JASA ARTICLE

Singing in virtual versus real rooms: Is it the same?

Pasquale Bottalico,[1,2,a)] Carly J. Wingfield,[2] Charles J. Nudelman,[1,3] Angelo Farina,[4,b)] Joshua D. Glasner,[5] and Yvonne Gonzales Redman[1,2]

[1]*Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, Illinois 61820, USA*

[2]*School of Music, University of Illinois Urbana-Champaign, Champaign, Illinois 61820, USA*

[3]*Department of Communication Sciences and Disorders, Syracuse University, Syracuse, New York 13244, USA*

[4]*Dipartimento di Ingegneria dei Sistemi e delle Tecnologie Industriali, Universitá di Parma, Parma 43121, Italy*

[5]*Department of Otolaryngology – Head and Neck Surgery: Lions Voice Clinic, University of Minnesota, Minneapolis, Minnesota 55455, USA*

ABSTRACT:

Previous research suggests that singers adjust their vocal production in response to different acoustic environments. This study investigated how virtual and real-room acoustics influence singers' vocal performance by analyzing vibrato rate, vibrato extent, and quality ratio. Ten classically trained singers performed an unaccompanied aria in three real performance spaces and their virtual replications under four sensory conditions: real, audio-only, visual-only, and combined audiovisual virtual reality (VR). Results showed that vibrato extent and vibrato rate were moderately affected by sensory condition, where larger values in some virtual conditions are compared to real rooms. However, the magnitude of these differences was within the range of just noticeable differences, suggesting that they may not be perceptually salient. perceived singing voice supportiveness, obtained from a survey, was significantly reduced in conditions lacking auditory feedback, underscoring the role of acoustic cues for singers. Overall, the findings suggest that VR-based auralizations can approximate the experience of real acoustic environments for singers' perceived and acoustic outcomes, although the effects on proprioception and voice support warrant further investigation. Given the small sample size, these findings are preliminary and should be confirmed in studies with larger singer populations. https://doi.org/10.1121/10.0039807

(Received 3 March 2025; revised 10 October 2025; accepted 12 October 2025; published online 3 November 2025)

[Editor: Manuj Yadav] Pages: 3645–3655

I. INTRODUCTION

The acoustics of a room significantly influence what we hear, perceive, and how we produce speech and voice. For singers, variations in acoustics are far more impactful, shaping not only what they hear but also how they perform.[1–6] Singers frequently comment on how the acoustics of performance spaces affect their emotional state, technical delivery, and overall confidence. These anecdotal accounts have been investigated for decades and laid the groundwork for formal studies on the relationship between spatial acoustics and vocal performance.[7]

Despite their reliance on optimal acoustics for performances, singers often rehearse in spaces extremely different from performance venues. Teaching studios, for example, are typically small and acoustically dry, lacking the reverberation and spaciousness of concert halls. This mismatch between rehearsal and performance environments can lead to significant challenges, including unintentional adjustments to vocal technique,[2–6,8] diminished confidence, and heightened performance anxiety.[9,10] The discrepancy also limits the ability of singers to rehearse with the auditory and visual cues they will encounter in a performance setting, which can leave them underprepared for the sensory demands of live performance.

Advances in virtual reality (VR) technology present an exciting opportunity to address this gap. By combining immersive audio and visual stimuli, VR has the capability to replicate the conditions of live performance venues. This technology, already widely used for training in fields such as surgery and flight simulation, offers a new approach to enhancing vocal pedagogy and performance preparation. Before VR can be adopted as a training tool for singers, an important long-term goal is to determine whether it can accurately replicate real-world performance environments and elicit comparable vocal production responses.

Regarding vocal production, research has extensively examined vibrato, an essential feature of classical singing. Vibrato is a periodic pulsation in pitch, intensity, and timbre that enhances tonal richness and emotional expressiveness.[11,12] It is characterized by two primary parameters: vibrato extent, which quantifies the frequency deviation around the mean pitch during a vibrato cycle, and vibrato rate, or the number of vibrato cycles per second.[11] Another critical measure in vocal production is the quality ratio (QR), which reflects the strength of the singer's formant.[13] Lower QR values indicate a more prominent singer's formant, a hallmark of operatic singing.

[a)]Email: pb81@illinois.edu

[b)]Deceased.

Studies on vocal performance in real acoustic spaces have demonstrated that singers adjust their technique in response to varying acoustic feedback. For instance, Ternström[2] found that room acoustics significantly affected choir vocal production. Of the three choirs examined, in the more reverberant spaces, the youth and adult choirs reduced power, while in the less reverberant basement room, they increased it. The boys choir, meanwhile, maintained power but adjusted spectral tilt for a brighter tone in the less reverberant room. All choirs shifted to higher formant frequencies in absorbent environments, suggesting unconscious adaptation to acoustics. Research by Bottalico *et al.*[8] showed that singers alter their vibrato extent with greater extents observed in spaces with higher clarity and shorter early decay times (EDTs). Similarly, Schärer Kalkandjiev and Weinzierl[14,15] documented adjustments in tempo and timbre among instrumentalists performing in different concert halls, highlighting the influence of room acoustics on musical performance.

Whereas studies conducted in real acoustic spaces offer valuable insights, they are often limited by logistical and methodological challenges such as the difficulty of controlling environmental variables and restricted access to performance venues. Luizard and Bernadoni[4] and Luizard *et al.*[5] encountered these limitations and subsequently extended their research into virtual acoustic environments,[6] which provided greater control over external factors such as time of day and singers' psychological states. Their findings add further evidence to support the notion that singers adjust their vocal production to differing acoustic environments; however, they did not observe consistent patterns across participants, as singers employed individualized strategies. Similarly, Brereton *et al.*[16] developed a virtual singing studio designed to simulate concert hall acoustics for the study of vocal performance. Although innovative, the absence of visual stimuli in this setup reduced the realism of the simulation, which, in turn, limited the extent of vocal adaptation.

Fischinger *et al.*[3] collected objective and subjective data from a mixed adult choir performing under contrasting virtual acoustic conditions, finding that tempo and timing precision declined in the most reverberant virtual room. Similarly, Ueno *et al.*[17] and Kato *et al.*[18] examined five musicians—four instrumentalists and one baritone—under simulated room acoustic conditions, demonstrating that musicians adapt their performance to different virtual acoustic environments. Of particular relevance to the present study, Kato *et al.*[18] found that vibrato extent decreased in rooms with longer reverberation times (RTs).

Whereas all of these studies have examined singing in real and virtual acoustic environments, little research has explored singing in combined audiovisual virtual environments. Findings from speech science underscore the importance of visual feedback in vocal adaptation,[19,20] although these studies did not focus on singing. Addressing this gap, the present study investigates differences in vocal production when singing in real rooms versus singing in their virtual replicas. Specifically, it examines three key vocal parameters—vibrato rate, vibrato extent, and QR—across a range of sensory conditions, including real rooms, audio-only (A_only) virtual rooms, visual-only (V_only) virtual rooms, and combined audiovisual (comb_AV) virtual rooms. By comparing these conditions, the study aims to determine whether VR can effectively replicate live performance spaces and elicit comparable vocal adjustments.

