Perceptive and objective evaluation of calibrated room acoustic simulation auralizations

Barteld N. J. Postma and Brian F. G. Katz

Citation: The Journal of the Acoustical Society of America **140**, 4326 (2016); doi: 10.1121/1.4971422
View online: https://doi.org/10.1121/1.4971422
View Table of Contents: https://asa.scitation.org/toc/jas/140/6
Published by the Acoustical Society of America

Perceptive and objective evaluation of calibrated room acoustic simulation auralizations

Barteld N. J. Postma[a)] and Brian F. G. Katz
Audio Acoustics Group, LIMSI, CNRS, Universite Paris-Saclay, 91405 Orsay, France

(Received 30 June 2016; revised 31 October 2016; accepted 22 November 2016; published online 19 December 2016)

Recently, auralizations have become more prevalent in architectural acoustics and virtual reality. However, there have been few studies examining the perceptual quality achievable by room acoustic simulations and auralizations. Such studies have highlighted potential problems in creating perceptually equivalent simulations when compared to measured auralizations in terms of parameter estimation. In order to accomplish realistic auralizations, calibration of the geometrical acoustics model can be considered a necessary step. In situations where the studied space exists, well-calibrated auralizations can be employed for multiple purposes, such as multi-modal virtual reality explorations, studies of the acoustical influence of renovations, and historic research. Using this case type as a base, a perceptual study evaluating state-of-the-art binaural auralizations has been carried out. Three test sites of different complexity and acoustics were selected: the abbey church Saint-Germain-des-Prés, the cathedral Notre-Dame de Paris, and the Théâtre de l'Athénée. Models were calibrated according to omni-directional source-receiver measurements for reverberation and clarity parameters. In the subjective listening test, measured and simulated binaural auralizations were compared according to eight acoustic perceptual attributes. Results showed that the methodical calibration procedure employed in combination with attention to control factors led to ecologically/perceptually valid auralizations.
[http://dx.doi.org/10.1121/1.4971422]

[NX] Pages: 4326–4337

I. INTRODUCTION

Recently, auralizations have become more prevalent in architectural acoustics and virtual reality. Auralization is a term introduced as an analogy to visualization to describe rendering audible (imaginary) sound fields.[1] Room acoustic auralization is a technique based on capturing or modeling spatial room impulse responses (RIRs) and convolving this with anechoic music, speech, or other suitable signals. Geometrical acoustics (GA) software are often employed to numerically compute the RIR of complicated geometries. Wave-based methods are computationally intensive, requiring complex geometrical models and complex input data[2] and have been shown to have limitations in the accurate simulation of late reverberation in complex spaces.[3]

As with any scientific simulation, it is important to have confidence in the quality of the GA models and the resulting auralizations. This can often be ensured through the use of calibration and validation procedures. Calibration of GA models, with the goal of achieving realistic auralizations, can be performed on the basis of room-acoustical parameter comparisons between measurement and simulation. Previous studies have evaluated the reliability of GA software to predict various measured room acoustical parameters.[3–7] However, there are numerous perceptual aspects associated with room acoustics that are not well captured by standard room acoustical parameters and may result in significant differences in the perceived acoustic space.

There are multiple incentives to obtain perceptually valid simulated auralizations. Calibrated models can be employed to generate a set of RIRs for use in a virtual reality simulation allowing the user to move around in the space with a reasonable degree of confidence in the validity of the acoustic simulation. Subsequent modifications can be made to the GA model in an effort to estimate the impact of modifications, with the premise that the calibrated GA model is a validated reference point. For example, in the context of proposed renovations, these virtual reality simulations could be used to investigate and demonstrate the acoustical implications of modifications to the architectural and acoustical properties of the space or to examine potential PA (public address and alarm) systems, including the effects of speaker position or type. In the context of historical studies, modifications can be made to the GA model to revert rooms to previous states in their history, such as prior to important renovations.[8,9] In addition, as acousticians become more familiar with calibrating models for a given GA software it can be hoped that subsequent model simulations for new constructions will also be more reliable.

To carry out auralization comparisons, besides ensuring the correct material property attributes, one needs to ascertain that other elements of the simulation are configured comparable to the measurement as well. Specific attention is made here with regards to binaural auralizations which provide the most natural reconstruction of the listening experience. The use of a different source or receiver directivity or binaural HRTF can lead to coloration and other significant perceptual differences. On the other hand, preparation steps

a)Electronic mail: bart.postma@limsi.fr

are required for the measured binaural room impulse response (BRIR) as well. Compensation must be made for the frequency response characteristics of the measurement system and for differences in signal-to-noise ratio (SNR) between measurement and simulation as the simulated BRIR is free of background-noise.

As the focus of this study concerns auralizations, previous studies which have compared measured to simulated auralizations are discussed here. Lokki and co-workers[10,11] compared real-head binaural recordings to simulated auralizations of a lecture room using subjective listening tests. The main aim was to validate auralizations of the DIVA system. Simulations were performed with the DIVA real-time auralization system. The radiation characteristics of the sound source were measured and radiation filters were designed to fit the measured frequency response. The applied HRTFs were measured from the same person who did the real-head recordings. Recordings were also made to capture the background noise which was then added to the simulated auralizations. Compensation for the frequency characteristics of the recording system was not carried out. Recordings and simulated auralizations were compared according to the following perceptual attributes: *source location*, *externalisation*, *sense of space*, and *timbre*. It was concluded that it was possible to create natural sounding auralizations with physics-based room acoustic modeling and advanced digital signal processing. However, some differences between recordings and auralizations were found, especially with a transient-like stimulus signal.

Choi and Fricke[12] attempted to validate computer models of two concert halls by comparing recordings to simulated auralizations. The main aim was to determine whether computer generated auralizations were judged in the same way as recorded music in actual halls. BRIR measurements were carried out and anechoic stimuli were played and recorded in the considered rooms. In parallel, GA models of the spaces were created using ODEON. Measured and simulated BRIRs for five source-receiver combinations per hall showed several differences which exceeded the just noticeable difference (JND) for room-acoustical parameters. Particularly, estimation of C80 seemed problematic with 7 out of 10 source-receiver combinations overestimating it by more than 1 JND [$\text{JND}_{C80} = 1$ dB (Ref. 13)] and in some cases by more than 4 JND. In subsequent listening tests, recordings were compared to simulated auralizations where HRTFs differed between recording and simulation. Compensations for the frequency characteristics of the recording system were carried out, however, no compensation was performed for differences in SNR. Subjective listening tests indicated that there was a significant difference in *preference* between measured and simulated auralizations, indicating that these differed perceptually.

Yang and Hodgson[14] attempted to validate computed auralizations for use in speech-intelligibility studies. For this purpose, two classrooms (one with "extensive acoustical treatment" and the other with "no acoustical treatment") were measured and simulated using the GA software CATT-ACOUSTIC. Simulated and measured RIRs showed several differences exceeding JNDs for parameter results. C50 was underestimated for one classroom and overestimated for the other by more than 1 JND for five of the six source-receiver combinations, with a maximum difference of 5.6 JND. Subsequently, speech-intelligibility tests were carried out in the classrooms in background noise free conditions and with added background noise. In parallel, the same speech-intelligibility tests were performed in a sound isolation booth using the simulated auralizations in comparable background noise conditions. No compensation was performed for the frequency characteristics of the measurement system. The HRTF differed between the classrooms and simulated auralization tests. In the noisy conditions the SNR was equal, however, no compensation was performed in the noise-free conditions. The simulated auralizations test and the test in the classrooms gave comparable results for the "extensively acoustically treated" room in background noise free conditions. However, in the case of the "no acoustical treatment" classroom in the noise free condition and both cases with added background noise, results did not agree between actual classroom and simulated conditions.

