Uncertainties of room acoustics simulation due to directivity data of musical instruments

Ernesto Accolti, Javier Gimenez, and Michael Vorländer

***Abstract*—Simulations and auralization methods in the field of room acoustics require the directivity of sound sources in third-octave frequency bands, which simplifies the calculation algorithms but introduces uncertainty. However, this uncertainty is not well known. A better understanding of the uncertainty is relevant because it can lead to improved accuracy of simulations, or at least allow users to quantify the uncertainty. In this article, a dataset of measurements of musical instruments is analyzed to study the uncertainty of the directional factor in bands. First, the directional factors of each partial is estimated as a reference. Then, different methods for estimating the directional factor in third-octave bands are analyzed, and the method of averaging directional indices is selected due to its lower uncertainty and error. Finally, the uncertainty propagation due to the selected method compared to the reference is studied based on predictions of sound level and loudness temporal profiles of about 50000 played notes of 41 musical instruments in a concert hall. The effects of source-receiver positions and instruments are shown through probability density function estimations of the sound level differences and excess loudness ratios. About 50% of the studied data shows sound level differences close to ± 5 dB and loudness excess ratio close to about $\pm 35\%$. This work provides a method for building future datasets of directivity in bands and a better understanding of the uncertainty of room acoustics simulations.**

I. Introduction

The radiation pattern of musical instruments depends on the played note, its frequency, its dynamic, its articulation, and its playing method, among other factors. In order to simplify some of these effects, the radiation pattern is usually estimated in frequency bands, which reduces computations but incorporates uncertainty.

Knowledge about the uncertainty of the source directional factor is an issue in room acoustics simulations and auralization methods. The effect of directional uncertainties is particularly important in direct sound and first reflections, since it influences considerably in the perceived sound quality compared to higher order reflections or late reverberation.

The input data regarding directivity in current commercial and academic software for simulation and auralization are the directional indices for each frequency band at a set of directions. These datasets are based on average of the spherical harmonics coefficients of the measured sound energy in each frequency band [1].

Some pioneering studies on the effect of third-octave band averaging focused on the similarity of the shape of partials' radiation patterns [2], [3]. The effects of band averaging are perceived during listening experiments and noticed in predicted room acoustical parameters [4]. The uncertainty effects due to other factors such as ambisonics encoding order on perceived differences has been investigated [5], [6]. Besides, some studies adapting directivity data for other simulation methods (e.g. wave-based) explore how to improve the source directivity including some musical instruments [7].

Apart from these studies, the uncertainty of directivity data has not received much attention in the literature. Moreover, no comprehensive investigation regarding averaging in bands has been previously carried out as far as the authors know.

The main motivation of this paper is the lack of uncertainty data of the directional index as an input parameter for room acoustics simulations and auralization. Although the path to study this uncertainty seems straightforward, definitions and estimation methods are not standard nor systematic.

A common practice adopted in simulation software is modeling the directional index in (1/3) third-octave bands. Some software also allow the use of (1/1) octave bands. However, as a rule of thumb, larger bands may fail to model sound diffusion (i.e. decreasing the precision of the diffusion model) and narrower bands may increase computational costs (i.e. requiring more instances of the ray tracing model).

In this paper, the uncertainty of using a unique directional factor for each third-octave band is analyzed based on the farfield sound pressure level of musical instruments at uniform sampled points of a sphere. Then, the propagation of this uncertainty to the sound pressure level and loudness at receivers in a simulated concert hall is analyzed. These two analyzed sets can be understood as input directivity data and output simulated sound signals. In order to carry out a comprehensive investigation on the topic, the probability density function (PDF) of both the source directivity index and the sound pressure level for simulated rooms are estimated.

The remainder of this paper is organized as follows. The conceptual framework and preliminaries of notation are introduced in Sec. II. Subsequently, Sec. III analyzes the errors and uncertainties of sound level in the free-field due to the averaging in bands with three methods. Afterward, the most accurate averaging method is applied to a comprehensive dataset to simulate sound signals in a concert hall in Sec. IV and analyze their uncertainty propagation in Sec. V. Finally, Sec. VI contains the discussions and Sec. VII contains the conclusions.

Ernesto Accolti and Javier Gimenez are with the Instituto de Automática (Institut of Automation) from both the National University of San Juan (UNSJ) and the National Scientific and Technical Research Council (CONICET), Argentina. Email: eaccolti@inaut.unsj.edu.ar. Michael Vorländer and E. A. are with the Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany. E. A. acknowledge the support by the Alexander von Humboldt Foundation.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes probability density plots of 40 musical instruments grouped by tone. Contact eaccolti@inaut.unsj.edu.ar for further questions about this work.

II. Conceptual framework

Datasets of radiation patterns typically contain measurements of the sound pressure generated by a musician playing chromatic scales in the center of a microphone array in a free-field environment. Sounds of almost all musical instruments involve periodic vibration processes, where individual vibrations combine to form a harmonic series of partial tones [8].

Let a_ℓ^z be the ℓ-th order partial tone of a note z for which the lowest frequency f_1^z is the fundamental frequency. Then, the frequencies of the partial tones are $f_\ell^z = \ell f_1^z$, $\ell = 2, \ldots, B$, where B is the number of considered harmonics. These partial tones are extracted from the Fourier transform of the note.

The sound pressure of each partial a_ℓ^z on a spherical surface of fixed radius r_0 in the farfield is

$$p_\ell^z(\theta, \phi) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} p_{nm}(a_\ell^z) Y_n^m(\theta, \phi), \tag{1}$$

where θ and ϕ are elevation[1] and azimuth angles, respectively; and $Y_n^m(\theta, \phi)$ are the Spherical Harmonics (SH) functions of degree n and order m, which form an orthogonal basis on the radiation patterns' set. In addition, the SH coefficients are

$$p_{nm}(a_\ell^z) = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} p_\ell^z(\theta, \phi)(Y_n^m(\theta, \phi))^* \cos\theta d\theta d\phi, \tag{2}$$

estimated with

$$\widehat{p}_{nm}(a_\ell^z) = \sum_{j=1}^{J} p_\ell^z(\theta_j, \phi_j)(Y_n^m(\theta_j, \phi_j))^*, \tag{3}$$

where the superscript $*$ indicates the complex-conjugate operator, and (θ_j, ϕ_j), $j = 1, \ldots, J$, is a representative sample of directions. Thus, the sound pressure can be approximated by

$$\widehat{p}_\ell^z(\theta, \phi) = \sum_{n=1}^{N} \sum_{m=-n}^{n} \widehat{p}_{nm}(a_\ell^z) Y_n^m(\theta, \phi), \tag{4}$$

where N is the order of the approximation. From now on, the hat will be avoided in the notations for simplicity.

The dataset [9] provides estimates of these coefficients for 41 musical instruments considering $r_0 = 2.1$ m, $B = 10$, $J = 32$, and $N = 4$. For this, musicians were asked to play their instruments in the center of a regular and spherical array of microphones. The physical center of the microphone array is aligned with the acoustic center of each instrument using the center-of-mass approach [10] below 0.5 kHz, and the phase symmetry approach above 0.5 kHz [11].

Useful definitions for room acoustics are the directivity factor defined as the ratio between the intensity in each direction and the average intensity, and the directional factor defined as the ratio between the pressure in each direction and the pressure at a reference direction [12], [13]. The definition adopted in this article for the directional factor is

$$\Gamma_\ell^z(\theta, \phi) = \frac{p_\ell^z(\theta, \phi)}{p_\ell^z(\theta_{\text{ref}}, \phi_{\text{ref}})}, \tag{5}$$

and for directional index is

$$L_\ell^z = 20 \log_{10}\left(\Gamma_\ell^z\right). \tag{6}$$

[1]The horizontal plane is defined as $\theta = 0$

In practice, the partials are organized by frequency bands and a unique and representative directional factor for each band is generated. The third-octave bands are defined in this article (according to the IEC 61260-2014, ISO 3 and ISO 266 standards) as $\mathcal{B}_i = [f_c(i) \times 2^{-1/6}, f_c(i) \times 2^{1/6})$, where $f_c(i) = 1000 \times 2^{(i-18)/3}$ is the central frequency that represents the band. For each musical instrument and for each dynamic (pianissimo and fortissimo), define the set $A_i = \{a_\ell^z | f_\ell^z \in \mathcal{B}_i\}$ for averaging purpose in Sec. III.