In addition to objective vocal measures, the study also explores singers' subjective perceptions of the various real and virtual rooms. Factors, such as perceived acoustic supportiveness and peacefulness, are critical for understanding how singers experience different settings and how these perceptions influence their performance. Previous research has shown that singers' evaluations of acoustic environments can vary widely, influenced by sensory inputs and individual preferences.[8,15] This study builds on these findings by incorporating an acoustic perception survey to capture and quantify the subjective experiences of singers in real and virtual rooms. In this study, the following research questions are addressed:

(1) To what extent do singers' vocal parameters—vibrato rate, vibrato extent, and QR—vary across different sensory conditions (real, A_only, V_only, and audiovisual VR), independent of specific room acoustics?; and
(2) to what extent do singers' subjective perceptions, specifically regarding acoustic supportiveness and peacefulness, differ across sensory conditions?

This research addresses a critical gap at the intersection of acoustics, vocal performance, and VR, advancing our understanding of the sensory factors that shape singing.

II. EXPERIMENTAL METHOD

A. Participants

The use of human subjects for this research was approved by the Office for the Protection of Research Subjects at the University of Illinois Urbana-Champaign [Institutional Review Board (IRB) No. 24-0549]. Five female and five male singers (average age 25.9 years old) volunteered to take part in the experiment. The age, gender, and voice type of the ten participants are reported in Table I. The singers were all graduate students in Western classical singing, with an average duration of consistent private voice lessons equal to 9.3 yrs.

B. Room description

The three rooms used for this study were all housed within the Tina Weedon Smith Memorial Hall, a historic building completed in 1920 and located on the southwest corner of the main quadrangle of the University of Illinois Urbana-Champaign campus. The three rooms—Smith Recital Hall, Smith Memorial room, and room 204—were selected as three different sized rooms—large, medium, and small, respectively—where singers often perform or rehearse.

The Smith Recital Hall is a large concert venue with mahogany walls that seats 802 guests on the main floor and

https://doi.org/10.1121/10.0039807

TABLE I. Voice type and experience of participants.

Identification	Age	Gender	Voice Type	Years of Experience
1	33	Female	Soprano	10
2	28	Male	Baritone	12
3	24	Female	Soprano	9
4	33	Male	Tenor	15
5	24	Male	Baritone	8
6	24	Female	Soprano	10
7	25	Male	Tenor	7
8	22	Female	Soprano	8
9	21	Female	Soprano	8
10	25	Male	Baritone	6

balcony. The Smith Memorial room is a medium-sized venue modeled after a baroque drawing room with marble floors and columns, plaster walls, three crystal chandeliers, and a large rug covering most of the floor. It seats 50 audience members and is used for many smaller recitals and chamber performances. Room 204, a teaching studio, was selected as the small room to model the space where singers typically rehearse and have voice lessons.

Each room was replicated visually and acoustically to provide the singers with an immersive VR experience. These virtual replications of the rooms were presented using a VR head mounted display (HMD) worn by participants in a whisper room (interior dimensions, 226 cm $\times$ 287 cm, $h = 203$ cm). RT ($T30$) for mid frequencies (500–2000 Hz) was measured at 0.07 s, and ambient noise was measured at 25 dB(A). The VR equipment involved a Meta Quest 2 (Meta Platforms, Menlo Park, CA) VR HMD for the visual stimulus, paired with open-backed headphones (HD600, Sennheiser, Wedemark, Germany) for the audio stimulus. The visual graphics displayed to the singers were 360° photos of the rooms, taken from the perspective of the singer on stage and captured by an X3 camera (Insta 360, Shenzhen, Guangdong, China).

C. Room acoustic measurements

To characterize the acoustic properties of the rooms, extensive measurements were taken, following ISO standard.[21] Impulse responses (IRs) were recorded using an impulsive sound source (model BAS006, Larson Davis, Provo, UT) and analyzed with an XL2 audio and acoustic analyzer (NTi Audio, Tigard, OR). The microphone used was a calibrated M2211 from NTi Audio. The impulsive sound source was located in the position of a performer (i.e., at the front of each room in the center), and the receivers were located in different positions throughout the rooms, where a voice teacher/audience member would be present.

From the IRs, RT ($T30$), EDT, and clarity ($C80$) were calculated with Aurora,[22] which is a plug-in for Audacity software (Muse group, Limassol, Cyprus). $T30$ is a common measure of RT and defined as two times the amount of time it takes for the sound source to reduce by 30 dB, excluding the first 5 dB of decay. RT is calculated by octave band, however, when a single number is needed, the average between 500 and 1000 Hz is used.[21] EDT refers to the duration in which the sound level decreases by 10 dB sound pressure level (SPL) after the sound source has ceased to operate, mulpiplied by six.[21,23] $C80$ is defined as the ratio of sound energy present within the first 80 ms of an IR when early reflections occur to the sound energy present thereafter. $C80$ is expressed in decibels, in which high $C80$ values correspond to higher clarity of music.[21] These parameters were averaged over the 500 Hz and 1 kHz octave bands, according to the international standard ISO 3382-1.[21] Relevant acoustic characteristics of each room are outlined in Table II.

In this study, room acoustic parameters, such as $T30$, EDT, and $C80$, were reported to provide a basic characterization of the acoustic environments used for the recordings. These parameters were selected to offer readers essential information about reverberation and clarity without introducing complexity unrelated to the primary focus of the investigation. Unlike previous work, where the direct influence of room acoustics on vocal production was analyzed in depth,[8] the present study aimed to assess whether singers produce comparable vocal outcomes in real rooms and their virtual replicas.

To characterize the acoustic properties of the rooms from the perspective of the performers, oral-binaural impulse responses (OBRIRs) were measured. Within each room, the following factors influence a singer's autophonic perception:

(1) Head diffraction—the propagation of sound from a point in front of the singer's mouth around the head and to the ears;[24]
(2) the room's IR—the acoustic response of the room to a brief, high-intensity sound source;[25] and
(3) head-related transfer function (HRTF)—the filtering of sound at the entrance to the ear canal, depending on the angle of incidence.

To model the combination of these effects, OBRIRs were measured using a head and torso simulator (HATS).[26,27] A sine sweep signal was emitted from the HATS' mouth to simulate the singer's voice. The sound propagated through the room, underwent head diffraction, and was filtered by the HRTF before reaching the ears. The resulting binaural signals were captured by microphones located in the HATS' ears. The IRs were then computed, using Aurora, by convolving the recorded sweeps with the inverse of the emitted sweep.[28] From these IRs, the voice support parameter (STv) was calculated following the method proposed by Pelegrín-García.[27] The resulting STv values in the real rooms were as follows: -13.28 dB for the Smith Recital Hall, -10.69 dB for the Smith Memorial room, and -4.98 dB for room 204. These values provide an objective measure of the reflected-to-direct sound energy ratio received at the performer's ears in each environment.