Table I provides an overview of which factors were controlled between measurement and simulation in these different studies. Even though the main objectives of these previous studies differed from the current study, their results highlight potential problems in creating GA models to obtain perceptually valid auralizations. First, several control factors differed between measurement and simulation. Second, objective and subjective differences were identified between the measurement and simulation, clarity parameters seem especially problematic. To explore the second issue, a prior study examined the ability to objectively calibrate a GA model to measured data of a real space, resulting in a methodical calibration procedure for GA models.[15] This calibration procedure was validated by means of objective parameter comparisons for omnidirectional source and receivers. It is therefore important to examine how other perceptual parameters, especially spatial parameters, vary between measured and simulated binaural room auralizations.

The current study proposes to compare the perceptual quality of binaural auralizations for several rooms and source-receiver combinations relative to measured data using a state-of-the-art simulation software and a methodical calibration procedure. The proposed study is designed to examine the following hypotheses.

TABLE I. Control factors between measurement and simulation in previous auralization comparison studies. A ✓ indicates that the factor was controlled (i.e., equal) between measurement/recording and simulation, a × indicates a difference. † indicates that control factor was equal for the noisy conditions, however differed for the noise free conditions.

Factor	Lokki and co-workers (Refs. 10, 11)	Choi and Fricke (Ref. 12)	Yang and Hodgson (Ref. 14)
Source directivity	✓	✓	✓
Meas. system's freq. response	×	✓	×
SNR	✓	×	†
HRTFs	✓	×	×

H1: GA model calibration according to a set of omni-directional source-receiver combinations, bringing clarity and reverberation parameters within 1 tolerance threshold (ϵ), results in valid binaural auralizations.

H2: GA model calibration for an ensemble of discrete points (source and receiver positions) provides sufficient confidence in the quality of the simulated RIRs at other positions.

To accomplish this study, the GA models of three test rooms were created and calibrated, described in Sec. II. Subsequently, the measured and simulated auralizations were compared using a subjective listening test. Section III presents the protocol of the listening test and Sec. IV describes the results. Section V discusses the hypotheses and compares the results to previous studies.

II. CREATION AND CALIBRATION OF THE GA MODELS

Binaural objective parameter results and auralizations were evaluated for the Notre-Dame de Paris cathedral and the Saint-Germain-des-Prés abbey church. In a room-acoustical sense, these sites are comparable, as they have generally well-distributed absorption and reverberation times which are not strongly dependent on scattering. To provide an acoustic contrast, auralizations were also studied for the Théâtre de l'Athénée, a small horseshoe style theatre built at the end of the 19th century. The GA model creation and calibration procedure was carried out for the three rooms. An overview of the seven-step procedure is summarized here, based on Postma and Katz.[15]

(1) RIR measurements are carried out in the studied venue. The results of these measurements are used as a reference for the calibration.
(2) The geometrical model is created and the geometry remains unchanged during calibration.
(3) Preliminary (but realistic) acoustical properties are assigned to all surfaces, resulting in a GA model.
(4) Since stochastic implementations of Lambert scattering in GA software leads to run-to-run variations, the GA model's repeatability is quantified. These variations are then taken into account when simulations and measurements are compared.
(5) The sensitivity of the GA model to variations of scattering coefficients is quantified. For this purpose, simulations are run of the initial GA model followed by simulations with all scattering coefficients set to 0%, then to 99%, with absorption coefficients unchanged.
(6) Acoustical surface properties (absorption and scattering coefficients) are adjusted, taking into account the determined sensitivities, to arrive at global mean differences between measurement and simulated results for reverberation and clarity parameters within a tolerance threshold (ϵ).
(7) Acoustic properties of local key surfaces are adjusted to minimize the standard deviation (SD) of pairwise differences in reverberance and clarity parameters.

For simplicity and uniformity in the current study, the JNDs of the ISO 3382 standard[13] ($\mathrm{JND}_{EDT} = 5\%$, range $= 1.0 - 3.0$ s; $\mathrm{JND}_{C80} = 1$ dB, range $= \pm 5$ dB) were selected as calibration target thresholds (ϵ) for the different parameters. It is noted that these JND values vary as a function of room usage, but they are chosen as a base model tolerance reference value for the purpose of this study. In a specific instance, a more suitable threshold should be used which is specifically appropriate to the room's function.

As previous studies have reported significant variations in RIR analysis algorithm implementations,[16,17] it is important to use a single analysis tool for estimating the acoustic parameters when comparing results from measured and simulated RIRs. In addition, as it is the exported RIRs that are listened to, and not the energy decay responses calculated and analyzed within the software for parameter calculations, it is reasonable that it is to these RIRs that the model is calibrated.

A. Associated measurements

Following the first step of the calibration procedure, acoustic measurements were performed according to the following details.

- Measurement positions—Figures 1(a)–1(c) show the measurement plans, indicating source and receiver position which were chosen to represent typical usage. In addition to RIR measurements for calibration, BRIRs were also measured for the subsequent listening test; their positions are indicated by the numerated positions. For the numbered receivers in the Théâtre de l'Athénée and Saint-Germain-des-Prés church, no omni-directional microphones were positioned.
- Signal—The exponential swept sine method was employed. The sweep frequency went from 20 Hz to 20 kHz, duration of 10 s for the Théâtre de l'Athénée and Saint-Germain-des-Prés church, while a 20 s sweep was used for the Notre-Dame cathedral, due to its longer reverberation time and larger volume which requires greater SNR for the same source level.
- Sound source—The stimuli signal was sent to an amplifier (Servo 120 a, SAMSOM) and sequentially to miniature ($ø = 0.12$ m) dodecahedral omni-directional sound sources (model 3D-032, Dr-Three; variation in directivity within ± 3 dB up to 4 kHz, evaluated per 1/3rd octave band).
- Microphones—Four omni-directional microphones (model 4006, DPA) were employed to cover audience areas in the Théâtre de l'Athénée and Saint-Germain-des-Prés church. The input signal measurement chains of the Notre-Dame cathedral are described in more detail in Postma and Katz.[9] For the purposes of the perceptual test, a dummy head (Neuman KU 80, equipped with model 4060, DPA) was included. The dummy head orientation in the Notre-Dame cathedral and Saint-Germain-des-Prés church was towards **S2** and in the Théâtre de l'Athénée towards the center of the stage.
- Post-processing—Post measurement deconvolutions were carried out using MATLAB. The resulting RIRs were analyzed using the in-house MATLAB impulse response analysis