III. Averaging bands

Three alternative methods to generate a unique and representative directional factor for each band are shown in this section. Then, a comparison of the performance of the three methods in free-field is presented.

A. Averaging SH coefficients

The average SH coefficients for the i-th band are [1, eq. 16]

$$\overline{p}_{nm}(i) = \frac{1}{\#A_i} \sum_{a_\ell^z \in A_i} p_{nm}(a_\ell^z), \tag{7}$$

where $\#$ indicates cardinality. Then, for each band i, the sound pressure is estimated as

$$\overline{p}_i(\theta, \phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \overline{p}_{nm}(i) Y_n^m(\theta, \phi), \tag{8}$$

and the directional factor based on averaging SH coefficients

$$\Gamma_{\overline{p}_i}(\theta, \phi) = \frac{\overline{p}_i(\theta, \phi)}{\overline{p}_i(\theta_{\text{ref}}, \phi_{\text{ref}})}. \tag{9}$$

This method is the state of the art. Notice that it is equivalent to average the sound pressure (i.e. an average on a_ℓ^z of (1) gives (8) which is the numerator of (9)).

B. Averaging directional factor

Alternatively, the directional factor can be estimated based on averaging the directional factor itself, i.e.,

$$\overline{\Gamma}_i(\theta, \phi) = \frac{1}{\#A_i} \sum_{a_\ell^z \in A_i} \Gamma_\ell^z(\theta, \phi). \tag{10}$$

C. Averaging directional index

A third option is to average directional indices, i.e.,

$$\Gamma_{\overline{L}_i}(\theta, \phi) = 10^{\frac{\overline{L}_i(\theta, \phi)}{20}}, \tag{11}$$

with

$$\overline{L}_i(\theta, \phi) = \frac{1}{\#A_i} \sum_{a_\ell^z \in A_i} L_\ell^z(\theta, \phi). \tag{12}$$

D. Uncertainty estimation

The uncertainty due to the method used to estimate the unique directional factor is computed with

$$e_\ell^z(\theta, \phi) = 20 \log_{10} \frac{\Gamma_{\text{method}}(\theta, \phi)}{\Gamma_\ell^z(\theta, \phi)}, \tag{13}$$

where Γ_{method} takes the value of $\Gamma_{\overline{p}_i}$, $\overline{\Gamma}_i$, or $\Gamma_{\overline{L}_i}$. Notice that the relation between these three averages is not linear.

E. Analysis of averaging methods

In order to perform a comparative analysis, the averaging methods are applied for the whole dataset. A total of 2×10^4 directions (θ, ϕ) sampled with a recursive zonal equal area sphere partitioning algorithm [14] are considered. In Sec III-E1 the average elements are introduced with an example. Then, in Sec. III-E2 the distribution of the errors are compared.

1) Preliminary results: Fig. 1 shows views of the directional factors for just three third-octave bands and just 30 directions in the horizontal and vertical planes. These directional factors are estimated with the three different methods: averaging the SH coefficients as in eq. (9), averaging the directional factors as in eq. (10), and averaging the directional indices as in eq. (11). In addition, all the partials that fall in each band for all the notes (i.e. the played notes go from F#3 to F6) are shown as reference.

The values of $\#A_i$ varies from one third octave band to another. For instance, in the 250 Hz band, fall just the four fundamental tones (first order partials) for musical notes A#3, B3, C4, and C#4, with frequencies 233 Hz, 247 Hz, 262 Hz, and 277 Hz, respectively (dots in first column of Fig. 1). In the 500 Hz band fall a total of nine partials (dots in second column of Fig. 1); including four fundamental tones (i.e. partials corresponding to the notes A#4, B4, C5, C#5, with frequencies 466 Hz, 494 Hz, 524 Hz, and 554 Hz, respectively), four second order partials (i.e. partials corresponding to the notes A#3, B3, C4, C#4 with frequencies 466 Hz, 494 Hz, 524 Hz, and 554 Hz, respectively), and one third order partial (i.e. partial corresponding to the note F#3 with frequency $175 \times 3 = 525$ Hz). In the 1000 Hz band fall a total of nineteen partials (dots in the third column of Fig. 1) including two fundamental tones, four second order partials, four third order partials four forth order partials, four fifth order partials, and one sixth order partial.

These preliminaries are mainly shown for introducing the whole problem with most of its particularities. For example, the estimates based on $\overline{\Gamma}_i$ and $\Gamma_{\overline{L}_i}$ rather than on $\Gamma_{\overline{p}_i}$ are closer to the estimates on partials (Fig. 1). Although estimates based on $\overline{\Gamma}_i$ and $\Gamma_{\overline{L}_i}$ are very similar in the 250 Hz frequency band, the estimate based on $\Gamma_{\overline{L}_i}$ is closer to the data of the partials, as can be deduced from the next subsection.

2) Uncertainty distribution in bands: The three methods were applied in order to obtain the uncertainty e_ℓ^z as in (13) with $\Gamma_{\overline{p}_i}$, $\overline{\Gamma}_i$, and $\Gamma_{\overline{L}_i}$ for the 41 musical instruments at both pianissimo and fortissimo subsets. The number of uncertainties e_ℓ^z computed for each instrument at each dynamic level is $2 \times Z \times 10^5$, as there are 2×10^4 directions, 10 partials and Z played notes.

Fig. 2 shows the probability density function (PDF) [2] of e_ℓ^z grouped by bands (i.e. each band considers the partials in the subset A_i for all the 2×10^4 directions) for a trumpet. The variance, bias, and interquartile range of these PDFs quantify the uncertainty of the corresponding estimation.

[2]PDFs are estimated by the kernel method with 0.1 dB (Gaussian) bandwidth which is similar to the extended uncertainty of the measurement system