D. Auralization and convolution

Auralization can be defined as the process of simulating room acoustics as a binaural listening experience at a given position in a room.[29] In the context of vocal performance,

TABLE II. RT ($T30$), EDT, and clarity ($C80$) were measured with an omnidirectional source located on the stage and a receiver in the audience. All parameters were averaged over the 500 Hz and 1 kHz octave bands. The average (minimum − maximum) values measured in the audience, Volume, number of seats, and number of measurement points are listed.

Room	Volume (m^3)	Seats	Points	$T30$ (s)	$C80$ (dB)	EDT (s)
Room 204	114.5	2	5	0.39 (0.26–0.81)	17.71 (12.72–22.4)	0.40 (0.17–0.57)
Smith Memorial room	400	56	9	1.04 (1.07–1.11)	2.00 (0.65–3.45)	0.98 (1.09–1.18)
Smith Recital Hall	6600	802	27	1.75 (1.56–1.87)	−0.83 (−2.56–2.87)	1.85 (1.72–2.15)

this requires accurately modeling how a singer hears their own voice within a performance venue.

The VR portion of this study was conducted in a whisper room, where participants performed vocal tasks while wearing open-backed headphones (HD600, Sennheiser, Wedemark, Germany) that reproduced the auralized acoustic environments in real time. A microphone (M2211, NTi Audio, Tigard, OR) positioned 30 cm in front of the singer's mouth was used to capture the voice signal, which was convolved with processed OBRIRs. The recorded OBRIRs were processed to filter out the systemic influences of the microphone, headphones, and HATS' ear canal.

To ensure the fidelity of the real-time auralization, these components were removed through inverse filtering, following established procedures.[30] Additionally, the direct sound component was removed from the final IRs as the use of open-backed headphones allowed participants to naturally perceive their own direct sound from mouth to ears without artificial mediation. This approach ensured that the OBRIRs represented only the reflected components of the acoustic environment, avoiding redundant or a distorted representation of the direct signal. As a result, the auralization accurately simulated the acoustic experience of each room while preserving the participant's natural autophonic feedback. The system gain was calibrated by measuring OBRIRs in the whisper room while the auralization system was active. The gain was adjusted to minimize discrepancies between the original and reproduced OBRIRs, with specific attention to maintaining the energy ratio between the direct path from the HATS' mouth and the reflected components.

In conclusion, the ORBIRs used in the auralization were obtained using the convolution formula

$$H_{\mathrm{room}}(t) = H_{1_{\mathrm{rev}}} * \mathrm{Inv}(H_{2_{\mathrm{dir}}}) * \mathrm{Inv}\,(H_3),$$

where $H_{\mathrm{room}}(t)$ is the final room IR, $H_{1_{\mathrm{rev}}}$ is the reverberant portion of the recorded IR, $\mathrm{Inv}(H_{2_{\mathrm{dir}}})$ is the inverse-filtered direct path from mouth-to-M2211 microphone, and $\mathrm{Inv}(H_3)$ is the inverse-filtered headphone-to-ear canal signal. The inverse filtering was implemented using Kirkeby filters.[30,31]

The derived ORBIRs were used in real-time auralization, implemented using convolution plug-ins in Reaper software (Cockos Inc., Rosendale, NY). The convolution signal was delivered through the open-backed headphones (HD600, Sennheiser, Wedemark, Germany). This procedure ensured that participants experienced a realistic simulation of the acoustic conditions of each room while maintaining their direct sound path.

Finally, the real-time auralization process can be expressed as

$$y_{\mathrm{rev}}(t) = x(t) * F_{1_{\mathrm{dir}}} + (x * H_{\mathrm{room}}(t)) * F_{2_{\mathrm{dir}}} * F_3,$$

where x is the participant's vocal signal, $F_{1_{\mathrm{dir}}}$ is the direct mouth-to-ear path, $F_{2_{\mathrm{dir}}}$ is the direct mouth-to-M2211 signal, and F_3 accounts for the headphone-to-ear canal transfer function. This convolution scheme preserves the direct sound and models the room response as filtered through the equipment and performer's auditory anatomy.

The system latency was first addressed in REAPER by enabling the "use audio driver reported latency" option in the audio preferences, which ensured automatic compensation. For greater accuracy, we followed standard loopback testing procedures to verify and fine-tune the latency compensation. Real-time convolution was performed using the ReaVerb plugin. To ensure minimal latency during auralization, we set the fast Fourier transform (FFT) size to 512 and enabled the "ZL" (zero latency) and "LL" (low latency threading) options. These settings were selected to ensure tight synchronization between the singers' live vocal output and the auralized feedback heard through open-back headphones. The measured and manually adjusted latency was also subtracted from the OBRIRs prior to convolution, ensuring that the timing of early reflections remained consistent with the real acoustic environments.

As part of this process, the STv was computed for the real and auralized conditions to quantify the reflected-to-direct energy ratio. The results confirmed a close match between real and simulated environments: −13.28 dB (real) and −13.22 dB (auralized) in the Smith Recital Hall, −10.69 dB (real) and −10.23 dB (auralized) in the Smith Memorial room, and −4.98 dB (real) and −4.72 dB (auralized) in room 204. These comparisons support the accuracy of the auralization procedure in reproducing the room-specific acoustic feedback experienced by performers.

E. Protocol

All ten singers gathered on one evening in October 2023 to perform live in each of the three rooms—first, in the Smith Recital Hall, followed by the Smith Memorial room, and, finally, room 204. All participants were music students at the University of Illinois Urbana-Champaign and familiar with the rooms used in the study through prior group activities and general exposure. However, none of the participants had previously performed solo or rehearsed individually in these rooms. No practice sessions or extended warm-ups were conducted in the rooms prior to the recordings, allowing the study to capture the singers' natural vocal production responses on entering each acoustic environment. Because the auralizations were based on a static IR without dynamic head tracking, participants were instructed to minimize head movements and maintain a steady head orientation during performances in all virtual conditions. They performed an excerpt of G. Giordani's "Caro mio ben" without accompaniment in their chosen key (Eb major for the five sopranos and two tenors, and C major for the three baritones). Two performances were given in each room, first, without the VR headset (real) and, then, a second time while wearing the VR headset (real_V), which displayed a 360-degree image of the same room. This step was taken to ensure that no significant vocal changes occurred simply because of wearing the headset. After two performances in each room, the participants were led to a quiet space to fill out a survey, rating their acoustic perceptions, previously used in Redman *et al*.[32] The survey consisted of 21 visual analog scales, each measuring a contrasting pair of descriptive attributes (e.g., "unsupported–supported," "dull–brilliant," and "quiet–noisy") on a 10 cm line. The questions addressed multiple aspects of the singing experience, including perceived vocal ease, acoustic support, clarity, reverberance, and the suitability of the space for unamplified performance.