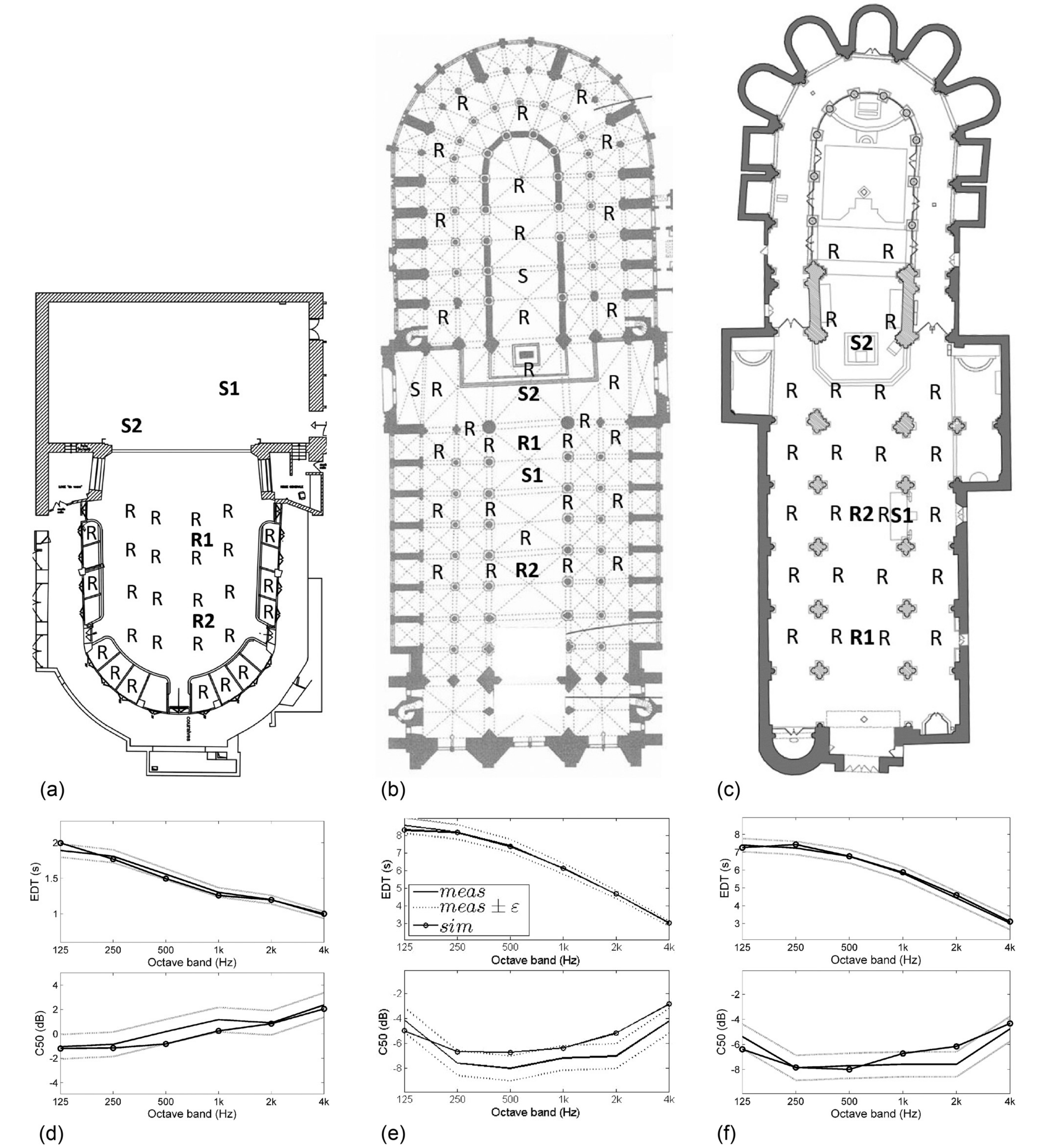

FIG. 1. Measurement plans of (a) Théâtre de l'Athénée, (b) Notre-Dame cathedral, and (c) Saint-Germain-des-Prés church. **S** and **R** represent source and receiver positions (**S#** and **R#** were employed in the listening test). Comparison between simulated and measured mean ($\pm 1\epsilon$) EDT and C50 for (d) Théâtre de l'Athénée, (e) Notre-Dame cathedral, and (f) Saint-Germain-des-Prés church.

(IRA) toolkit of the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI). For the purpose of this study, six parameters were calculated according to the standard:[13] T20, EDT, C50, C80, IACC early ($IACC_e$), and IACC late ($IACC_l$).

As noted above, the binaural receiver positions in the Théâtre de l'Athénée and Saint-Germain-des-Prés church employed in the subjective listening test were different from the omni-directional receiver positions employed for the calibrations, contributing to the investigation of **H2**.

B. GA model calibration

CATT-ACOUSTIC (v.9.0.c:3, TUCT v1.1a:4) was employed to create the GA models and perform simulations. The geometry

TABLE II. Approximate details and dimensions of the geometrical models of the Théâtre de l'Athénée, Notre-Dame cathedral (ND), and Saint-Germain-des-Prés church (SGdP).

Room	Polygons	Volume (m^3)	Floor plan area (m^2)
Athénée	1300	2500	300
ND	14 700	84 000	4,800
SGdP	2200	22 200	1800

of the Théâtre de l'Athénée was determined from architectural plans and sections [see Fig. 1(a)]. The geometries of the Notre-Dame cathedral and Saint-Germain-des-Prés church were determined from 3D laser scan point clouds as well as architectural plans and sections [Figs. 1(b) and 1(c)]. Table II presents a summary of the geometrical model details.

Surface materials were determined from visual inspection. Initial absorption coefficients were adopted from publicly available databases.[18–20] Scattering coefficients were generally modeled using the CATT-ACOUSTIC option *estimate* which provides a simple estimation of this frequency dependent coefficient based on a given characteristic depth representative of the surface's roughness.

The choice of calculation algorithm depends on the geometry of the space. As the Notre-Dame cathedral and Saint-Germain-des-Prés church have fairly even absorption and reverberation times which are not strongly dependent on scattering, simulations were run with *Algorithm 1: Short calculation, basic auralization, and transition order 1 with 250 000 and 150 000 rays, respectively*. In contrast, absorption distribution in the Théâtre de l'Athénée is not uniform. As such, simulations were performed using *Algorithm 2: Longer calculation, detailed auralization with 100 000 rays*.

During step 6 of the calibration procedure, reverberation parameters T20 and EDT were calibrated by adjusting the absorption coefficients before adjusting scattering coefficients to calibrate clarity parameters C50 and C80. This step was performed with the baseline requirement to keep the material properties within physically viable values. To provide an overview of the difference between measurement and calibrated GA models, Figs. 1(d)–1(f) compare the mean measured EDT and C50 of all source-receiver combinations for the omni-directional microphones to those of the simulations. Simulated reverberation parameter EDT is within 1ϵ of the measured values across all frequency bands for the three rooms. The Théâtre de l'Athénée model estimated the C50 within 1ϵ of the measured value across all octave bands. The Notre-Dame model overestimated the C50 by slightly more than 1ϵ in the 500 and 2000–4000 Hz octave bands. The Saint-Germain-des-Prés model overestimated the C50 by slightly more than 1ϵ in the 2000 Hz octave band. T20 results were similar to EDT results and C80 results were comparable to C50 results. The slight overestimation of clarity parameters could not be corrected while both maintaining reverberation parameter calibration and keeping scattering properties within physically viable values.

In total, four source-receiver combinations were excluded from the average parameter analysis due to close source-to-receiver proximity. Calibration of these source-receiver combination was difficult due to the perfect omni-directional directivity of the simulated sound source while the measured sound sources have slight variations in directivity. As the receivers in the Théâtre de l'Athénée were positioned relatively far from the sources, no outliers for this room were observed. Receivers in the Notre-Dame cathedral were positioned in close proximity to **S2** and the source at the west-side of the altar. As a result, these two simulated source-receiver combinations exhibited variations relative to measured values EDT [single number frequency average (500–1000 Hz) ΔEDT $= -2.25$ s] and C50 [single number frequency average (500–1000 Hz) ΔC50 $= +7.6$ dB]. In the Saint-Germain-des-Prés church, the two source-receivers combination including **S2** and receivers closest to this source underestimated EDT [single number frequency average (500–1000 Hz) ΔEDT $= -0.68$ s] and overestimated C50 [single number frequency average (500–1000 Hz) ΔC50 $= +2.0$ dB] relative to measured results. As **S1** was positioned in the pulpit, receivers close to this source were (partially) shielded from the direct sound and included in the mean parameter analysis.

