of envelopment created by a multichannel recording could be enhanced if the side walls of the listening room are treated in such as way as to provide diffuse reflections of sounds emanating from loudspeakers on opposite walls [10]. This idea becomes particularly relevant for 3D audio formats with dedicated side channels (i.e., loudspeakers positioned at +/−90 degrees), such as 22.2 or Dolby Atmos. It is possible then that the small differences in LEV associated with the rooms used in the current study are a result of their acoustical design: the rooms that better reflect and diffuse reverberant sound from within the recordings give listeners a better sense of envelopment or spaciousness.

In the current study, listeners can be broadly separated into two different "experience" groups: those with previous experience or proficiency with audio production and hearing 3D audio reproduced through a 22.2-channel system and those without either of said types of experience. Howie et al. have examined the effect of different types of previous experience on listener performance [42] and skill level [55] in 3D audio evaluation. As with the current study, Howie et al. used immersive recordings of orchestral music as stimuli. In [42], it was found that among several different types of previous experience, audio production experience was the best predictor of listener consistency in making preference or ranking judgments of 3D audio stimuli. For those trained or working as audio engineers, even in stereo, it is normal to divide and analyze a sound scene based on many complex factors covering a wide range of spatial, timbral, dynamic, and musical considerations. It follows logically, therefore, that listeners with this type of previous experience would be better able to use complex and multi-modal concepts such as LEV to differentiate between listening environments or sound scenes, particularly if they were already familiar with immersive audio reproduction.

For inexperienced or "naïve" listeners, who are generally unfamiliar with this type of advanced sound scene analysis, timbre might be a more effective or natural attribute for differentiating between sound scenes, immersive or otherwise; timbre is one of the few factors a consumer can actively listen for or control when comparing loudspeakers or adjusting a filter on a stereo system. In the current study, it is not possible to say which of the two previous experience types, audio production or immersive audio listening, had a greater impact on the perception of the experience group, since they are fundamentally correlated within the data. In Howie et al.'s study [42], previous experience "hearing 3D audio" was not found to have an effect on listener performance. However, the focus of that study was consistency of making evaluations, not how the experience modulated a listener's perception of the sound scene or listening space. In the current study, it follows somewhat more logically that previous experience hearing immersive sound scenes through a 22.2 system could be a factor in "training" subjects to focus more or equally on spatial attributes, such as LEV, rather than purely timbral differences between sound scenes. This begs the question: is LEV an auditory attribute that is understood differently in comprehension of immersive reproduced sound field, depending on previous training or experience? Further experiments would be required to examine this question adequately.

5 CONCLUSION

This study investigated the effect of the listening environment (listening room acoustics and loudspeaker) on listener perception of loudspeaker-based music reproduction. Three listening environments were compared in terms of both their physical and perceptual characteristics. The acoustics of the reproduction rooms used to create the stimuli for this study differed in terms of physical dimensions, reverberation time (T30), and acoustic clarity (C80) values. Results of a subjective listening test show that for reproduction of two different orchestral music sound scenes, an immersive 3D audio system reduced room-induced effects on listeners' recognition of the target music pieces, as compared with a traditional two-channel, stereo reproduction system. Immersive sound reproduction (22.2 in this study) was found to preserve enhanced spatial attributes of the reproduced orchestral music, regardless of the listening environment.

However, for both the immersive and stereo playback conditions, listeners' perceptual ratings related to the spectral fidelity of stimuli were modulated by the differing room acoustics of each listening environment. In addition listeners' previous experience with audio production or loudspeaker-based 3D audio reproduction may have differentiated their perceptual responses, resulting in an idiosyncratic cognitive base difference between the experienced and non-experienced groups. The experienced group appears to better understand small inter-differences in the four spatial attributes under investigation. For the experienced group, LEV was the primary factor in distinguishing between the four listening environments, while spectral fidelity was the primary distinguishing factor for the non-experienced group.

6 ACKNOWLEDGMENT

This paper was partially supported by Yamaha Corporation and was carried out under the Cooperative Research Project Program of the RIEC, Tohoku University. The author extends special thanks to Prof. Richard King (McGill University), Dr. Toru Kamekawa (Tokyo University of the Arts), and Dr. Shuichi Sakamoto for granting the use of orchestral music recordings and access to the listening rooms used in this study.

7 REFERENCES

[1] B. Blesser and L.-R. Salter, *Spaces Speak, Are You Listening? Experiencing Aural Architecture* (MIT Press, Cambridge, MA, 2006).

[2] M. A. Kolar, *Archaeological Psychoacoustics at Chavín de Huántar, Perú*, Ph.D. thesis, Stanford University, Stanford, CA (2013).