For the second part of the study, singers performed the same excerpt under nine different virtual conditions and one additional control condition. The recordings were performed in a sound-attenuating double-walled whisper room. The nine conditions were three simulations of each of the three rooms: V_only, A_only, and comb_AV. For the comb_AV conditions, participants wore a VR headset (Meta Quest 2, Meta Platforms) displaying a 360-degree photo of the simulated room and open-backed headphones (HD600, Sennheiser, Wedemark, Germany), which simulated the acoustics of the same room. For the V_only conditions, participants wore the VR headset displaying the image of the simulated room and continued to wear the open-backed headphones but received no processed audio signal through the headphones. For the A_only conditions, participants received the acoustic feedback of the simulated room through the headphones and wore a blindfold for neutrality of visual input. As a baseline, a tenth control condition of no visual no-audio was recorded, in which the participants wore a blindfold and headphones but received no-audio filtering. The ten conditions were presented in a randomized order that was unique to each participant. After each condition, participants answered the same survey as in part one.

The survey and its analytical approach used in this study follow Redman *et al*.[32] Likewise, the acoustic parameters and subsequent analyses follow Bottalico *et al*.[8]

Acoustic recordings were captured using a calibrated measurement microphone (M2211, NTi Audio, Tigard, OR), which was calibrated prior to each session using a 1 kHz reference tone at 94 dB SPL. Following the indication of Svec and Granqvist,[33] the microphone was positioned 30 cm directly in front of the singer's mouth, which was consistent across all conditions. This placement ensured that the microphone remained within the critical distance of each room, favoring a high direct-to-reverberant energy ratio. This setup minimized the influence of room reflections on the recorded signal.

F. Statistical analysis

Statistical analyses of the objective voice parameters were conducted using linear mixed-effect models (LMMs). These analyses were performed with *R* software (version 3.6.0) and the lme4 package (version 1.1-10). Different models were built for the three response variables (vibrato extent, vibrato rate, and QR). The models were computed for each of the three voice parameters, where fixed effects (the sensory conditions) and random effects (rooms and the difference among singers) could be simultaneously taken into account. Tukey's *post hoc* pairwise comparisons with single-step correction were performed to examine the differences between all levels of the fixed factors of interest. (See the supplementary material for full pairwise comparisons among sensory conditions across all outcome measures.)

The analysis of the surveys was divided into two phases. In the first phase, the objective was to identify the set of significant affective impressions in the overall evaluation of the halls. A factor analysis (FA) was performed on the questionnaire. In this analysis, factors are represented as linear combinations of the original variables without inherent meanings. FA uncovers latent factors that explain observed correlations or covariances between variables. Assigning names to factors is a context-dependent, subjective process based on the examination of factor loadings.[34] The analisys was performed using the package psych (version 2.5.3). The second phase used LMMs to determine the relationship between sensory conditions and participants' subjective impressions.

III. RESULTS

A. Effect of sensory conditions on objective voice parameters

1. Effect of sensory condition on vibrato extent

The effect of sensory condition on vibrato extent was analyzed using a linear mixed-effects model, with sensory conditions as fixed effects, and participant identification (ID) and room as a random effect (Table III). Tukey *post hoc* comparisons were performed to explore specific differences between the sensory conditions.

The linear mixed-effects model examined the effect of sensory condition on vibrato extent while accounting for variability across participants and rooms. The results show a significant effect of sensory condition. Specifically, vibrato extent was significantly greater in the comb_AV, A_only, and V_only conditions compared to the real condition, where *p*-values were well below 0.001. The control condition also showed a trend toward increased vibrato extent relative to the

TABLE III. Linear Mixed-Effects (LME) model fit by Restricted Maximum Likelihood (REML) for the response variable *vibrato extent* with and condition as fixed factor. The estimate (β), standard error (SE), degrees of freedom (df), *t*-value, and *p*-value for each level are listed. Significance levels: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Condition	β	SE	df	*t*-value	*p*-value
(Intercept: real)	105.80	9.44	11.56	11.21	$< 0.001^{***}$
Real_V	0.02	2.73	1554.06	0.01	0.994
Combined_AV	29.68	2.65	1554.37	11.20	$< 0.001^{***}$
A_only	27.14	2.69	1554.55	10.10	$< 0.001^{***}$
V_only	34.76	2.72	1555.00	12.76	$< 0.001^{***}$
Control	23.61	9.05	2.73	2.61	$0.088^{\cdot}$

real condition ($p = 0.088$). Importantly, no significant difference was observed between the real condition and the real V condition (i.e., the real room with the VR headset worn but no virtual input), suggesting that wearing the headset did not measurably influence vocal production in this sample. The structure of the random effects revealed that variability attributable to individual singers was substantially higher than that associated with room differences, which likely reflects the limited number of rooms and participants as well as the diversity of voice types in the sample. Therefore, this result should not be interpreted as evidence that room acoustics exert a negligible influence but rather as an indication that inter-individual differences dominated within the scope of the present dataset. Consistent with the study's aim to assess overall effects of sensory condition rather than specific room characteristics, room was modeled as a random effect to account for variability across acoustic environments. The Shapiro-Wilk test was conducted to assess the normality of the residuals from the linear mixed-effects model. The results indicated no significant deviation from normality ($W = 0.998$, $p = 0.060$), suggesting that the residuals are approximately normally distributed. The DHARMa nonparametric dispersion test[35] also revealed no evidence of heteroscedasticity (dispersion $= 1.0047$, $p = 0.88$), supporting the assumption of homoscedasticity. Together, these results confirm that the model adequately met the assumptions underlying linear mixed-effects modeling.

Post hoc Tukey comparisons confirmed these findings, showing significant increases in vibrato extent in all virtual conditions (except real_V) relative to the real condition. Whereas the differences between comb_AV, A_only, and V_only conditions were not statistically significant, V_only produced significantly larger vibrato extent than A_only ($p = 0.042$), suggesting a possible influence of visual context on vocal modulation. This increase in the V_only condition is consistent with the absence of auditory feedback in that setting, as singers may compensate for the acoustically dead environment by increasing pitch modulation to enhance perceived vocal presence. Differences involving the control condition were not statistically meaningful. Overall, these results indicate that sensory manipulations—especially the absence or alteration of typical auditory and visual feedback—significantly influence vibrato extent, although individual variability remains substantial. *Post hoc* Tukey comparisons are listed in Table I of the supplementary material.

Figure 1 shows the mean vibrato extent (in cents) across different sensory conditions, where the black error bars represent the 95% confidence intervals. The *x* axis represents the six sensory conditions, whereas the *y* axis shows the mean vibrato extent (in cents). Each point on the line indicates the mean vibrato extent for each condition, where error bars capture the variability in the data. Additionally, the gray error bars represent the just noticeable difference (JND) for vibrato extent,[36] illustrating which conditions produced vibrato values perceptually close to the real condition. Figure 1 shows that there is not a noticeable difference in the mean vibrato extent between the real condition and other conditions, meaning a listener may not be able to perceive the difference in the performances with regard to vibrato extent.

2. Effect of sensory condition on vibrato rate

The effect of sensory condition on vibrato rate (Hz) was analyzed using a linear mixed-effects model, with sensory condition as a fixed effect and random intercepts for participant and room conditions (Table IV). Tukey *post hoc* comparisons were conducted to investigate specific differences between sensory conditions. Compared to the real condition, vibrato rate differed significantly in all sensory conditions except real_V, indicating that vibrato rate was generally sensitive to sensory context. The standard deviation attributable to participants ($SD = 0.27$) was relatively large compared to the standard deviation attributable to room ($SD = 0.06$), and both were smaller than the residual error ($SD = 0.38$). This suggests that although individual differences among singers contributed substantially to variability in vibrato rate, the variability across rooms was minimal. Therefore, room acoustics did not introduce major additional variance beyond the individual and sensory condition effects. The model residuals indicated no significant deviation from normality ($W = 0.995$, $p = 0.054$; dispersion $= 1.0056$, $p = 0.89$), suggesting that the model adequately met the assumptions underlying linear mixed-effects modeling.