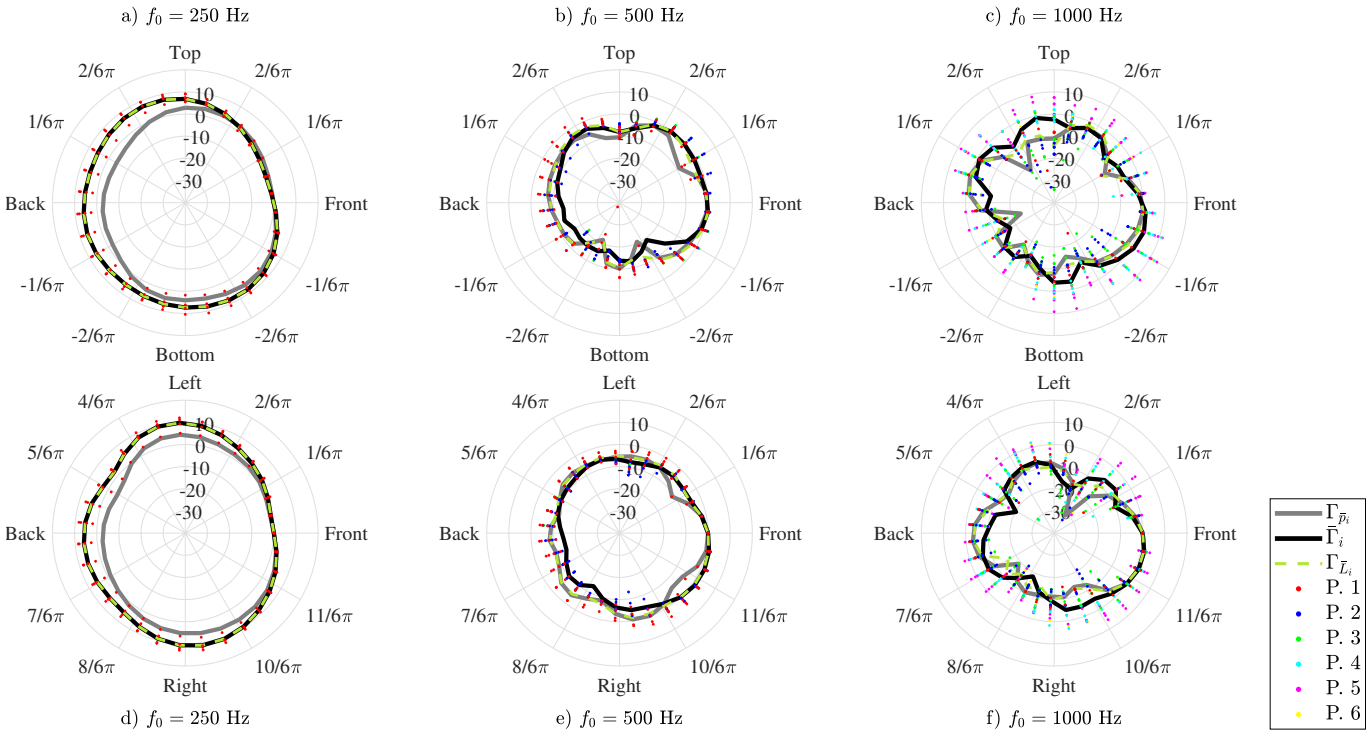

Fig. 1: Directional factor estimations for a trumpet in vertical plane (a-c) and horizontal plane (d-f). Solid gray: $\Gamma_{\overline{p}_i}$ estimation averaging the SH coefficients with eq. (9), solid black: $\overline{\Gamma}_i$ estimation averaging directional factors with eq. (10), and segmented green: $\Gamma_{\overline{L}_i}$ estimation averaging the directional indices with eq. (11). Dots: partials (see partial order in references).

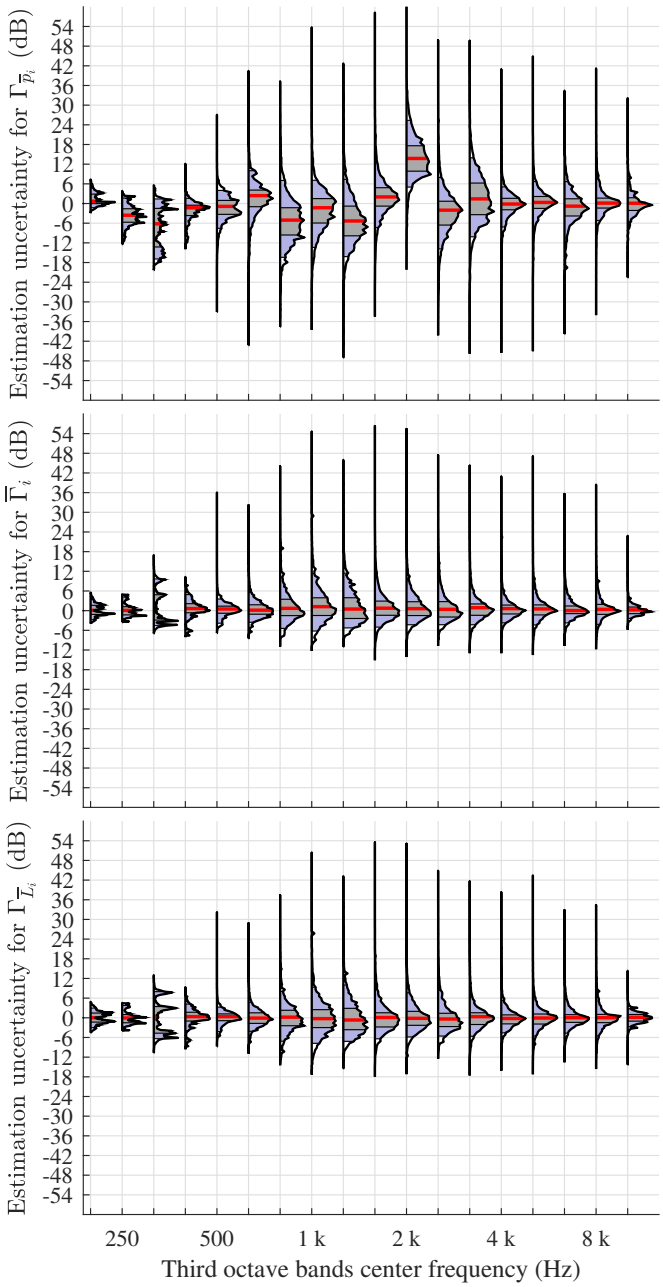

Fig. 2: PDF of the uncertainties due to third octave band average of directional factors of a trumpet. Dark shaded zone: interquartile range; light shaded zone: 5%–95% range; red line: median

F. Comparison of the three averaging methods

Fig 3 shows the mean $\overline{e}_\ell^z$ and interquartile range $IR(e_\ell^z)$ grouped by instrument and musical dynamic. As previously suggested, averaging directional index yields the smallest uncertainty.

IV. Directivity performance in a simulated hall

The simulation signal $\boldsymbol{y}$ in a room can be modeled as a linear system by the convolution of an anechoic audio signal $\boldsymbol{x}$ with a room impulse response $\boldsymbol{h}$, i.e.

$$\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{h}. \tag{14}$$

The source directivity effect is included in $\boldsymbol{h}$. Hence, due to homogeneity property of linear systems, the $\boldsymbol{h}$ with less uncertainty will render an output-simulated signal $\boldsymbol{y}$ with less uncertainty. Since the ultimate goal of this paper is to investigate the smallest uncertainty of $\boldsymbol{y}$ due to the uncertainty of directivity as input data, just the method of averaging directional indices $\Gamma_{\overline{L}_i}$ is explored (i.e. due to its small uncertainty).

The structure of this section is as follows. The modeling of the impulse response $\boldsymbol{h}$ is divided in three subsections. First, the room is simulated in Sec. IV-A. Subsequently, the directional factor for each band and for each note partial are estimated in Sec. IV-B. Then, the data obtained in the two previous subsections is used to generate the impulse responses in Sec- IV-C. Afterward, the input signals $\boldsymbol{x}$ are introduced in Sec. IV-D and the calculation of the output signals $\boldsymbol{y}$ is shown in Sec. IV-E. A total of about 30 hours of audio files (individually played notes) are generated in Sec. IV-E and post-processed in Sec. IV-F in order to be analyzed in Sec. V.

A. Room model and configuration

The sound pressure in a room is simulated at four receiver positions and two source positions. Fig. 4 shows a scheme of the room model, source poses (S_1 and S_2), and receiver poses (R_1, R_2, R_3, and R_4).

The room is a box-shaped concert hall with one balcony level. The number of surfaces is small for two reasons: keeping the room shape general and calculating reflections in a reasonable time. The absorption and scattering coefficients for the audience correspond to upholstered seats and the other surfaces are typical average coefficient for hard surfaces in concert halls (both coefficients are available in the software used). The inner volume of the hall is about 20000 m^3 and the area of the interior surfaces is about 5400 m^2.

A simulation with the image source method (ISM) up to the 4^{th} order [13] is carried out through the software RAVEN [15] using omnidirectional sources and receivers. The directional behavior of the source is modeled with the two different approaches in Sec. IV-B.

The reflection coefficients' impulse response vector $\boldsymbol{\rho}_\kappa$, the arrival time t_κ, and the direction of arrival $(\theta_\kappa, \phi_\kappa)$ of each ray κ are computed by RAVEN. ISM software compute the main output which is the room impulse response vector as follows

$$\boldsymbol{h} = \sum_\kappa \boldsymbol{\rho}_\kappa * \boldsymbol{\gamma}(\theta_\kappa, \phi_\kappa) * \mathrm{sinc}(f_\mathrm{s}(\boldsymbol{t} - t_\kappa)), \tag{15}$$

where $\boldsymbol{t} = [0, 1/f_\mathrm{s}, 2/f_\mathrm{s}, \cdots, T]$ is a time vector sampled at f_s sampling rate, T is the duration, $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$ is the sinc function, and $\boldsymbol{\gamma}$ is the directional-factor impulse response. In this case $\boldsymbol{\gamma} = 1$ is bypassed because this paper applies directivity outside RAVEN as shown in Sec. IV-C.

B. Directional factor modeling

The directional index L_ℓ^z is estimated with (6) for each partial a_ℓ^z. Then, the average directional index $\overline{L}_i$ is estimated with (12) for each third-octave band i of each musical instrument.