III. LISTENING TEST

With the conclusion of the omni-directional receiver calibration procedure for the selected acoustic parameters, the measured and simulated binaural receiver RIRs employed in the listening test were analyzed. The binaural GA-based simulation incorporated the previously measured HRTF of the same dummy head used during the measurement. Notre-Dame cathedral receiver position **S1R1** (Ref. 21) was omitted from the listening test as the dummy head was orientated away from the sound source. The following section describes the necessary post-processing of the measured RIRs for use in auralizations. The resulting measured and simulated BRIRs were convolved with three anechoic stimuli. Finally, an overview of the test protocol is presented, building upon a previous preliminary study.[22]

A. Preparation of the measured BRIR

Prior to commencing the subjective listening test, some additional processing was required for the measured BRIRs. First, the frequency response characteristics of the measurement system have been compensated for by creating an equalization filter. The measurement chain (one omni-directional microphone only) was installed in an anechoic room (IRCAM, Paris) and the IR of the omni-directional speaker was measured at 5° increments in the horizontal plane. The resulting IRs were time-windowed to 512 pt, to remove any reflection artifacts, from which the fast Fourier transform was calculated and the mean over all directions of the magnitude determined. A filter was generated to match the inverse of this response using the recursive filter design yulewalk method. Non-linear frequency weighting followed a bark scale approximation, constraining the filter's level of detail to follow human hearing sensitivity.

The results of a preliminary test[22] indicated that the measured binaural auralizations were judged as "brighter" than their simulated counterpart. No such difference was

TABLE III. Measured and simulated frequency averaged objective parameters (EDT, T20, C50, C80, $IACC_e$, $IACC_l$) per binaural receiver position. (Differences greater than 1ϵ are in boldface.)

			Reverberation EDT(s) [$\epsilon = 5\%$] (500–1000 Hz)			Clarity C50(dB) [$\epsilon = 1.0$] (500–1000 Hz)			Spatiousness $1–IACC_e$ [$\epsilon = 0.075$] (500–4000 Hz)		
	Position	Distance S-R (m)	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.	Meas.	Sim.	Diff.
Athénée	S1R1	7.4	1.68	1.31	**+0.37**	−0.3	−1.5	**+1.2**	0.520	0.506	+0.014
	S1R2	12.8	1.65	1.38	**+0.27**	−0.5	−0.7	+0.2	0.660	0.692	−0.032
	S2R1	6.7	1.37	1.33	+0.04	0.7	−1.5	**+2.2**	0.598	0.644	−0.046
	S2R2	11.3	1.35	1.29	+0.06	0.3	0.1	+0.2	0.740	0.715	+0.025
ND	S1R2	17.4	5.52	6.53	**−1.01**	−6.5	−6.5	−0.0	0.531	0.216	**+0.315**
	S2R1	7.9	6.76	6.59	+0.18	−2.4	−0.1	**−2.3**	0.214	0.159	+0.055
	S2R2	31.1	7.44	7.78	−0.34	−11.3	−10.9	−0.4	0.312	0.256	+0.056
SGdP	S1R1	14.7	5.65	5.56	+0.09	−7.1	−7.6	+0.5	0.649	0.688	−0.039
	S1R2	4.0	5.04	4.64	**+0.40**	−1.3	−0.4	−0.9	0.613	0.488	**+0.125**
	S2R1	33.7	7.11	7.23	−0.12	−11.6	−10.1	**−1.5**	0.537	0.269	**+0.268**
	S2R2	19.9	6.98	6.80	+0.18	−9.8	−8.0	**−1.8**	0.234	0.140	**+0.094**
			T20(s) [$\epsilon = 5\%$] (500–1000 Hz)			C80(dB) [$\epsilon = 1.0$] (500–1000 Hz)			$1–IACC_l$ [$\epsilon = 0.075$] (500–4000 Hz)		
Athénée	S1R1	7.4	1.42	1.38	+0.04	1.6	1.7	−0.1	0.878	0.686	**+0.192**
	S1R2	12.8	1.44	1.42	+0.02	1.9	1.9	+0.0	0.860	0.625	**+0.235**
	S2R1	6.7	1.38	1.45	−0.07	3.0	1.6	**+1.4**	0.858	0.666	**+0.192**
	S2R2	11.3	1.40	1.49	**−0.09**	3.0	2.5	+0.5	0.805	0.613	**+0.192**
ND	S1R2	17.4	6.51	6.82	−0.31	−5.3	−5.3	−0.0	0.894	0.644	**+0.250**
	S2R1	7.9	6.19	6.14	+0.05	−1.8	0.7	**−2.5**	0.928	0.658	**+0.270**
	S2R2	31.1	6.57	6.69	−0.12	−10.1	−9.1	−1.0	0.887	0.595	**+0.292**
SGdP	S1R1	14.7	6.44	6.35	+0.09	−5.4	−5.7	+0.3	0.911	0.692	**+0.219**
	S1R2	4.0	6.46	6.37	+0.09	0.0	1.4	**−1.4**	0.907	0.690	**+0.217**
	S2R1	33.7	6.68	6.55	+0.13	−9.5	−8.3	**−1.2**	0.896	0.642	**+0.272**
	S2R1	19.9	6.12	6.30	−0.18	−7.8	−6.0	**−1.8**	0.907	0.618	**+0.289**

observed for the omni-directional RIRs. This coloration was most likely due to the mean spectral response of the binaural microphone setup which was not equalized for the measured response, contrary to the HRTF measurement protocol, which eliminates the frequency responses of the speaker and microphone, used for integrating the HRTF in the simulated responses. Therefore, an additional equalization filter was applied to the measured BRIRs. Example BRIRs from the three test rooms were analyzed by comparing the spectral differences of the late part of the reverberation. The average of these spectral differences across the three rooms was used to create an inverse filter, as previously described, which was then applied to the measured BRIRs.

Subsequently, differences in SNR between frequency bands were compensated for. The RIR was decomposed into 1/3rd-octave band components (spanning 100–16 000 Hz). The noise floor was detected by determining the SNR for each 1/3rd-octave band. The signals were then windowed at the point 5 dB above the noise floor, eliminating the trailing noise. The decay rate (reverberation time) was then calculated over the entire window, and this decay rate was used to synthesize the continuation of a noise-free reverberant tail. This reverberant tail was created by multiplying a vector of random values by an exponential vector based on the decay rate, this synthesized RIR was filtered to the considered 1/3rd octave band. Since a reliable decay estimate reasonably requires at least 15 dB, 1/3rd-octave bands with $SNR < 20$ dB were discarded (muted in the final RIR). This typically resulted in omitting the lowest two 1/3rd-octave bands: 100 and 125 Hz (these were also omitted from the simulated RIRs). An equal power cross-fade between the measured and synthesized responses was applied over the last 10 dB decay of the measured and the first 10 dB decay of the synthesized response to provide a smooth transition between decay sessions, limiting audible artifacts.

No effort was made to recreate the correlation between left and right ear synthesized tails. Processing left the perceivable correlation between left and right channel unaffected as the absolute differences between pre- and post-processed single number frequency average $IACC_l$ (mean absolute difference: 0.004, SD: 0.003) were well within one JND [$JND_{IACC} = 0.075$ (Ref. 13)].