Post hoc comparisons revealed that vibrato rate was significantly slower in all virtual conditions compared to the real condition. Specifically, the comb_AV, A_only, and

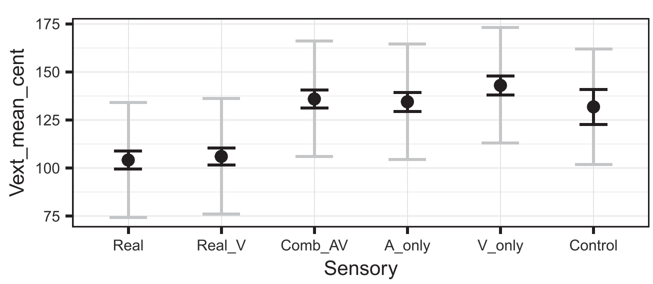

FIG. 1. Plot showing the mean vibrato extent [Vext-mean (cents)] across six sensory conditions. Error bars indicate 95% confidence intervals. The gray error bars represent the just noticeable difference (JND) threshold of ± 30 cents, showing that the different sensory conditions produced vibrato extents that are not perceptibly different to a listener.

TABLE IV. Linear Mixed-Effects (LME) model fit by Restricted Maximum Likelihood (REML) for the response variable vibrato rate (Hz) and condition as fixed factor. The estimate (β), standard error (SE), degrees of freedom (df), t-value, and p-value for each condition are listed. Significance levels: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Condition	β	SE	df	t-value	p-value
(Intercept: real)	5.74	0.094	11.60	60.85	< 0.001***
Real_V	0.03	0.032	1554.08	0.90	0.367
Combined_AV	−0.28	0.031	1554.59	−9.13	< 0.001***
A_only	−0.31	0.031	1554.72	−10.00	< 0.001***
V_only	−0.31	0.032	1555.27	−9.78	< 0.001***
Control	−0.30	0.078	3.65	−3.86	0.022*

V_only conditions all resulted in significantly reduced vibrato rates compared to the real condition (all $p < 0.001$), with differences of approximately −0.28 to −0.31 Hz. The control condition also showed a significantly slower vibrato rate compared to the real condition ($\beta = -0.30$, $p = 0.0014$). No significant differences were observed between the real and real V (real with virtual visual) conditions ($p = 0.94$), indicating that no measurable effect of the headset was detected within this sample, although such effects cannot be ruled out. Furthermore, pairwise contrasts among virtual sensory conditions (e.g., comb_AV versus A_only or V_only; A_only versus V_only) were all nonsignificant ($p > 0.9$), suggesting that any differences among these sensory manipulations were smaller than could be reliably detected with the present sample. Similarly, no significant differences were found between the control and any of the other virtual conditions. Overall, these results suggest that vibrato rate was consistently reduced in virtual environments relative to the fully real condition, but subtle differences among the virtual sensory modalities did not lead to differential effects. *Post hoc* Tukey comparisons are listend in Table II of the supplementary material.

Figure 2 shows the mean vibrato rate (Hz) across different sensory conditions, where error bars represent the 95% confidence intervals. The x axis represents the six sensory conditions, whereas the y axis shows the mean vibrato rate. Each point on the line indicates the mean vibrato extent for each condition, with error bars capturing the variability in the data. Additionally, a shaded gray interval highlights the vibrato rate values that fall within a 0.35 Hz range (±0.35 Hz) from the mean vibrato extent in the real condition. This shaded interval represents the JND for vibrato rate,[37,38] illustrating which conditions produced vibrato values perceptually close to the real condition. From Fig. 2, it is evident that the vibrato rate in all conditions is within the JND from the real condition.

3. Effect of sensory condition on QR

The effect of sensory condition on voice QR was analyzed using a linear mixed-effects model with sensory condition as a fixed effect, and participant ID and room as random intercepts (Table V). The model residuals indicated no significant deviation from normality ($W = 0.998$, $p = 0.061$; dispersion = 0.98702, $p = 0.86$), suggesting that the model adequately met the assumptions underlying linear mixed-effects modeling. The model revealed no statistically significant differences between the real condition and any of the other sensory conditions, although some effects approached significance. Specifically, the A_only condition ($\beta = 1.05$, $p = 0.060$) and the comb_AV condition ($\beta = 0.91$, $p = 0.10$) showed trends toward higher QR compared to the real condition. However, no significant differences were detected between the real with virtual visual (real V), V_only, or control conditions and the real condition ($p = 0.17$), although this may reflect limited power rather than true equivalence. The estimated variance caused by participant differences (SD = 2.40) was substantially higher than the variance attributable to room acoustics (SD = 0.27), and the residual variability was also large (SD = 6.75). This suggests that individual differences among singers accounted for a notable portion of the overall variability in QR, whereas differences among rooms contributed minimally. Overall, although trends suggest that some virtual conditions may slightly increase QR, particularly A_only and combined_AV, these effects did not reach conventional levels of statistical significance. *Post hoc* comparisons confirmed that the there was no statistical difference among sensory conditions (Table III of the supplementary material).

Figure 3 shows the mean QR (dB) across different sensory conditions, where error bars represent the 95% confidence intervals. The x axis represents the six sensory conditions, whereas the y axis shows the mean QR (dB).

TABLE V. Linear Mixed-Effects (LME) model fit by Restricted Maximum Likelihood (REML) for the response variable QR (dB) and condition as fixed factors condition. The estimate (β), standard error (SE), degrees of freedom (df), t-value, and p-value for each condition are listed. Significance levels: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Condition	β	SE	df	t-value	p-value
(Intercept: real)	15.88	0.87	13.57	18.20	< 0.001***
Real_V	0.77	0.57	1554.38	1.35	0.178
Combined_AV	0.91	0.55	1556.49	1.65	0.100
A_only	1.05	0.56	1556.80	1.88	0.060
V_only	0.18	0.57	1558.51	0.32	0.752
Control	1.05	0.86	21.69	1.23	0.232

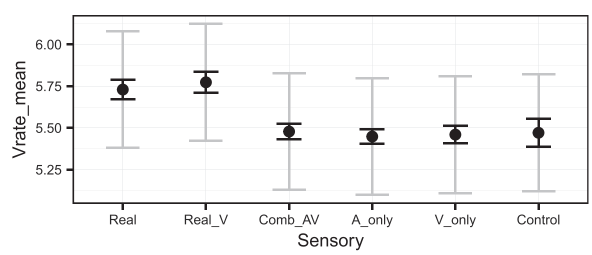

FIG. 2. Plot showing the mean vibrato rate [Vrate mean(Hz)] across six sensory conditions. Error bars indicate 95% confidence intervals. The gray error bars represent the JND threshold of ±0.35 Hz, showing that the different sensory conditions produced vibrato extents that are not perceptibly different to a listener.