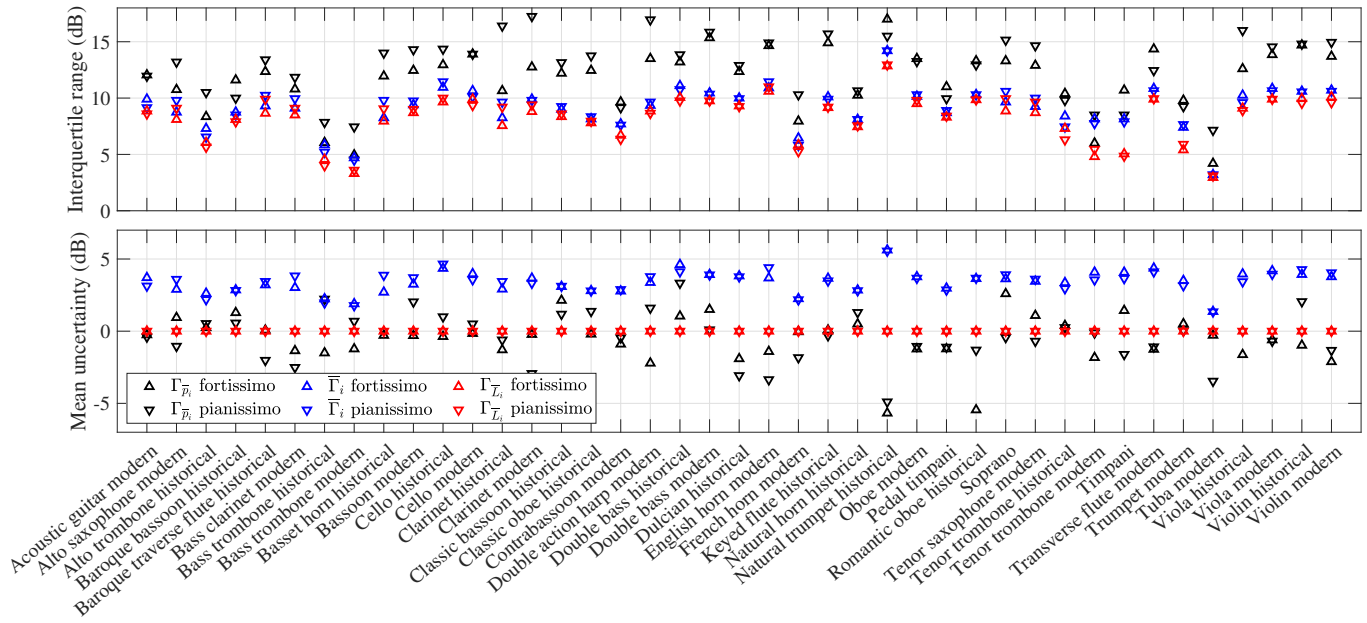

Fig. 3: Source directivity input uncertainties due to three band averaging methods

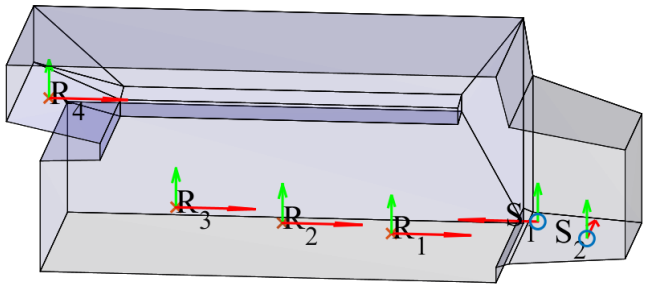

Fig. 4: Room setup. Audience dimensions are about 40 m long, 23 m wide, and 18.6 m high. Stage dimensions are about 10 m long, 18 m wide, and 12.5 m high.

Notice that L_ℓ^z applies to partial ℓ from note z, whereas $\overline{L}_i$ applies to all partials in band i from all notes.

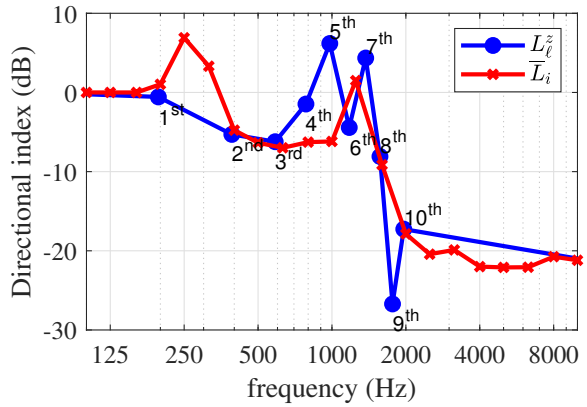

Fig. 5: Directional indices in the S_1–R_2 direction. Red crosses: $\overline{L}_i$ for third-octave bands; blue circles: L_ℓ^{G3} for partials of note G3. The partials' order (1^{st}, 2^{nd}, $\cdots$, 10^{th}) is written

Fig. 5 shows the directional index of a trumpet with these two approaches (L_ℓ^z and $\overline{L}_i$) for the direction from source S_1 to receiver R_2. L_ℓ^z applies just for z corresponding to a G3 note and $\overline{L}_i$ applies for all the played notes. The amount of partials per band increases with the frequency from just one partial every three bands at low frequencies to more than one partial per band in high frequencies.

The difference between $\overline{L}_i$ and L_ℓ^{G3} in Fig. 5 is the difference due to the averaging method for note G3. It is around 1 dB for the 1^{st} partial or fundamental, negligible for the 2^{nd} and 3^{rd} partials, and between 5 dB to 10 dB for the 4^{th} to the 7^{th} partials.

Let $f_\mathrm{s} = 44100$ Hz be the sampling rate and $\boldsymbol{f}_\mathrm{d} = f_\mathrm{s}[-N_\mathrm{d}/2, \ldots, -1, 0, 1, 2, \ldots, N_\mathrm{d}/2 - 1]^\mathrm{T}/N_\mathrm{d}$ a vector with $N_\mathrm{d} = 2^{15}$ elements. Then, $\boldsymbol{\Gamma}_\mathcal{B}$ is estimated with a linear interpolation into $\boldsymbol{f}_\mathrm{d}$ of a vector with elements $\Gamma_{\overline{L}_i}$ for the frequency corresponding to each band i. Afterward, the N_d-points directional factor impulse response based on bands is

$$\boldsymbol{\gamma}_\mathcal{B}(\theta_\kappa, \phi_\kappa) = \mathfrak{F}^{-1}[\boldsymbol{\Gamma}_\mathcal{B}(\theta_\kappa, \phi_\kappa)], \tag{16}$$

where $\mathfrak{F}^{-1}$ is the inverse discrete Fourier transform. In turn, let $\boldsymbol{\Gamma}_z$ be the linear interpolation into $\boldsymbol{f}_\mathrm{d}$ of a vector with elements Γ_ℓ^z for each partial of note z. Then, the N_d-points directional factor impulse response based on partials is estimated as

$$\boldsymbol{\gamma}_z(\theta_\kappa, \phi_\kappa) = \mathfrak{F}^{-1}[\boldsymbol{\Gamma}_z(\theta_\kappa, \phi_\kappa)]. \tag{17}$$

C. Room impulse responses of directional sources

The room impulse response is estimated based on the ISM outputs of RAVEN for each ray κ (i.e. ρ_κ, t_κ, and $(\theta_\kappa, \phi_\kappa)$). These outputs, generated with omnidirectional sources (see Sec. IV-A) are reused for each source–receiver combination.

The summation of the rays is implemented outside RAVEN in order to apply the directional factor impulse responses $\boldsymbol{\gamma}_z(\theta_\kappa, \phi_\kappa)$ from (17) corresponding to each ray κ by using

(15). In addition, $\boldsymbol{\gamma}_{\mathcal{B}}$ from (16) is used accordingly to estimate the impulse response based on bands. The number of impulse responses calculated for each instrument is Z for estimations based on partials ($\boldsymbol{h}_z$) whereas it is just one for estimations based on bands ($\boldsymbol{h}_{\mathcal{B}}$).