Table III provides a comparison of the measured and simulated single number frequency average objective parameters (EDT, T20, C50, C80, $IACC_e$, $IACC_l$). T20 between measurements and simulations agreed well for all positions except one in the Théâtre de l'Athénée, estimating the measurement within 1ϵ. The clarity parameters exhibited more occurrences of differences greater than 1ϵ. Position **S2R1** in the Notre-Dame Cathedral exhibited the maximum differences between measured and simulated C50 and C80, with 2.3ϵ and 2.5ϵ, respectively. The pews positioned in the nave

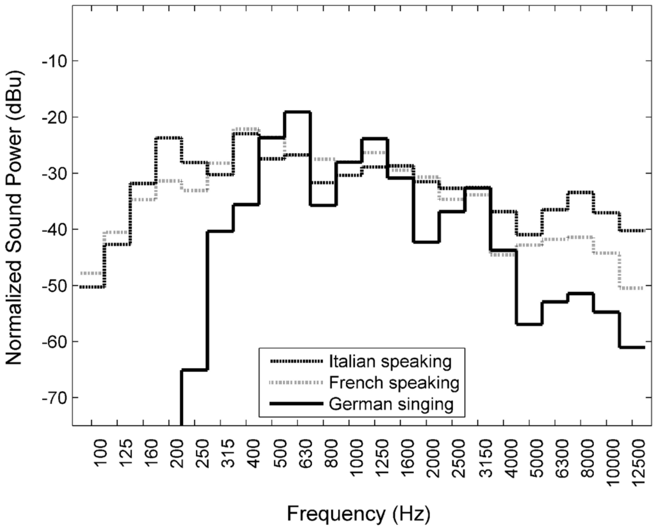

FIG. 2. 1/3-octave power spectrum for the anechoic test stimuli.

and first side aisles were modeled as floor with high scattering coefficients. In the cathedral, the pews have a height of ∼0.9 m and exhibit a regular and repeating geometry. This difference could have influenced the C50 and C80 for combinations with short source-receiver distances. Besides this maximum difference, more than half of the source-receiver combinations exhibited simulated clarity parameters within 1ϵ. Finally, differences between measured and simulated $IACC_e$ and $IACC_l$ exceeded 1ϵ for the majority of the positions, in some cases by more than 3ϵ. It should be repeated that the ϵ employed for IACC metrics was originally determined from concert hall conditions, and is therefore probably too strict for cathedral spaces. However, it should be appropriate for the drier theatrical space.

B. Stimuli

The resulting measured and simulated BRIRs were convolved with three anechoic audio extracts appropriate to the acoustic function of the different rooms. For the Théâtre de l'Athénée: a French speaking male reciting a translated extract of *Hamlet*, by W. Shakespeare, and an Italian speaking male reciting an extract of *Non recidere, forbice, quel volto* by E. Montale. These anechoic stimuli were recorded in the anechoic room (IRCAM, Paris) using an omnidirectional microphone (model 4006, DPA) at 4 m distance. For the Notre-Dame cathedral and Saint-Germain-des-Prés church: a female soprano singing *Abendempfindung* by W. A. Mozart, for details of the anechoic recording system see Ref. 23. Figure 2 depicts the spectral composition of the chosen extracts, each with a duration of ∼13 s. The root mean square (RMS) of the measured and simulated convolutions was used for level normalization.

An alternative normalization could have been to use A-weighted levels. As the auralizations employ the same anechoic recording and similar BRIRs these methods are roughly equivalent. This is confirmed by comparisons of the A-weighted levels of the normalized measured and simulated auralizations. For all auralization pairs, this difference was less than 1 dBA (mean absolute difference 0.48 dBA, SD: 0.27 dBA).

C. Listening test protocol

The resulting binaural auralizations were compared in a subjective listening test. This test was set up as an AB comparison. In order to study the influence of the selected anechoic sound sample, two stimuli were used for the Théâtre de l'Athénée, resulting in eight auralization pairs (2 stimuli × 4 positions: **S1R1**, **S1R2**, **S2R1**, **S2R2**). Only one stimuli was selected for the other rooms, resulting in three auralization pairs for the Notre-Dame cathedral (**S1R2**, **S2R1**, **S2R2**), and four for the Saint-Germain-des-Prés church (**S1R1**, **S1R2**, **S2R1**, **S2R2**), comprising 15 tested pairs. Additionally, four configurations were repeated to monitor the repeatability of perceptual responses. One pseudo pair (A ≡ B) was tested which had the explicit function of determining the reliability of participant responses. This resulted in a total of 20 pairs. Initially, participants were given three training pairs under supervision to ensure they understood the task. Results for these training pairs were not analyzed.

Participants were given written instructions before commencing the subjective listening test. They were not informed of the nature of the recordings (i.e., simulated or measured). They were asked to rate the similarity of samples according to eight perceptual attributes with the associated definitions.

- Reverberance (reverb)—The perception of the decay of sound. More reverberance is associated to a longer decay.
- Clarity—The degree to which discrete elements in the recorded musical performance stand apart from one another. If clarity is high, it is easy to spot individual notes in a musical piece, or individual phonemes in speech. If clarity is low, individual sounds merge, blend, or at the extreme can be confused and muddled.
- Distance (dist.)—The perceived distance to the sound source in the recording space.
- Tonal balance (ton. bal.)—Tonal balance represents changes in timbre, or frequency balance. It is qualified here as a comparison of the ratio of high to low frequency components so that more "tonal balance" indicates more high frequency content.
- Coloration (col.)—Coloration represents modifications in the timbre of the sound source from its original timbre. With less coloration, the recording sounds more natural.
- Plausibility (plaus.)—Given the assumption that the binaural recordings were made in a church (singing voice) or in a theatre (spoken voice): Does the recording sound reasonable to you?
- Apparent source width (ASW)—Apparent source width describes the perceived width of a sound image. The source may sound "narrow" (in the extreme case it is as if the sound is coming from a point). On the contrary, the source can also sound very "wide."
- Listener envelopment (LEV)—Listener envelopment describes the spatial distribution of the reverberant sound field. Higher listener envelopment means a more uniform

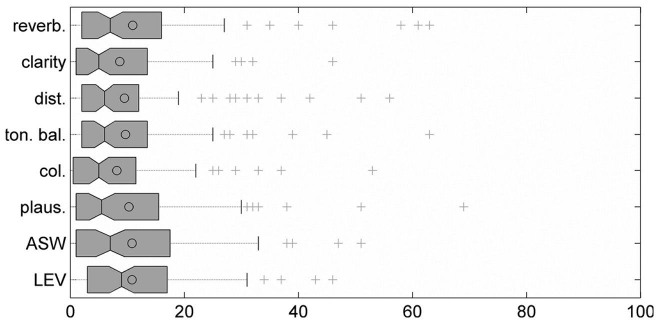

FIG. 3. Difference (magnitude) of subject parameter ratings for repeated pair conditions. Box limits represent 25% and 75% quartiles, (+) outliers, (○) median, and notch (|) mean values. (See Sec. III C for acoustical attribute acronym descriptions.)

distribution, while less listener envelopment means a more localized or directional reverberant sound.

Participants responded using a continuous graphic slider scale (100 pt), with the end parts labeled "A is much more…" and "B is much more…" corresponding to end values of –50 and +50, respectively, with a center 0 response indicating no perceived difference. Presentation order and AB correspondence to simulation and measurement were randomized. Participants were able to listen to the compared pairs as many times as desired. Auralizations were presented via headphones (Sennheiser model HD 600) at an RMS level of 75 dBA.

In contrast to the preliminary test,[22] where participants could only start listening at the beginning of the auralization, the test interface now allowed participants to be able to select the starting play point during listening. Additionally, the definitions for *coloration* and *tonal balance* had been revised.

27 test subjects participated, selected as having experience in either acoustics or vocal/instrumental performances. Before commencing the listening test, an audiogram was performed. One participant was immediately excluded due to poor hearing. Of the remaining 26 participants (mean age: 39.2 years, SD: 13.1) 21 were tested in an isolation booth located at LIMSI, ambient noise level <30 dBA, and 5 were tested in a silent office located at the Institut National d'Histoire de l'Art (THALIM) (ambient noise level ∼31 dBA).