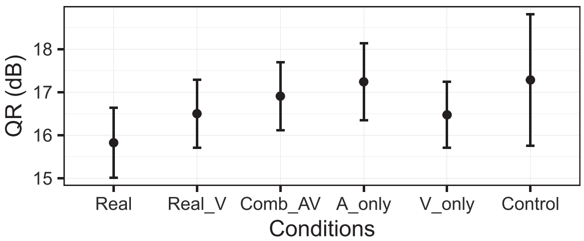

FIG. 3. Plot showing the mean QR (dB) across six sensory conditions. Error bars indicate 95% confidence intervals.

From Fig. 3, it is evident that the mean QR values appear to be consistent across sensory conditions with overlapping confidence intervals. Although slight variations in mean QR values are visible, these differences seem to be not robust enough to indicate meaningful changes.

B. Effect of sensory conditions on singers' perception

The singers' perception was analyzed following the methodology of Redman *et al.*[32] A FA was performed using the minimum residual (minres) method to minimize the residuals between the observed and model-implied correlation matrices. The solution was obtained using ordinary least squares (OLS) estimation. The cumulative variance explained by the first two extracted factors was 60% (Table VI). Factor scores were calculated using the regression method implemented in the fa( ) function from the psych package in *R*, which estimates participants' scores on each factor based on the factor loading structure while accounting for item covariances and uniqueness. The numerical values reported in Table VI represent the factor loadings, indicating the strength and direction of association between each perceptual item and the corresponding factor. The contribution of each original survey item to the two factors was analyzed to define the conceptual meaning of each dimension, supporting previous findings from Redman *et al.*[32] The following two factors were identified:

- **Factor 1**: *singing voice supportiveness*, representing singers' perception of voice support provided by the room along with their overall evaluation of the environment; and
- **factor 2**: *peacefulness and concentration in the venue*, capturing the perception of room quietness, reverberance, and how these factors affect voice concentration and clarity.

The effect of sensory condition on singing voice supportiveness was analyzed using a linear mixed-effects model with sensory condition as a fixed effect and participant ID and room as random intercepts (Table VII). The means and 95% confidence intervals for each condition are presented in Fig. 4. Tukey-adjusted *post hoc* comparisons were conducted to examine specific pairwise differences across sensory conditions (Table IV of the supplementary material). The fixed effect of sensory condition was not statistically significant overall. None of the individual conditions (comb_AV, A_only, V_only, or control) differed significantly from the real condition. For example, singing voice supportiveness scores in the control condition were slightly lower than those in the real condition (estimate = −0.20), but this difference did not reach statistical significance ($p = 0.33$). Similarly, all other comparisons showed nonsignificant differences (all $p = 0.62$). The random effects revealed moderate variability attributable to participants (SD = 0.38), whereas the variance associated with room was relatively small (SD = 0.08), indicating limited influence of room-level differences on the perceived supportiveness in this model. The model residuals indicated no significant deviation from normality ($W = 0.99379$, p-value = 0.1532; dispersion = 0.9723, $p = 0.92$), suggesting that the model adequately met the assumptions underlying linear mixed-effects modeling.

A linear mixed-effects model was used to assess the effect of sensory condition on peacefulness and concentration in the venue, with participant ID and room included as random effects. No significant differences were found between sensory conditions (all $p = 0.32$). *Post hoc* Tukey tests confirmed that none of the pairwise comparisons reached significance. Random effects indicated that

TABLE VI. FA of the perceptual survey items. Values represent the factor loadings for each item on the two extracted dimensions: MR1 (singing voice supportiveness) and MR2 (peacefulness and concentration). Factor loadings indicate the strength of association between each survey item and the corresponding factor.

Items (extremes)	MR1	MR2
Ease of singing (difficult-easy)	0.77	
Pleasantness of singing (unpleasant-pleasant)	0.79	
Reverberance while singing (dry-reverberant)	0.91	
Noise perception in the space when not singing (quiet-noisy)		−0.60
Peacefulness in the space when not singing (disruptive-peaceful)		0.62
Liveliness in the space when not singing (dull-live)	0.84	
Voice support (unsupported-supported)	0.63	
Voice brilliance (dull-brilliant)	0.90	
Voice fullness (thin-full)	0.82	
Voice focus (diffused-concentrated)		0.63
Voice weight (light-heavy)	0.82	
Voice power (weak-powerful)	0.88	
Voice gain (muted-amplified)	0.69	
Voice self-perception (difficult-easy)	0.85	
Loudness of own voice (weak-strong)		
Voice intonation (flat-sharp)		
Voice timbre (dark-bright)		
Voice clarity (muddy-clear)		0.62
Pleasantness of room feedback while singing (not at all-very much)	0.86	
Room size for singing (very small-very large)	0.80	
Pleasantness of non-amplified singing (not at all-very much)	0.74	
% **Variance explained**	49	11

https://doi.org/10.1121/10.0039807

TABLE VII. Linear Mixed-Effects (LME) model fit by Restricted Maximum Likelihood (REML) for the response variable singing voice supportiveness and the fixed factors condition and room. The estimate (β), standard error (SE), degrees of freedom (df), t-value, and p-value for each condition are listed. Significance levels: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Condition	β	SE	df	t-value	p-value
(Intercept: real)	0.300	0.254	8.43	1.181	0.270
Combined_AV	−0.053	0.193	114	−0.275	0.784
V_only	−0.927	0.193	114	−4.792	<0.001***
A_only	0.019	0.193	114	0.097	0.923
Control	−1.014	0.415	4.39	−2.447	0.065

variability was primarily attributable to individual differences (SD = 0.38), whereas room contributed minimally (SD = 0.08).

IV. DISCUSSION

The aim of the present study was to determine if vocal production and perception are affected when performing under various VR conditions. For each of the three real rooms, three different virtual sensory conditions were presented to the singers: A_only with no visual, V_only with no auralization, and comb_AV. Additionally, there was one control condition of no audio and no visual. Vibrato extent, vibrato rate, and QR were used to analyze the effect of sensory condition and room size on voice production. Singer perception of various acoustic environments was studied by completing a FA, which resulted in two latent factors: singing voice supportiveness and peacefulness and concentration in the venue.

Sensory condition significantly affected vibrato extent. Whereas virtual conditions differed from the real condition, not all contrasts exceeded the JND (∼30 cents). Thus, although statistical differences were detected, some changes may not be perceptible to listeners. Comparing the two real-room conditions (with and without the VR headset), no significant difference in vibrato extent was found. Whereas this result does not demonstrate equivalence, it suggests that any potential influence of wearing the headset was small relative to within-singer variability. Sensory condition also influenced vibrato rate, although the effects were modest. Average vibrato rates in the real (5.7 Hz) and comb_A (5.5 Hz) conditions fall within normative ranges reported in the literature,[39–42] situating these results within the broader context of singing voice research. As with vibrato extent, differences between real and virtual conditions did not exceed perceptual thresholds, suggesting limited perceptual relevance. No systematic difference was observed between performances with and without the VR headset, indicating that no measurable effect of headset use was detected; however, such differences cannot be excluded given the sample size.