The convolution ($*$) in these cases are implemented in frequency domain using the algorithm called *FFT filter* [16, pp. 268]. The convolution between $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ multiplies N_d sized vectors in the frequency domain, whereas the convolution between that result and the sinc function multiplies $T \times f_\mathrm{s}$ sized vectors in the frequency domain. For that purpose, the $\mathrm{sinc}(\boldsymbol{t} - t_\kappa)$ function is estimated in the frequency domain as $e^{j\boldsymbol{\omega} t_\kappa}$, with $\boldsymbol{\omega} = 2\pi f_\mathrm{s} \boldsymbol{t}/T - f_\mathrm{s}/2$ as in [17, Sec. 3.3].

Fig. 6 shows the envelope of the resulting impulse responses with the two approaches for the same trumpet note. The envelopes are estimated by the Hilbert transform of 500 Hz, 1000 Hz, and 2000 Hz octave bands as in [18].

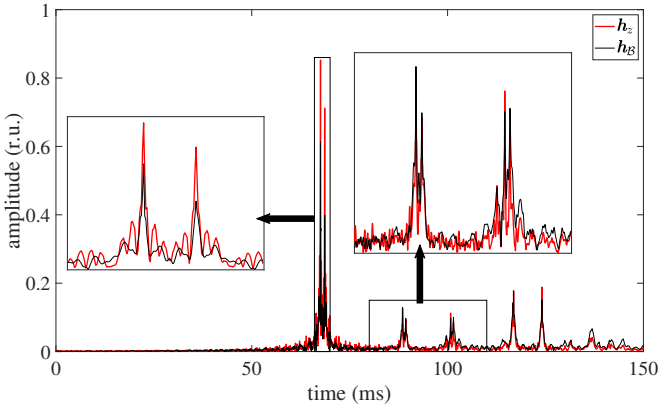

Fig. 6: Envelope of room impulse responses with partials directional factor $\boldsymbol{h}_z$ and bands directional factor $\boldsymbol{h}_{\mathcal{B}}$

The directional factor impulse response based on bands $\boldsymbol{\gamma}_{\mathcal{B}}(\theta_\kappa, \phi_\kappa)$ has more energy than that based on partials $\boldsymbol{\gamma}_z(\theta_\kappa, \phi_\kappa)$ for some rays and the opposite applies for other rays. In other words, some peaks are greater for impulse responses $\boldsymbol{h}_{\mathcal{B}}$ based on bands (e.g. Fig. 6 left zoom square) and some others peaks are greater for impulse responses $\boldsymbol{h}_z$ based on partials (e.g. Fig. 6 right zoom square). These preliminaries show that the relation of these differences with the source–receiver poses combination is not simple and that predicting the uncertainty at the output based on the directivity input data without generating impulse responses is not straightforward.

D. Input signals

The anechoic input signals $\boldsymbol{x}_z$ are sound pressure signals of isolated notes of about 2 s duration extracted from the frontal microphone of the same dataset [9]. The signals were recorded in the anechoic chamber of Technical University Berlin (room volume $V = 1070$ m^3, and certified lower frequency limit $f_g = 63$ Hz) using a surrounding spherical array of 32 microphones with radius 2.1 m.

During the recordings, musicians were looking at a reference direction ($\phi = 0$ and $\theta = 0$) with the main sound emitting part of their instrument centered inside the array. The signals in this part of the study are extracted from recordings at microphone number 4 that was situated close to the reference direction ($\phi = 0$ and $\theta = -\pi/18$).

The instruments were played by professional musicians at two dynamics: pianissimo and fortissimo. A chromatic scale non-legato was played, and then each note separated as if they were recorded one by one. The total number of notes is 3305 (for the 41 musical instruments at two dynamics) [1], [9].

E. Output signal

The reference output $\boldsymbol{y}_z$ based on partials is computed by

$$\boldsymbol{y}_z = \boldsymbol{x}_z * \boldsymbol{h}_z, \tag{18}$$

and the evaluated output $\boldsymbol{y}_{\mathcal{B},z}$ based on bands is computed by

$$\boldsymbol{y}_{\mathcal{B},z} = \boldsymbol{x}_z * \boldsymbol{h}_{\mathcal{B}}. \tag{19}$$

For each given musical instrument and dynamic, $\boldsymbol{h}_{\mathcal{B}}$ (based on bands) is the same whereas $\boldsymbol{h}_z$ (based on partials) is different for each note. The convolution ($*$) in this case is implemented with an FFT convolution algorithm [19, pp. 139].

The complete set of output signals includes the 3305 notes at 2 source positions and 4 receivers positions. Considering the 2 simulation approaches (i.e. based on bands and based on partials), the total number of generated signals is $3305 \times 2 \times 4 \times 2 = 52880$ yielding about 30 hours of audio.

F. Descriptors

In order to compare the simulations generated with the input directivity averaged in bands with the reference generated with the input directivity of each partial, several descriptors can be used. In this paper two time varying descriptors are calculated: the sound pressure level and the loudness.

The sound pressure level vectors $\boldsymbol{L}_z$ and $\boldsymbol{L}_{\mathcal{B},z}$ are estimated applying a moving average filter to the reference output signals $\boldsymbol{y}_z$ and the evaluated output signals $\boldsymbol{y}_{\mathcal{B},z}$, respectively. The moving average filter mimics a sound level meter with fast time weighting and linear frequency weighting[3].

The loudness vectors $\boldsymbol{N}_z$ and $\boldsymbol{N}_{\mathcal{B},z}$ are estimated with the monaural dynamic loudness model [20] applied respectively to $\boldsymbol{y}_z$ for reference estimations based on partials and $\boldsymbol{y}_{\mathcal{B},z}$ for estimations based on bands. Although actual loudness depends on directions of arrival, the monaural model avoids the influence of the uncertainty of head-related transfer functions.

The note signals $\boldsymbol{x}_z$ are about 2 s long and the time varying sound level and loudness are down-sampled to a 2 ms time period[4]. In consequence, the size of each vector $\boldsymbol{L}_z$, $\boldsymbol{L}_{\mathcal{B},z}$, $\boldsymbol{N}_z$, and $\boldsymbol{N}_{\mathcal{B},z}$ is about 1000 elements long for each note z.

Fig. 7 shows the results for sound pressure level and loudness estimations at receiver position R_2 for one note played in a trumpet at source position S_1. The differences are about 2 dB and about 1 sone.

[3]The fast weighting correspond to a 125-ms time constant for the moving average filter and the linear weighting is a common terminology to indicate no frequency compensation as opposed to A- or C-weighting compensations

[4]The sampling period for the dynamic loudness model is 2 ms and is in the same order of magnitude of the down-sampling applied by sound level meters

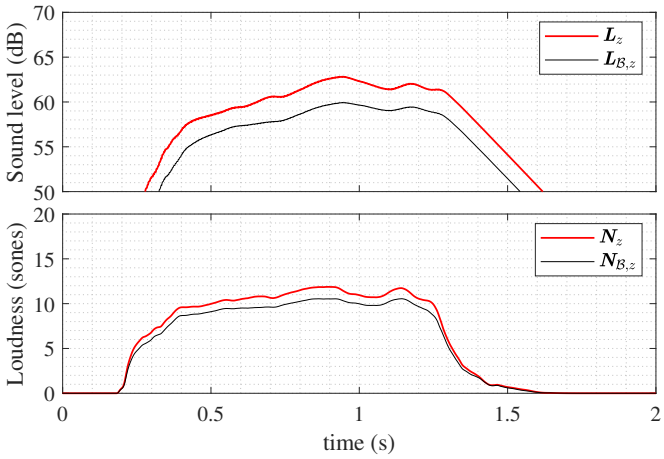

Fig. 7: Time analysis for a trumpet fortissimo G#3.

V. Effect at receivers of directivity in bands

The tails of the PDFs in Fig. 2 show some large differences which are not dominant in an actual auralization scenario. This section shows results which may be useful to approach the effects of that uncertainty on auralization.

Sound level differences of the time-varying $L_{\mathcal{B}}$ based in bands referred to L_z based on partials are computed as

$$\boldsymbol{\lambda} = \boldsymbol{L}_{\mathcal{B},z} - \boldsymbol{L}_z, \tag{20}$$

for the complete set of output signals. In addition, the loudness excess ratios of the time-varying $\boldsymbol{N}_{\mathcal{B}}$ based on bands referred to $\boldsymbol{N}_z$ based on partials is estimated as

$$\boldsymbol{\eta} = (\boldsymbol{N}_{\mathcal{B},z}/\boldsymbol{N}_z - 1) \times 100\%. \tag{21}$$

As shown in previous section, 26440 pairs of output signals were generated. Considering the 2 ms down-sampling of the descriptors, the total number of sound level differences plus loudness ratios for the complete set is about 52.88×10^6.