IV. RESULTS

Initial attention was given to the reliability of the subjects, determined from the results of the pseudo pair. Analysis of the absolute data for this pair showed that one participant gave outlier responses (exceeding $\sim\pm 2.7\sigma$ and 99.3% coverage if the data are normally distributed) for all acoustics attributes indicating that the task was not well understood, and was therefore excluded from further analysis.

A. Perceptual repeatability tolerance

Analysis of the remaining 25 participants' results focused first on the repeatability of responses, determined from the absolute difference between the four repeated configurations (see Fig. 3). The mean difference between repetitions over attributes was 9.5 pt, with individual attribute repetition values being *reverberance* = 10.7, *clarity* = 8.4, *distance* = 9.1, *tonal balance* = 9.2, *coloration* = 7.8, *plausibility* = 9.7, *ASW* = 10.5, and *LEV* = 10.3. These values were used as tolerance ranges to judge whether a subjective acoustic attribute differed perceptually between measurement and simulation for the given test protocol and subjects.

B. Perceptual attributes

Combined results were compared [see "total" boxplot in Fig. 4(a)]. All perceptual attribute results were within the repeatability tolerance ranges established in Sec. IV A. Subsequently, results are analyzed for each individual space [see Fig. 4(a)]. No perceptual differences were observed for the attributes *reverberance*, *tonal balance*, *coloration*, *plausibility*, *ASW*, and *LEV*. However, the simulated auralizations of the Saint-Germain-des-Prés church were judged "clearer" and "closer" than the measured auralizations, while Théâtre de l'Athénée and Notre-Dame judgements were within the repetition tolerances for these attributes.

Results were further examined by room in detail. For all positions in the Théâtre de l'Athénée, averages were within tolerance ranges [see Fig. 4(b)]. Concerning the Notre-Dame cathedral, position **S2R1** results showed the simulated auralization was perceived as "clearer" and "closer" [see Fig. 4(c)], while all other equality judgements were within tolerance ranges. Saint-Germain-des-Prés results showed more differences; the simulated auralization of positions **S1R2** and **S2R2** were judged "closer" as well as **S1R2**, **S2R1**, and **S2R2** were considered "clearer" [see Fig. 4(d)], while all other equality judgements were within the tolerance ranges.

In comparison to the preliminary test,[22] the results of the Saint-Germain-des-Prés church have significantly improved for *tonal balance*,[24] due to the binaural dummy-head recording equalization. In contrast to the preliminary test results, simulated auralizations in the current test were found to be "clearer" and "closer." These differences could be due to the added ability of participants to start auralization playback at different points in the extract, allowing them to focus on different elements, instead of having to start listening at the beginning of the auralizations each time. The remaining attribute results were comparable to the preliminary test results.

C. Objective attributes

Subjective results were compared to the objective parameters presented in Table III. Between measurement and simulation, only a single T20 instance was found to slightly exceed 1ϵ, where all test results for *reverberance* were within the tolerance range. As auralizations with spoken voice were employed for the Théâtre de l'Athénée, C50 results are compared to the subjective attributes *clarity* and *distance*.[25] C50 results indicated that measured auralizations for positions **S1R1** and **S2R1** would be expected to be

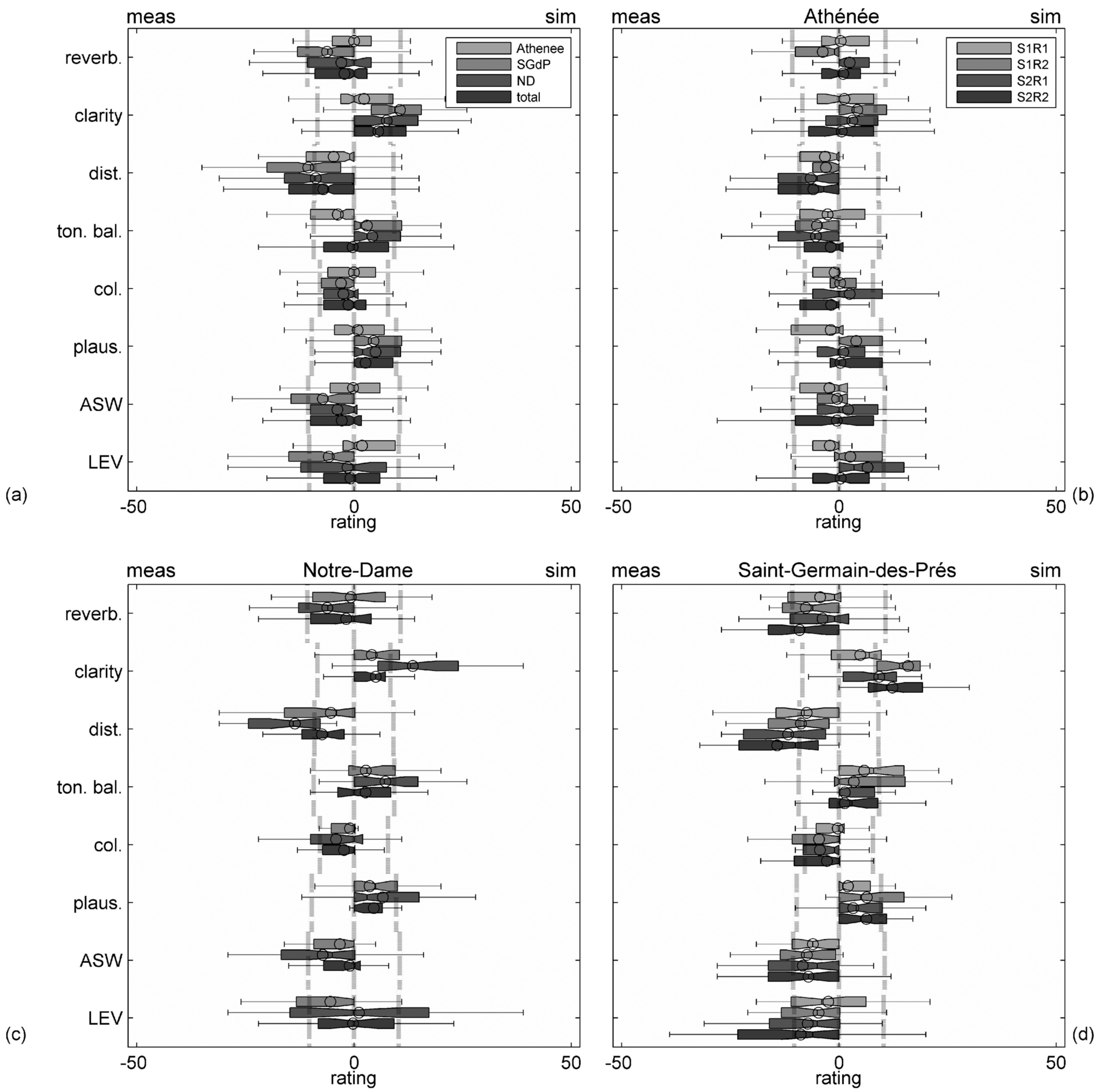

FIG. 4. Subject results on similarity of measured and simulated BRIR auralizations. **S** and **R** represent source and receiver positions. (a) Combined results and results by position for (b) Théâtre de l'Athénée, (c) Notre-Dame cathedral, and (d) Saint-Germain-des-Prés church. Center dashed line represents a neutral response. Dashed vertical bars represent individual attribute repeatability mean values. (See legend of Fig. 3 for boxplot notations and Sec. III C for attribute acronym description.)

perceived "clearer" and "closer," however, listening test results showed no significant difference between measurements and simulations. It should be noted that the direct correspondence between perceived *clarity* and the parameters C50 and C80 is not well defined, and that examining these parameters individually as a function of frequency may be overemphasising their quality as a metric.