QR did not differ significantly across sensory conditions. Descriptively, singers tended to produce slightly higher QR values in the largest, most reverberant room and when singing in virtual conditions without visual stimuli, but these trends showed substantial overlap and likely lack perceptual salience. Examination of random effects indicated that room and singer contributed variability, with singer-specific differences comparable to residual error. This suggests that performers adapted idiosyncratically to the experimental manipulations, which is consistent with prior reports of individual variability in responses to room acoustics.[3–6,17] Because room was modeled as a random factor, the present analysis cannot be used to draw inferences about specific room sizes or acoustics beyond noting that they contributed some unexplained variability.

The perception survey confirmed that acoustic feedback plays a central role in judgments of singing voice supportiveness. In particular, the control and V_only conditions, which lacked acoustic input, yielded significantly lower ratings. This finding underscores the importance of reflected sound energy for singers' sense of vocal ease and is consistent with previous work by Daşdöğen *et al.*,[19] who observed greater vocal effort and reduced comfort in comparable no-audio conditions. By contrast, ratings in the auralized virtual environments did not differ significantly from those in the real room. The absence of significant differences is consistent with the hypothesis that VR can approximate real acoustic feedback. However, this finding should not be interpreted as proof of perceptual equivalence as subtle differences may exist below the detection threshold of the present study. These results suggest that at the group level, simulated acoustics can provide a level of perceived support comparable to real environments.

Perceived peacefulness and concentration in the venue showed no significant effects of sensory condition. Here, too, the fixed-effect estimates were small relative to the between-singer variability, suggesting that this perceptual dimension may be less sensitive to experimental manipulations or singers differ substantially in how they interpret and apply this judgment. Response variability was notably higher in the control condition, indicating that singers were less consistent when deprived of acoustic and visual cues. This further supports the hypothesis that environmental feedback—whether real or virtual—stabilizes singers' perceptual evaluations.

In summary, VR conditions yielded small production differences but preserved singers' perceptual sense of support, highlighting the value of VR as a practical extension of real acoustic environments.

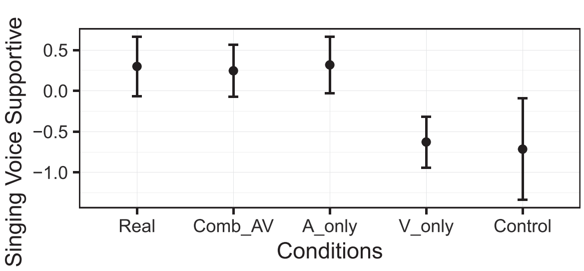

FIG. 4. Mean of the singing voice supportiveness across five sensory conditions. Error bars indicate 95% confidence intervals.

Several methodological and practical considerations should be acknowledged when interpreting the present findings. Although the order of the virtual conditions was randomized for each participant to reduce order effects, the real-room conditions necessarily followed a fixed order as a result of logistical constraints. This design choice may have introduced potential sequence effects that could not be fully controlled. Additionally, variability between the real and virtual performance sessions—such as differences in vocal dosage across conditions and the time of day at which participants performed—may have influenced vocal production outcomes. Participants may also have been affected by the lack of acoustic feedback in the treated sound booth, which differs substantially from typical performance environments. Furthermore, although participants did not report discomfort, motion sickness (i.e., cybersickness)[43] was not systematically assessed in this study. Future work should include standardized measures of VR-induced motion sickness to ensure participant comfort and better understand its potential impact on vocal behavior and performance outcomes. Finally, the relatively small sample size limits the statistical power of the study and may have reduced sensitivity to more subtle effects, underscoring the need for replication with larger participant groups.

Future work should build on these results by investigating a wider range of singer populations (e.g., amateur versus professional singers or different stylistic traditions) and examining long-term adaptation to virtual practice environments. Continued technological development and systematic validation could ultimately make high-quality VR simulations a valuable tool for training and providing access to performance spaces that are otherwise unavailable to singers.

V. CONCLUSIONS

This study examined the effects of sensory conditions and room simulations on the vocal production and acoustic perception of ten Western classical singers across three real rooms and their virtual replications. Vibrato extent and vibrato rate showed moderate influences of sensory condition, although the magnitude of some differences between real and virtual environments was within the range of JND, suggesting that they may not be perceptually salient to listeners. Importantly, no significant changes in vibrato extent, vibrato rate, or QR were observed when singers wore a VR headset in the real room. Whereas this finding cannot be taken as definitive evidence of equivalence, the absence of systematic shifts provides cautious support for the idea that headset use does not substantially disrupt vocal production. In line with prior work by Bottalico *et al.*[8] and Redman *et al.*,[32] the present findings reinforce the role of environmental context—visual and acoustic—in shaping singers' experiences of vocal production and perceived singing voice supportiveness. Notably, auralized virtual acoustics, particularly when combined with visual cues, elicited perceptions of vocal support comparable to those reported in real acoustic spaces. Although the present results provide encouraging evidence that virtual auralizations can approximate real performance conditions, the small sample size limits statistical power and generalization; future work, including larger and more diverse cohorts, is needed to validate these conclusions.

SUPPLEMENTARY MATERIAL

See the supplementary material for complete pairwise comparisons of sensory conditions across all acoustic and perceptual measures.

ACKNOWLEDGMENT

This work was made possible through the generous support of the Kate Neal Kinley Fellowship. We extend our sincere gratitude to the singers for their kind cooperation and participation in this study. Additionally, we appreciate the support provided by the University of Illinois Urbana-Champaign College of Fine and Applied Arts. We also thank the research assistants in the SpAA Laboratory, Asritha Tunuguntla and Ricardo Perez Guerrero, for their valuable contributions to data collection and curation. We are deeply saddened to acknowledge the passing of one of the authors of this paper, Professor Angelo Farina, who was an internationally recognized expert in auralization and room acoustics. His expertise and guidance were instrumental in designing and implementing the auralization procedures for this study. This work is dedicated to his memory and legacy in the field.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The use of human subjects for this research was approved by the Office for the Protection of Research Subjects at the University of Illinois Urbana-Champaign (IRB No. 24-0549).

DATA AVAILABILITY

Due to the data protection guidelines of the University of Illinois Urbana-Champaign, the raw audio recordings produced are available for scientific purposes only from the first author upon reasonabe request. Behavioral data are openly available in Singing-in-Virtual-versus-Real-Rooms-Is-It-the-Same- at https://github.com/SpAA-LAB/Singing-in-Virtual-versus-Real-Rooms-Is-It-the-Same-.

[1]V. L. Jordan, *Acoustical Design of Concert Halls and Theatres: A Personal Account* (Applied Science, London, UK, 1980), pp. 81–83.
[2]S. Ternström, "Long-time average spectrum characteristics of different choirs in different rooms," Voice (UK) **2**, 55–77 (1993).