For presentation purposes, these values are shown in the next subsections with PDF estimations grouped by source-receiver combination in Sec. V-A, by instrument in Sec. V-B, and by note name in Sec. V-C. [5]

A. Source-receiver effect

Fig. 8 shows the PDFs of sound level differences $\boldsymbol{\lambda}$ grouped by source-receiver. The extreme values in Fig. 8 can be considered outliers since they represent data points with differences of about 2 ms in duration, which may not be perceptible during the playback of running music. Some data in the tails in these distributions is likely in similar case than this extreme values (i.e. may be not perceived due to its short duration).

Tab. I shows the mean difference $\overline{\lambda}$, the interquartile range u_{25-75}, and the 5%–95% range u_{5-95} for each source–receiver combination. The mean uncertainty $\overline{\lambda}$ grows about 1 dB with source–receiver distance. Besides the ranges are similar across source-receiver combinations with a tolerance of ± 0.6 and ± 1 for u_{25-75} and u_{5-95}, respectively.

[5]Kernel bandwidth: 0.1 dB for level and 1% for excess loudness ratio

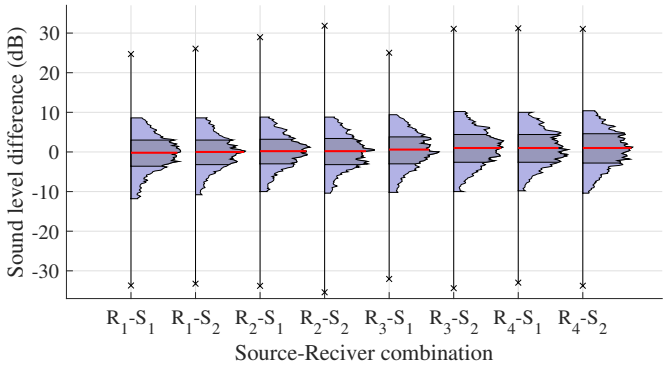

Fig. 8: PDFs of sound level differences $\boldsymbol{\lambda}$ grouped by source-receiver combination. Dark shaded zone: interquartile range; light shaded zone: 5%–95% range; red line: median; black $\times$: maximum and minimum

TABLE I: Uncertainty distribution (dB) for different source–receiver combinations

Source	1	2	1	2	1	2	1	2
Receiver	1	1	2	2	3	3	4	4
$\overline{\lambda}$	-0.4	-0.1	0.2	0.0	0.4	0.9	0.9	0.8
u_{25-75}	7.0	6.6	6.6	7.0	7.2	7.4	7.4	7.8
u_{5-95}	20.4	19.4	18.8	19.2	19.6	20.2	19.8	20.8

B. Results for each musical instrument

Fig. 9 shows the PDFs of sound level differences $\boldsymbol{\lambda}$ and loudness excess ratios $\boldsymbol{\eta}$ grouped by instruments. The PDFs are based on data of different sizes due to the different number of notes played by each instrument (e.g. 352×10^3 values for the guitar and 280×10^3 values for the alto saxophone).

The results show multimodal rather than Gaussian distributions. The multiple modes may occur due to factors such as source–receiver combination. This is clearer in Sec. V-C when grouping by note names.

Fig. 10 shows the interquartile range and mean of the sound level differences $\boldsymbol{\lambda}$ for each instrument. The interquartile ranges are between about 2 dB to 8 dB, and the main uncertainties are between about -7.5 dB to 7.5 dB, with exception to the case of Viola modern fortissimo which is -11.5 dB.

Fig. 11 shows the interquartile range and the mean of the loudness excess ratios $\boldsymbol{\eta}$ for all instruments. The interquartile range is less than 50% for most instruments at fortissimo dynamic.

C. Results for each note

The uncertainty for a set of notes from a set of instruments is relevant for the design and generation of auralization material. These uncertainties can be estimated from the PDFs of the level differences grouped by notes for each instrument which are provided in the supplementary material.

Fig. 12 shows the PDFs of sound level differences $\boldsymbol{\lambda}$ for an Acoustic guitar playing fortissimo, grouped by notes. Each distribution is estimated based on about 8000 values.

The first lower octave and the first third part of the second octave (i.e. from E2 to A#3) show level differences $\boldsymbol{\lambda}$ bounded at [-5, 10] dB for almost the 100% of the data. However, as far

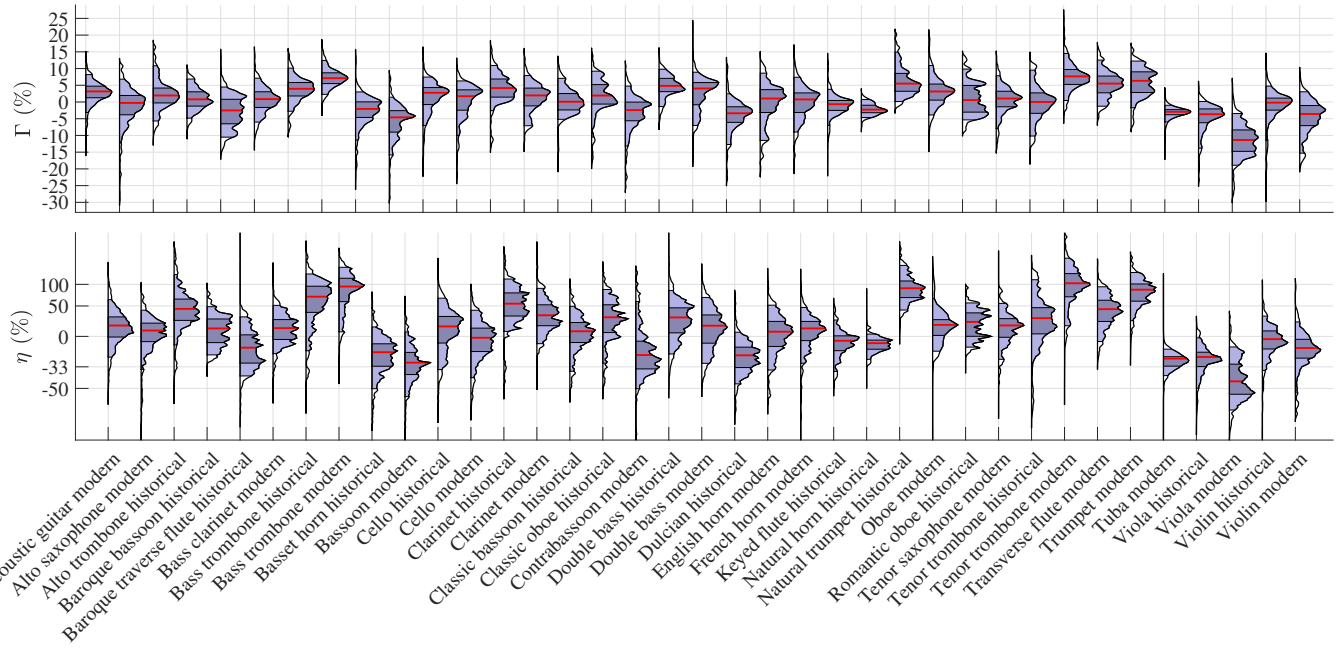

Fig. 9: Level differences $\boldsymbol{\lambda}$ and loudness excess ratio $\boldsymbol{\eta}$ distributions for simulations based on bands versus simulations based on partial. Dark shaded zone: interquartile range; light shaded zone: 5%–95% range; red line: median