As singing stimuli were employed for the Notre-Dame cathedral and Saint-Germain-des-Prés church, C80 was compared to *clarity* and *distance* in these rooms.[26] Correlation for the Notre-Dame results is good, with position **S2R1** being judged "clearer" and "closer" for the simulated auralization with C80 being 2.5ϵ higher. Correlation between subjective results and C80 for the Saint-Germain-des-Prés is good as well, with the simulated auralizations for the positions **S1R2**, **S2R1**, and **S2R2** having a perceptually higher C80 while also being judged "clearer" and "closer" in the listening test.

Correlation between *ASW* and IACC_e as well as *LEV* and IACC_l is rather poor. Differences in IACC_e exceeded 1ϵ for one position in the Notre-Dame cathedral and three in the Saint-Germain-des-Prés church while no significant differences for *ASW* were observed in the subjective listening test. Differences in IACC_l exceeded 1ϵ for all positions in the three rooms, however, no perceptual differences for *LEV* were observed.

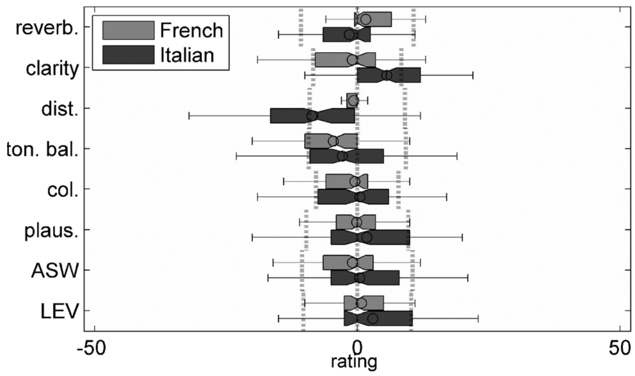

FIG. 5. Perceptual results for the Théâtre de l'Athénée seperated by stimuli. Center dashed line represents a neutral response. Dashed vertical bars represent individual attribute repeatability mean values. (See Fig. 3 for boxplot notations and Sec. III C for attribute acronyms.)

D. Stimuli effect

Subjective results as a function of stimuli were analyzed. Figure 5 shows that the majority of attributes were judged equally between the *French speaking* and *Italian speaking* stimuli. A repeated measures analysis of variance ($\alpha = 0.05$ level) indicated significant differences for *reverberance*, *clarity*, and *distance* (*reverberance*: $F = 5.2$, $p = 0.02$; *clarity*: $F = 24.9$, $p < 10^{-2}$; *distance*: $F = 32.1$, $p < 10^{-2}$; *tonal balance*: $F = 1.2$, $p = 0.28$; *coloration*: $F = 0.3$, $p = 0.57$; *plausibility*: $F = 1.3$, $p = 0.25$; *ASW*: $F = 0.8$, $p = 0.36$; *LEV*: $F = 2.0$, $p = 0.16$). The *reverberance* judgement difference was mainly due to position **S1R2**. Since this perceptual attribute is related to T20, this parameter's results were compared between measurement and simulation in 1/3rd octave bands where the two anechoic stimuli recordings differed more than 5 dB in spectral content, namely, 200, 4000, 6300, 8000, 10 000, and 12 500 Hz (see Fig. 2) with ΔT20(*measurement – simulation*) being $\Delta[+0.87, -0.14, -0.08, 0.00, -0.02, -0.18]$, respectively. As the simulation underestimated the T20 significantly in the 200 Hz 1/3rd octave band and the *Italian speaking* stimuli contained more energy in this 1/3rd octave band, it can be hypothesized that this resulted in the observed difference as a function of stimuli.

Additionally, differences regarding *clarity* and *distance* were mainly due to judgements at position **S2R2**. Since both of these perceptual attributes are related to C50, this parameter's results were compared between measurement and simulation in the same 1/3rd octave bands. It was observed that C50 was overestimated in the simulation for these 1/3rd octave bands by an average of 1.4 dB with a SD of 0.9 dB. In these 1/3rd octave bands, the *Italian speaking* stimuli contained more energy with the early-to-late energy being overestimated in the simulation. It can be hypothesized that the combination of these factors resulted in the observed differences in *clarity* and *distance* judgements as a function of stimuli.

V. DISCUSSION

This study presented the results of a subjective listening test which compared simulated and measured binaural auralizations. Hypothesis **H1** considered calibration of a set of omni-directional source-receiver combinations by bringing the clarity and reverberation parameters within a selected tolerance range, here taken as 1 JND from the ISO 3382 standard, results in valid binaural auralizations. The validity of **H1** was shown by comparing measured to simulated auralizations that resulted in perceptually equivalent binaural auralizations for the calibrated GA models of the Théâtre de l'Athénée, Notre-Dame cathedral, and Saint-Germain-des-Prés church.

These binaural auralizations were compared according to eight perceptual acoustic attributes commonly used in room acoustic evaluations. Listening test results showed no significant perceptual differences regarding *reverberance*, *tonal balance*, *coloration*, *plausibility*, *ASW*, and *LEV*. However, some trends in perceptual attribute differences were found which slightly exceeded participant repeatability tolerances, specifically *distance* and *clarity* for several positions in the Notre-Dame and Saint-Germain-des-Prés models. These differences could be associated to small variations in clarity parameter results between measurement and simulation RIRs which were unable to achieve the calibration goal of $<1\epsilon$ error for clarity while still adhering to other calibration procedure conditions. The geometrical approximations used for the model construction, specifically the means of simplified modelling of the seating, could be a source of common error in these two GA models. However, these perceptual differences were not observed in the preliminary listening test, where participants were required to listen to the entire extract from beginning when judging. We assume that the observed differences are due to the ability to carry out more detailed listening on specific sections of the extracts, rather than global ratings as was the case in the previous study. It is therefore recommended that for critical auralization listening tests, participants be able to select which elements of the extract they listen to, so that they can focus on key points where differences may be more evident.

Additionally, reverberation and clarity parameters corresponded well to the perceptual attributes *reverberance* and *clarity*. However, results showed no significant differences for *ASW* and *LEV* while differences between measurement and GA simulations exceeded ϵ reference values for IACC_e and IACC_l.

Hypothesis **H2** considered if calibration for a sufficient ensemble of discrete points provides sufficient confidence in the quality of the simulated RIR at other positions was validated in part. The positions of the studied binaural source-receiver combinations in the Théâtre de l'Athénée and Saint-Germain-des-Prés differed from the calibrated omni-directional combinations and provided perceptually valid auralizations indicating the validity of this assumption. The use of binaural auralization presents an additional extrapolation as the data used for calibration employed omni-directional microphones, not the measured BRIRs. However, as these binaural receiver position were within the measurement grid, this hypothesis should be studied with radically different positions in order to be fully validated.

Furthermore, the influence of anechoic stimuli selection was examined in auralizations of the Théâtre de l'Athénée using two different talkers. Results indicated that significant differences in some attribute judgments existed between the anechoic stimuli, linked to fine differences in spectral content. Therefore, in order to achieve more generalizable results, it is recommended that subjective tests employ multiple stimuli with varying frequency content.