[3]T. Fischinger, K. Frieler, and J. Louhivuori, "Influence of virtual room acoustics on choir singing," Psychomusicol.: Music, Mind, Brain **25**(3), 208–218 (2015).
[4]P. Luizard and N. H. Bernardoni, "Changes in the voice production of solo singers across concert halls," J. Acoust. Soc. Am. **148**(1), EL33–EL39 (2020).
[5]P. Luizard, J. Steffens, and S. Weinzierl, "Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions?," J. Acoust. Soc. Am. **147**(2), EL132–EL137 (2020).
[6]P. Luizard, N. H. Bernardoni, and C. Böhm, "Using virtual acoustics and electroglottography to study the adaptation of singing voice production," in *Forum Acusticum 2020*, Lyon, France (2020).
[7]A. C. Gade, "Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments," Acta Acust. Acust. **69**(5), 193–203 (1989).
[8]P. Bottalico, N. Łastowiecka, J. Glasner, and Y. Redman, "Singing in different performance spaces: The effect of room acoustics on vibrato and pitch inaccuracy," J. Acoust. Soc. Am. **151**(6), 4131–4139 (2022).
[9]B. Ackermann, "Managing the musculoskeletal health of musicians on tour," Med. Probl. Perform. Artists **17**(2), 63–67 (2002).
[10]D. T. Kenny, *The Psychology of Music Performance Anxiety* (Oxford University Press, Oxford, UK, 2011).
[11]J. Sundberg, *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb, IL, 1987), pp. 163–165.
[12]C. E. Seashore, "The natural history of the vibrato," Proc. Natl. Acad. Sci. U.S.A. **17**(12), 623–626 (1931).
[13]C. Manfredi, D. Barbagallo, G. Baracca, S. Orlandi, A. Bandini, and P. H. Dejonckere, "Automatic assessment of acoustic parameters of the singing voice: Application to professional Western operatic and jazz singers," J. Voice **29**(4), 517.e1–517.e9 (2015).
[14]Z. Schärer Kalkandjiev and S. Weinzierl, "The influence of room acoustics on solo music performance: An empirical case study," Acta Acust. Acust. **99**(3), 433–441 (2013).
[15]Z. Schärer Kalkandjiev and S. Weinzierl, "The influence of room acoustics on solo music performance: An experimental study," Psychomusicol.: Music, Mind, Brain **25**(3), 195–207 (2015).
[16]J. Brereton, D. Murphy, and D. Howard, "A loudspeaker-based room acoustics simulation for real-time musical performance," in *Audio Engineering Society Conference: UK 25th Conference: Spatial Audio in Today's 3D World*, March 25–27, York, UK [Audio Engineering Society (AES), New York, 2012.
[17]K. Ueno, K. Kato, and K. Kawai, "Effect of room acoustics on musicians' performance. Part I: Experimental investigation with a conceptual model," Acta Acust. Acust. **96**(3), 505–515 (2010).
[18]K. Kato, K. Ueno, and K. Kawai, "Effect of room acoustics on musicians' performance. Part II: Audio analysis of the variations in performed sound signals," Acta Acust. Acust. **101**(4), 743–759 (2015).
[19]Ü. Daşdöğen, S. N. Awan, P. Bottalico, A. Iglesias, N. Getchell, and K. V. Abbott, "The influence of multisensory input on voice perception and production using immersive virtual reality," J. Voice (published online 2023).
[20]C. J. Nudelman and P. Bottalico, "Investigating the impact of visual input on voice production in virtual reality," J. Voice (published online 2023).
[21]ISO 3382-1:2008, "Acoustics—Measurement of room acoustic parameters. Part 1: Performance spaces" (International Organization for Standardization, Geneva, Switzerland, 2008).
[22]S. Campanini and A. Farina, "A new Audacity feature: Room objective acustical parameters calculation module," in *Proceedings of the Linux Audio Conference 2009*, Parma, Italy, (April 16–19, 2009).
[23]L. M. Rantala and E. Sala, "Effects of classroom acoustics on teachers' voices," Build. Acoust. **22**(3-4), 243–258 (2015).
[24]D. Pelegrín-García, E. De Sena, T. van Waterschoot, M. Rychtáriková, and C. Glorieux, "Localization of a virtual wall by means of active echolocation by untrained sighted persons," Appl. Acoust. **139**, 82–92 (2018).
[25]S. Rosen and P. Howell, *Signals and Systems for Speech and Hearing* (Brill, Leiden, Netherlands, 2011), Vol. 29.
[26]M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**(1), 214–222 (1975).
[27]D. Pelegrín-García, "Comment on 'Increase in voice level and speaker comfort in lecture rooms' [J. Acoust. Soc. Am. **125**, 2072–2082 (2009)] (L)," J. Acoust. Soc. Am. **129**(3), 1161–1164 (2011).
[28]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, February 19–22, Paris, France [Audio Engineering Society (AES), New York, 2000.
[29]M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," J. Audio Eng. Soc. **41**(11), 861–875 (1993).
[30]A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, May 5–8, Vienna, Austria [Audio Engineering Society (AES), New York, 2007].
[31]O. Kirkeby, F. Orduna, P. A. Nelson, and H. Hameda, "Inverse filtering in sound reproduction," Meas. Control **26**(9), 261–266 (1993).
[32]Y. G. Redman, J. D. Glasner, D. D'Orazio, and P. Bottalico, "Singing in different performance spaces: The effect of room acoustics on singers' perception," J. Acoust. Soc. Am. **154**(4), 2256–2264 (2023).
[33]J. G. Svec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," Am. J. Speech-Lang. Pathol. **19**(4), 356–368 (2010).
[34]S. Amengual Gari, M. Kob, and T. Lokki, "Analysis of trumpet performance adjustments due to room acoustics," in *Proceedings of the International Symposium on Room Acoustics*, Amsterdam, Netherlands (September 15–17, 2019), pp. 65–73.
[35]F. Hartig and M. F. Hartig, "Package 'dharma,'" R package **531**, 532 (2017).
[36]R. Wooding and J. Nix, "Perception of non-vibrato sung tones: A pilot study," J. Voice **30**(6), 762.E15–762.E21 (2016).
[37]H. Järveläinen, "Perception-based control of vibrato parameters in string instrument synthesis," in *International Computer Music Conference*, Gothenburg, Sweden (September 16–21, 2002).
[38]J. D. Glasner and A. M. Johnson, "Effects of historical recording technology on vibrato in modern-day opera singers," J. Voice **36**(4), 464–478 (2022).
[39]J. Nix, "Shaken, not stirred: Practical ideas for addressing vibrato and nonvibrato singing in the studio and the choral rehearsal," J. Sing. **70**(4), 411–418 (2014).
[40]I. Titze, *Principles of Voice Production* (Prentice Hall, Iowa City, IA, 1994).
[41]J. Sundberg, "Acoustic and psychoacoustic aspects of vocal vibrato," Q. Prog. Rep. **35**(2–3), 45–68 (1994).
[42]J. Hakes, T. Shipp, and E. Doherty, "Acoustic properties of straight tone, vibrato, trill, and trillo," J. Voice **1**(2), 148–156 (1987).
[43]S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: A review," Front. Psychol. **10**, 158 (2019).