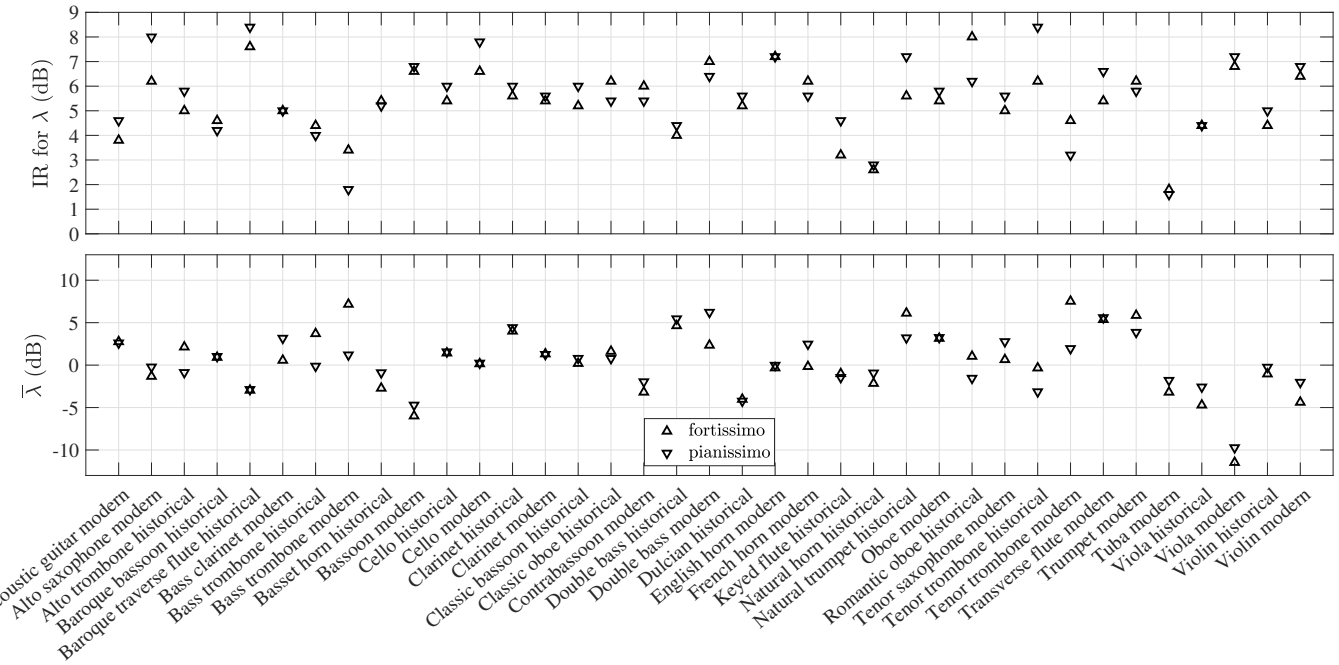

Fig. 10: Uncertainties in simulated levels in a room due to band averaging directional indexes

as the note pitch goes higher, the differences increase reaching some extreme values for certain notes (e.g. the 100% of the data is in the range [-14, -7] dB for the E5 note).

VI. Discussion

Results show that the uncertainty of both the directional factors due to band averaging and the predicted sound level can be significant. Moreover, the provided results may be relevant for the control and report of uncertainty during activities of design and generation of auralization material.

Systematic differences due to averaging methods were found (Fig. 3). From smaller to larger uncertainties, the estimations are: $\Gamma_{\overline{L}_i}$, $\overline{\Gamma}_i$, and $\Gamma_{\overline{p}_i}$. Hence, just $\Gamma_{\overline{L}_i}$ is recommended for future work. The interquartile range of $\Gamma_{\overline{L}_i}$ is about 6 dB for the trumpet (Fig. 2) and within $[3\ 13]$ dB for the complete dataset (Fig. 3).

These large uncertainties, especially on the tails of the distributions, are not dominant at all as regards sound level loudness in a hall. This was confirmed by evaluating the time-varying sound pressure level and loudness of the complete

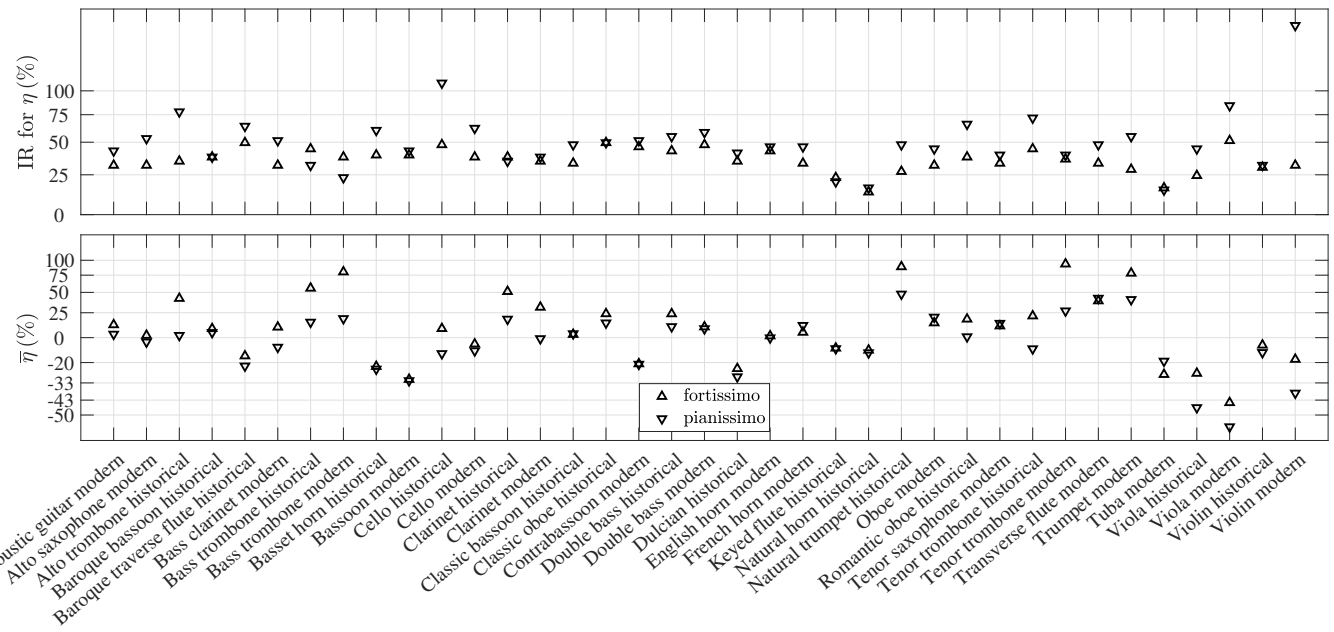

Fig. 11: Loudness differences in simulated rooms due to band averaging

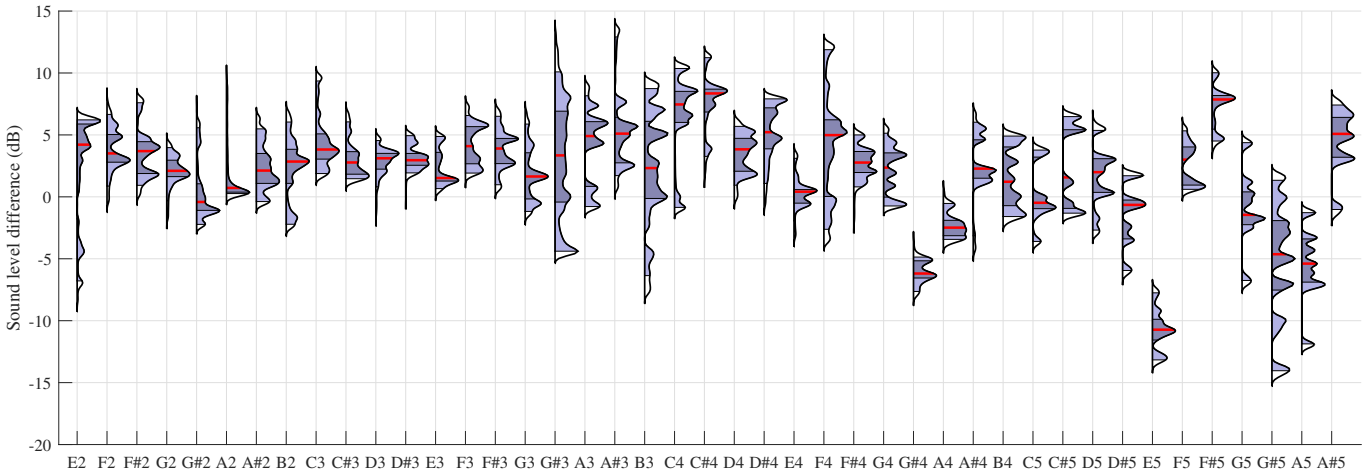

Fig. 12: PDFs of the sound level differences $\boldsymbol{\lambda}$. Acoustic guitar modern fortissimo. Dark shaded zone: interquartile range; light shaded zone: 5%–95% range; red line: median.

dataset in a hall at 8 source-receiver combinations. About 50% of these data show sound level differences $\boldsymbol{\lambda}$ close to ± 5 dB and loudness excess ratio $\boldsymbol{\eta}$ close to $\pm 35\%$. Both values reach about $\pm 10\,\text{dB}$ and $[-50,\ 75]\%$, respectively, for the 5%–95% range of the studied data. The uncertainty is not small; however, put in context, several of the largest differences may not be perceived. The upper bound of $\boldsymbol{\lambda}$ is barely exceeded for some instruments and the whole interval is biased for the modern transverse flute, the timpani, and the modern viola (i.e. the effect is within a [-19, -2] dB interval for the modern viola). Considering loudness, the transverse flute modern exhibits a biased PDF and the violin pianissimo can be considered an outlier.