[3] M. A. Kolar, "Situating Inca Sonics: Experimental Music Archaeology at Huanúco Pampa, Perú," in M. Stöckli and M. Howell (Eds.), *Flower World – Music Archaeology of the Americas / Mundo Florido – Arqueomusicología de las Américas*, vol. 6, pp. 13–46 (Ekho Verlag, Berlin, Germany, 2020).

[4] M. Forsyth, *Buildings for Music: The Architect, the Musician, and the Listener From the 17th century to the Present Day* (MIT Press, Cambridge, MA, 1985).

[5] J. Meyer, *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers* (Springer, New York, NY, 2009), 5th ed.

[6] M. Dickreiter, *Tonmeister Technology: Recording Environments, Sound Sources, and Microphone Techniques* (Temmer Enterprises, New York, NY, 1989).

[7] F. Klein, S. Werner, and T. Mayenfels, "Influence of Training on Externalization of Binaural Synthesis in Situations of Room Divergence," *J. Audio Eng. Soc.*, vol. 65, no. 3, pp. 178–187 (2017 Mar.). https://doi.org/10.17743/jaes.2016.0072.

[8] S. Zielinski F. Rumsey, and S. Bech, "On Some Biases Encountered in Modern Audio Quality Listening Tests–A Review," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 427–451 (2008 Jun.).

[9] N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, and S. H. Jensen, "Perception and Preference of Reverberation in Small Listening Rooms for Multi-Loudspeaker Reproduction," *J. Acoust. Soc. Am.*, vol. 146, no. 5, pp. 3562–3577 (2019 Nov.). https://doi.org/10.1121/1.5135582.

[10] F. E. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms* (Routledge, New York, NY, 2018), 3rd ed.

[11] ITU-R, "Multichannel Stereophonic Sound System With and Without Accompanying Picture," *Recommendation ITU-R BS.775-3* (2012 Aug.).

[12] S. E. Olive and W. L. Martens, "Interaction Between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction," presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), paper 7196.

[13] ITU-R, "Methods for the Subjective Assessment of Small Impairments in Audio Systems," *Recommendation ITU-R BS.1116-3* (2015 Feb.).

[14] AES Technical Committee on Multichannel and Binaural Audio Technology, "Multichannel Surround Systems and Operations," Tech. Rep. AESTD1001.1.01-10 (2000).

[15] S. Agrawal, A. Simon, S. Bech, K. Bærensten, and S. Forchhammer, "Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 10275.

[16] M. Slater, "A Note on Presence Terminology," *Presence-Connect*, vol. 3, no. 3, pp. 1–5 (2003 Jan.).

[17] A. Roginska and P. Geluso (Ed.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (Routledge, New York, NY, 2017).

[18] S. Oode, I. Sawaya, K. Ono, and K. Ozawa, "Three-Dimensional Loudspeaker Arrangement for Creating Sound Envelopment," *IEICE Tech. Rep. EA2012-46*, vol. 112, no. 125, pp. 7–12 (2012 Jul.).

[19] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 Multichannel Sound System and Its Application," presented at the *118th Convention of the Audio Engineering Society* (2005 May), paper 6406.

[20] M. Ashok, R. King, T. Kamekawa, and S. Kim, "Acoustic and Subjective Evaluation of 22.2- and 2-Channel Reproduced Sound Fields in Three Studios," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 10018.

[21] A. C. Gade, "Acoustics in Halls for Speech and Music," in T. D. Rossing (Ed.), *Springer Handbook of Acoustics*, Spinger Handbooks Series, pp. 317–366 (Springer, New York, NY, 2014), 2nd ed.

[22] W. Howie, R. King, D. Martin, and F. Grond, "Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9797.

[23] S. Sakamoto, F. Saito, Y. Suzuki, et al., "Construction of Two Anechoic Rooms With a New Experimental Floor Structure," in *Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (INTER-NOISE)* (Hamburg, Germany) (2016 Aug.).

[24] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *J. Audio Eng. Soc.*, vol. 50, no. 9, pp. 651–666 (2002 Sep.).

[25] M. Morimoto and K. Iida, "A Practical Evaluation Method of Auditory Source Width in Concert Halls," *J. Acoust. Soc. Jpn. (E)*, vol. 16, no. 2, pp. 59–69 (1995). https://doi.org/10.1250/ast.16.59.

[26] S. Choisel and F. Wickelmaier, "Evaluation of Multichannel Reproduced Sound: Scaling Auditory Attributes Underlying Listener Preference," *J. Acoust.*

Soc. Am., vol. 121, no. 1, pp. 388–400 (2007 Jan.). https://doi.org/10.1121/1.2385043.