Parameter estimations and subjective listening test results are compared to studies[10–12,14] discussed in Sec. I. Section III showed that using this calibration method, objective parameter estimation for the three considered GA models was better than previous studies. The binaural parameter estimations showed fewer positions exhibiting differences $>1\epsilon$ ($1\epsilon = 1$ JND) and a smaller maximum difference than those reported in previous studies. In addition, these studies failed to control for several factors between measurement and simulation where the employed auralizations in the presented listening test did. These being source/receiver directivity, HRTF, compensation for the frequency response characteristics of the measurement system, and differences in SNR between measurement and simulation. Even though the main objective of these studies differed from creating perceptually equal auralizations, a combination of these aspects resulted in simulated auralizations which varied from the measured auralizations. The simulated auralizations in this study were judged perceptually equal to the measured auralizations.

VI. CONCLUSION

The results presented here confirmed those from a preliminary study evaluating auralizations for omni-directional and binaural receivers. These combined results show that with a current commercially available GA software [specifically CATT-ACOUSTIC (v.9.0.c:3, TUCT v1.1a:4) in these studies], employing a methodical calibration approach based on few parameters evaluated for omni-directional source-receiver pairs, one is capable of producing perceptually equivalent spatial BRIRs when compared to measured BRIRs.

It should be noted that auralizations in this study used an omni-directional source directivity definition. As real sources have more complicated directivity patterns,[27] studies are currently exploring the effect of dynamic voice directivity with regards to dynamic direction and phoneme dependent radiation patterns.[28] Additionally, having validated the perceptual quality of acoustic auralization simulations, one can examine the impact of the addition of a visual model to the auditory perception of the virtual reality experience.[29]

ACKNOWLEDGMENTS

We would like to thank the personnel of the Théâtre de l'Athénée, Notre-Dame Cathedral, and Saint-Germain-des-Prés church for their assistance and patience during the measurements. Additional thanks to THALIM for their help in hosting the listening test and to all participants of the listening test for their time. Special thanks to Bengt-Inge Dalenbäck (CATT-ACOUSTIC) for the numerous and lengthy informative discussions, critical ear during the preparation of the auralizations, and comments in preparing this manuscript. This work was funded in part by the ECHO project (ANR-13-CULT-0004, echo-projet.limsi.fr). Partners include THALIM/ARIAS-CNRS, Bibliothèque nationale de France (BnF), and LIMSI-CNRS.

[1]M. Kleiner, B. Dalenbäck, and P. Svensson, "Auralization—An overview," J. Audio Eng. Soc. **41**, 861–875 (1993).
[2]S. Siltanen, T. Lokki, and L. Savioja, "Rays or waves? Understanding the strengths and weaknesses of computational room acoustics modeling techniques," in *International Symposium on Room Acoustics (ISRA)* (2010), pp. 1–6.
[3]P. Luizard, M. Otani, J. Botts, L. Savioja, and B. F. Katz, "Comparison of sound field measurements and predictions in coupled volumes between numerical methods and scale model measurements," Proc. Mtgs. Acoust. **19**, 015114 (2013).
[4]M. Vorländer, "International round robin on room acoustical computer simulations," in the *International Congress on Acoustics* (1995), pp. 689–692.
[5]I. Bork, "A comparison of room simulation software—The 2nd round robin on room acoustical computer simulation," Acta Acust. Acust. **86**, 943–956 (2000).
[6]I. Bork, "Report on the 3rd round robin on room acoustical computer simulation. Part I: Measurements," Acta Acust. Acust. **91**, 740–752 (2005).
[7]I. Bork, "Report on the 3rd round robin on room acoustical computer simulation, Part II: Calculations," Acta Acust. Acust. **91**, 753–763 (2005).
[8]B. Katz and E. Wetherill, "Fogg art museum lecture room, a calibrated recreation of the birthplace of room acoustics," in *Forum Acusticum* (2005).
[9]B. N. Postma and B. F. G. Katz, "Acoustics of Notre-Dame Cathedral de Paris," in *International Congress on Acoustics (ICA)*, 0269:1-10, Buenos Aires (2016), http://www.ica2016.org.ar/ica2016proceedings/ica2016/ICA2016-0269.pdf (Last viewed June 30, 2016).
[10]T. Lokki and H. Järveläinen, "Subjective evaluation of auralization of physics-based room acoustics modeling," in *International Conference on Auditory Display* (2001), pp. 1–6.
[11]T. Lokki and V. Pulkki, "Evaluation of geometry-based parametric auralization," in *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio* (2002), pp. 367–376.
[12]Y.-J. Choi and F. Fricke, "A comparison of subjective assessments of recorded music and computer simulated auralizations in two auditoria," Acta Acust. Acust. **92**, 604–611 (2006).
[13]ISO 3382-1, "Acoustics—Measurement of room acoustic parameters—Part 1: Performance spaces" (International Organization for Standardization, Geneva, Switzerland, 2009).
[14]W. Yang and M. Hodgson, "Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms," Acta Acust. Acust. **93**, 991–999 (2007).
[15]B. Postma and B. Katz, "Creation and calibration method of virtual acoustic models for historic auralizations," Virtual Real. **19**, 161–180 (2015).
[16]B. Katz, "International round robin on room acoustical response analysis software 2004," J. Acoust. Soc. Am. **4**, 158–164 (2004).
[17]D. Cabrera, J. Xun, and M. Guski, "Calculating reverberation time from impulse responses: A comparison of software implementations," Acoust. Aust. **4**(2), 1–10 (2016).
[18]L. Beranek, *Acoustics* (Acoustical Society of America, Melville, NY, 1986), pp. 1–643.
[19]L. Beranek, *Concert and Opera Halls: How they Sound* (Acoustical Society of America, Melville, NY, 1996), pp. 1–643.
[20]M. Vorländer, *Auralizations Fundamentals of Acoustics, Modeling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer-Verlag, Berlin, 2008), pp. 1–335.
[21]The notation employed here denotes the source-receiver combination of Source 2 and Receiver 1. This notation is employed throughout this paper.
[22]B. Postma, A. Tallon, and B. Katz, "Calibrated auralization simulation of the abbey of Saint-Germain-des-Prés for historical study," in the *International Conference on Auditorium Acoustics* (2015), Vol. 37, pp. 190–197, available at

https://www.researchgate.net/publication/283457384_Calibrated_auralization_simulation_of_the_abbey_of_Saint-Germain-des-Pres_for_historical_study (Last viewed June 30, 2016).
[23]T. Lokki, J. Pätynen, and V. Pulkki, "Recording of anechoic symphony music," in *Proceedings of Acoustics* (2008), pp. 6431–6436.
[24]In the preliminary test defined as *Coloration*.
[25]R. Thiele, "Richtungsverteilung und zeitfolge der schallruckwurfe in räumen" ("Directional distribution and sequence of the sound reflections in rooms"), Acustica **3**, 291–302 (1953).
[26]W. Reichardt, O. Abdel Alim, and W. Schmidt, "Definition und messgrunde eines objektiven masses zur ermittlung der grenze zwischen brauchbarer und unbrauchbarer durchsichtigkeit bei musikarbietung" ("Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances)," Acustica **32**, 126–137 (1975).
[27]B. Katz and C. d'Alessandro, "Directivity measurements of the singing voice," in *International Congress on Acoustics*, Madrid (2007), Vol. 19, pp. 1–6, https://www.researchgate.net/publication/236213215_Directivity_measurements_of_the_singing_voice (Last viewed June 30, 2016).
[28]B. N. Postma and B. F. Katz, "Dynamic voice directivity in room acoustic auralizations," in *German Annual Conference on Acoustics (DAGA)* (2016), pp. 352–355.
[29]D. Poirier-Quinot, B. N. Postma, and B. F. G. Katz, "Augmented auralization: Complimenting auralizations with immersive virtual reality technologies," in *International Symposium on Music and Room Acoustics (ISMRA)*, La Plata (2016), pp. 1–10.