The anechoic signal of the violin pianissimo was immersed in background noise. The estimations of partials' directivity may still be robust against broadband noise. However, the loudness calculation is affected by the noise energy at other frequencies than the partials and, thus, considered an outlier.

The source-receiver combination has an effect on the simulated sound level differences. The multimodal behavior of the distributions grouped by note likely occur due to the source-receiver combinations and is evident for some notes.

The results constitute a pilot comprehensive study on the uncertainty of geometrical acoustics simulations due to band averaging the directivity of musical instruments. The dynamic loudness model [20] is a model and a fast algorithm which providing a good estimation of loudness based on monaural signals. However, binaural loudness models may be worthwhile for the investigation of receivers' directivity.

It is not known yet if these uncertainties are audible in a dynamic scene with running music. Furthermore, the audibility may depend on the number and classification of instruments playing together (e.g. solo, an orchestra section, or a tutti), as well as other aspects of the repertoire.

Musical instruments of different kind clearly modify the expected uncertainty, but it remains unknown the effect of differences due to player, instrument construction techniques, instrument materials, or units of the same instrument. Although the results are strictly valid for the simulated hall, the conclusions can be extended to other halls with similar characteristics.

As a consequence of these results, the authors will continue with listening tests in order to investigate just noticeable differences (JND). The purpose of these future studies is to yield to a resolution in frequency bandwidth and angular sampling that would be required for staying below JNDs.

VII. Conclusions

A first conclusion is that directional factors based on averaging directional index $\Gamma_{\overline{L}_i}$ is recommended for building directivity datasets in fractional octave bands. A second outcome, in accordance with previous subjective studies and based on objective estimates, is that uncertainties due to third octave bands is likely perceivable depending on the repertoire. Moreover, the main contribution of this paper is to provide a first piece of knowledge on the uncertainty of the sound level and loudness in a hall due to band averaging of directional indices. Both the directional index as input data summarized in Fig. 3 and the sound level in a hall summarized in Fig. 10 and the supplementary material constitute a guideline for estimating the uncertainty of input data and its propagation in simulations and auralizations.

References

[1] N. R. Shabtai, G. Behler, M. Vorländer, and S. Weinzierl, "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 1246–1256, 2017.

[2] F. Hohl, *Kugelmikrofonarray zur Abstrahlungsvermessung von Musikinstrumenten (Spherical microphone array for capturing soundradiation from musical instruments),*. Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria, 2009. Master's thesis.

[3] F. Hohl and F. Zotter, "Similarity of musical instrument radiation-patterns in pitch and partial," in *Fortschritte der Akustik, DAGA*, 01 2010.

[4] F. Otondo and J. Rindel, "The influence of the directivity of musical instruments in a room," *Acta Acustica united with Acustica*, vol. 90, no. 6, pp. 1178–1184, 2004.

[5] C. Pörschmann and J. M. Arend, "A method for spatial upsampling of voice directivity by directional equalization," *Journal of the Audio Engineering Society*, vol. 68, no. 9, pp. 649–663, 2020.

[6] A. Quélennec and P. Luizard, "Pilot study on the influence of spatial resolution of human voice directivity on speech perception," *Acta Acustica*, vol. 6, p. 10, 2022.

[7] J. Ahrens and S. Bilbao, "Computation of spherical harmonic representations of source directivity based on the finite-distance signature," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 29, pp. 83–92, 2021.

[8] J. Meyer, *Acoustics and the Performance of Music*. Springer, 1st ed., 2009.

[9] S. Weinzierl, M. Vorländer, G. Behler, F. Brinkmann, H. V. Coler, E. Detzner, J. Krämer, A. Lindau, M. Pollow, F. Schulz, and N. R. Shabtai, *A Database of Anechoic Microphone Array Measurements of Musical Instruments*. Technische Universität Berlin. DepositOnce, 2017.

[10] I. B. Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of source measured by surrounding spherical microphone arrays," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2003–2015, 2011.

[11] N. R. Shabtai and M. Vorländer, "Acoustic centering of sources with high-order radiation patterns," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 1947–1961, 2015.

[12] Blauert, J. and Xiang,N., *Acoustics for Engineers*. Springer Berlin Heidelberg, 2008.

[13] H. Kuttruff, *Room acoustics*. CRC Press, sixth edition ed., 2016.

[14] P. Leopardi, "A partition of the unit sphere into regions of equal area and small diameter," *Electronic Transactions on Numerical Analysis*, vol. 25, pp. 309–327, 2006.

[15] D. Schröder and M. Vorländer, "RAVEN: A real-time framework for the auralization of interactive virtual environments," in *Forum Acusticum*, pp. 1541–1546, Aalborg Denmark, 2011.

[16] F. Miyara, *Software-based acoustical measurements*. Modern acoustics and signal processing, Springer, 2017.

[17] E. Accolti and F. di Sciascio, "On the use of time windows for the determination of sound strength parameter G from uncalibrated room impulse responses measurements," *Applied Acoustics*, vol. 178, p. 12, 2021.

[18] T. Hidaka and N. Nishihara, "Favorable reverberation time in concert halls revisited for piano and violin solos," *The Journal of the Acoustical Society of America*, vol. 151, no. 3, pp. 2192–2206, 2022.

[19] M. Vorländer, *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, 1st ed., 2008.

[20] J. Chalupper and H. Fastl, "Dynamic loudness model (DLM) for normal and hearing-impaired listeners," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 378–386, 2002.

Ernesto Accolti received the engineering degree in acoustics from the Technical University of Chile, in 2006, and the doctoral degree from the National University of Rosario, Argentina, in 2015. Since 2015 he is with the Institute of Automation (INAUT) which depends on the National University of San Juan (UNSJ), and the National Council of Scientific and Technical Research (CONICET), Argentina. He is Professor at the INAUT within the UNSJ, while also serving as Researcher for CONICET. His research interest is acoustics of music performance spaces. He is a member of the board of the Argentine Association of Acoustics (AdAA).

Javier Gimenez graduated in mathematics from the National University of San Juan (UNSJ), Argentina, in 2009; and obtained a Ph.D. in mathematics from the National University of Córdoba, Argentina, in 2014. He is currently an Adjunct Researcher of CONICET with the Institute of Automation, UNSJ-CONICET, and a full professor with the Department of Mathematics, Faculty of Engineering, UNSJ, Argentina. His research interests include SLAM algorithms in robotics and general mathematical applications in engineering.

Michael Vorländer graduated in physics in 1984, gained a doctor degree in 1989 at RWTH Aachen University, Germany, and a habilitation degree at Technical University Dresden, Germany, in 1995. He is Professor at RWTH Aachen University, Germany, for Technical Acoustics since 1996. His book "Auralization" (Berlin, Germany: 2nd ed., Springer 2020) is a reference on the field of Acoustic Virtual Reality. His current research interest is auralization including simulation techniques and signal processing. He was president of the European Acoustics Association, EAA, in the term 2004–2006 and president of the International Commission for Acoustics, ICA, in the term 2010–2013. He is Fellow of the Acoustic Society of America and was awarded with the Rayleigh Medal (IoA UK) in 2021.