[27] J. S. Bradley and G. A. Soulodre, "Objective Measures of Listener Envelopment," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2590–2597 (1995 Nov.). https://doi.org/10.1121/1.413225.

[28] N. Zacharov, C. Pike, F. Melchoir, and T. Worch, "Next Generation Audio System Assessment Using the Multiple Stimulus Ideal Profile Method," in *Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (Lisbon, Portugal) (2016 Jun.).

[29] W. M. Hartmann, "Localization of Sound in Rooms," *J. Acoust. Soc. Am.*, vol. 74, no. 5, pp. 1380–1391 (1983 Nov.). https://doi.org/10.1121/1.390163.

[30] N. Kopčo and B. Shinn-Cunningham, "Auditory Localization in Rooms: Acoustic Analysis and Behavior," in *Proceedings of the 32nd International Acoustical Conference—EAA Symposium*, pp. 109–112 (Banská Štiavnica, Slovakia) (2002 Sep.).

[31] K. Hamasaki, K. Hiyamah, T. Nishiguchi, and R. Okumura, "Effectiveness of Height Information for Reproducing the Presence and Reality in Multichannel Audio System," presented at the *120th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2006 May), paper 6679.

[32] S. Kim, Y. W. Lee, and V. Pulkki, "New 10.2-Channel Vertical Surround System (10.2-VSS); Comparison Study of Perceived Audio Quality in Various Multichannel Sound Systems With Height Loudspeakers," presented at the *129th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2010 Nov.), paper 8296.

[33] H. Shim, E. Oh, S. Ko, and S. H. Park, "Perceptual Evaluation of Spatial Audio Quality," presented at the *129th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2010 Nov.), paper 8300.

[34] S. Choisel and K. Zimmer, "A Pointing Technique With Visual Feedback for Sound Source Localization Experiments," presented at the *115th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2003 Oct.), paper 5904.

[35] N. Ford, F. Rumsey, and B. de Bruyn, "Graphical Elicitation Techniques for Subjective Assessment of the Spatial Attributes of Loudspeaker Reproduction – A Pilot Investigation," presented at the *110th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2001 May), paper 5388.

[36] J. Usher and W. Woszczyk, "Design and Testing of a Graphical Mapping Tool for Analyzing Spatial Audio Scenes," in *Proceedings of the AES 24th International Conference on Multichannel Audio: The New Reality International Conference on Multichannel Audio: The New Reality*, pp. 157–170 (Banff, Canada) (2003 Jun.), paper 11.

[37] J. Usher and W. Woszczyk, "Visualizing Auditory Spatial Imagery of Multi-Channel Audio," presented at the *116th Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2004 May), paper 6054.

[38] A. Paivio, "The Relationship Between Verbal and Perceptual Codes," in E. C. Carterette and M. P. Friedman (Eds.), *Handbook of Perception, Volume VIII: Perceptual Coding*, pp. 375–397 (Academic Press, New York, NY, 1978). https://doi.org/10.1016/B978-0-12-161908-4.50017-6.

[39] B. Martin, R. King, and W. Woszczyk, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part I," presented at the *141st Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (2016 Sep.), paper 9613.

[40] B. Martin, D. Martin, R. King, and W. Woszczyk, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part II," presented at the *147th Convention of the Audio Engineering SocietyConvention of the Audio Engineering Society* (2019 Oct.), paper 10265.

[41] S. Kim, N. Shime, and M. Ikeda, "Enhanced Presence of Electronic Orchestral Music (Part I): A Comparison of Loudspeakers and Their Perceptual Effect," presented at the *14th Regional Convention of the Audio Engineering Society, Japan Section Regional Convention of the Audio Engineering Society, Japan Section* (Tokyo, Japan) (2009 Jul.).

[42] W. Howie, D. Martin, S. Kim, T. Kamekawa, and R. King, "Effect of Audio Production Experience, Musical Training, and Age on Listener Performance in 3D Audio Evaluation," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 782–794 (2019 Oct.). https://doi.org/10.17743/jaes.2019.0031.

[43] S. Kim, R. King, T. Kamekawa, and S. Sakamoto, "Recognition of an Auditory Environment: Investigating Room-Induced Influences on Immersive Experience," in *Proceedings of the AES International Conference on Spatial Reproduction – Aesthetics and Science* (Tokyo, Japan) (2018 Jul.), paper P6-1.

[44] K. Inanaga, Y. Yamada, and H. Koizumi, "Headphone System With Out-of-Head Localization Applying Dynamic HRTF (Head Related Transfer Function)," presented at the *98th Convention of the Audio Engineering Society* (1995 Feb.), paper 4011.

[45] 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.).

[46] S. E. Olive, "Some New Evidence That Teenagers and College Students May Prefer Accurate Sound Reproduction," presented at the *132nd Convention of the Audio Engineering Society* (2012 Apr.), paper 8683.

[47] S. E. Olive, T. Welti, and E. McMullin, "The Influence of Listeners' Experience, Age, and Culture on Headphone Sound Quality Preferences," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), paper 9177.

[48] F. Rumsey, S. Zieliński, R. Kassier, and S. Bech, “On the Relative Importance of Spatial and Timbral Fidelities in Judgments of Degraded Multichannel Audio Quality,” *J. Acoust. Soc. Am.*, vol. 118, no. 2, pp. 968–976 (2005 Aug.). https://doi.org/10.1121/1.1945368.

[49] E. C. Poulton, *Bias in Quantifying Judgments* (Lawrence Erlbaum Associates, Hove, UK, 1989).

[50] D. Begault, “Preferences Versus References: Listeners as Participants in Sound Reproduction,” in *Proceedings of the Spatial Audio & Sensory Evaluation Techniques Workshop* (Guilford, UK) (2006 Apr.).

[51] H. L. Meiselman (Ed.), *Context: The Effect of Environment on Product Design and Evaluation* (Elsevier, Duxford, UK, 2019).

[52] S. Kim, M. de Francisco, K. Walker, A. Marui, and W. L. Martens, “An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique,” in *Proceedings of the AES 28th International Conference on The Future of Audio Technology – Surround Sound and Beyond International Conference on The Future of Audio Technology – Surround Sound and Beyond* (Piteå, Sweden) (2006 Jun.), paper 9-2.

[53] T. Hanyu and S. Kimura, “A New Objective Measure for Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections,” *Appl. Acoust.*, vol. 62, no. 2, pp. 155–184 (2001 Feb.).

[54] D. Griesinger, “Spatial Impression and Envelopment in Small Rooms,” presented at the *103rd Convention of the Audio Engineering Society Convention of the Audio Engineering Society* (1997 Sep.), paper 4638.

[55] W. Howie, D. Martin, S. Kim, T. Kamekawa, and R. King, “Effect of Skill Level on Listener Performance in 3D Audio Evaluation,” *J. Audio Eng. Soc.*, vol. 68, no. 9, pp. 628–637 (2020 Sep.). https://doi.org/10.17743/jaes.2020.0050.

THE AUTHORS

Sungyoung Kim

Will Howie

Sungyoung Kim received a B.S. degree from Sogang University, Korea, and Master of Music and Ph.D. from McGill University, Canada. He joined the Electrical, Computer, and Telecommunication Engineering Department of the Rochester Institute of Technology (RIT) as an assistant professor in 2012 (and has been an associate professor since 2018). His research interests are rendering and perceptual evaluation of spatial audio, digital preservation of aural heritage, and auditory training for hearing rehabilitation.

•

Will Howie is an audio engineer, music producer, and researcher, specializing in multichannel immersive audio. He is currently the Audio Producer for the Vancouver Symphony Orchestra and a Recording Engineer with CBC/Radio-Canada. Will holds a Ph.D. in Sound Recording from McGill University, where he taught undergraduate and graduate-level courses in audio production. His research focuses on production techniques for and perception of 3D audio.

A.1 ANOVA RESULTS OF FOUR SPATIAL ATTRIBUTE RATINGS

Table 4. The two-way mixed ANOVA results of four spatial attribute ratings for the 22.2-channel system. The results show that no significant interaction between the ROOM and GROUP factor for all four attributes. In addition, the GROUP factor significantly modulates the ASW, LEV, and (Spatial) Clarity ratings while the ROOM factor (as a listening environment) only significantly modulates the LEV rating. Please refer to SEC. 3.2 for detailed explanations.

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	22.963	<0.001	*	0.107
ROOM	3	282	0.462	0.709		0.002
GROUP:ROOM	3	282	1.213	0.305		0.007

ASW ratings (22.2-ch.)

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	4.644	0.034	*	0.025
ROOM	3	282	6.162	<0.001	*	0.031
GROUP:ROOM	3	282	4.631	0.169		0.009

LEV ratings (22.2-ch.)

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	15.134	<0.001	*	0.074
ROOM	3	282	1.841	0.140		0.010
GROUP:ROOM	3	282	1.257	0.289		0.007

CLARITY ratings (22.2-ch.)

Effect	DFn	DFd	F	p	$p < .05$	ges
GROUP	1	94	0.001	0.976		<0.001
ROOM	3	282	2.352.612	0.072		<0.001
GROUP:ROOM	3	282	1.255	0.290		<0.001

DEPTH ratings (22.2-ch.)