SURFACE SEGMENTATION FOR ACOUSTIC RADIANCE TRANSFER MODELING: A CASE STUDY

Matteo Scerbo[1,2], Per Hiselius[2], Cheol-Ho Jeong[1]
[1] *DTU Electro, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*
[2] *Ecophon, Saint-Gobain Ecophon AB, Box 500, SE-26503 Hyllinge, Sweden*

Abstract

In the context of acoustic radiance transfer, one piece of advice is frequently heard: reflecting surfaces should be segmented in patches of approximately equal area (uniformity). Another often-repeated suggestion is that individual surface patches should be compact, i.e., one should avoid using patches which are long but narrow (compactness). While many works in the literature discuss the aspects of patch size and number, none touch on these two recommendations, let alone the possible repercussions of their violation. It is therefore possible that abiding by these guidelines imposes unnecessary constraints, making models less efficient than they otherwise could be. This paper describes a case study which challenges such conventional assumptions, using environments from the benchmark for room acoustical simulation (BRAS) dataset. Each environment is simulated several times using different criteria for surface segmentation, including ones which deliberately infringe the aforementioned recommendations. Comparing the simulation results to ground-truth recordings shows that the performance of acoustic radiance transfer is not meaningfully affected by either of the challenged guidelines (uniformity and compactness). All of the tested surface meshes are available online, alongside the code to generate and simulate them.

Keywords: *room acoustic modeling, acoustic radiance transfer, polygon mesh processing*

1. Introduction

To predict room acoustics, the acoustic radiance transfer (ART) model relies on the segmentation of room boundaries into discrete surface patches [1]. It is well known that the level of discretization is the most important factor in the trade-off between ART's accuracy and efficiency: smaller patches lead to more reliable results, but they bear steep complexity scaling [2], [3]. Beyond this basic fact, no information is found in the literature regarding different choices of surface segmentation and their effects on simulation results. Given the number of surface patches' strong influence on the runtime of ART simulations, it would be beneficial to know which aspects of the segmentation approach do or don't affect simulation results in meaningful ways. This paper aims to present a preliminary study in this regard.

As a starting point for the investigation, we consider two guidelines which are often suggested by experts in the field. The first one is that all patches in the segmentation should ideally have similar surface areas. The second is that no one patch should be long and narrow; all patches should be compact. These recommendations were likely inherited from the fields of heat exchange and computer graphics, where the ART algorithm originated [1]. In those contexts, the first few reflections play a much more prominent role in the overall spatial distribution of energy. To assess the relevance of said criteria in the context of acoustics, ART models of several environments from the benchmark for room acoustical simulation (BRAS) dataset [4] were generated and tested using different surface segmentation approaches. The tested meshes and simulation code can be found at the DOI 10.5281/zenodo.21630205.

The remainder of this paper is structured as follows. Section 2 details the preparation of 3D meshes for simulation, including the different surface segmentation strategies under exam. Section 3 details the approach by which measured and simulated responses were processed and compared. Section 4 illustrates and discusses the comparison results. Section 5 concludes the paper.

2. Mesh preparation

The BRAS dataset includes four reverberant environments of increasing geometric complexity. They are denoted CR1 (coupled laboratory & reverberation chamber), CR2 (small seminar room), CR3 (medium music hall), and CR4 (large auditorium). Two sets

Corresponding author: *mattes@dtu.dk.*

of recordings from CR1 are provided: one with the connecting door mostly closed, one mostly open.

The 3D models of all environments are provided in the form of SketchUp models of substantial detail. In the context of ART, the high level of detail led to the unavoidable presence of some extremely small surface patches in each environment. In order to test the impact of surface segmentation strategies in a wider variety of meshing conditions, simplified versions of the CR1 and CR2 environment meshes were produced for this case study. These simplified meshes were generated manually by removing geometric details deemed inessential for late reverberation modeling. No manual simplification of CR3 and CR4 was conducted due to their much higher geometrical complexity. Conducting an automatic simplification of the meshes would require an algorithm designed specifically for acoustic models [5], and is left for future works. With a simplified mesh being unavailable, the CR4 environment was excluded from this case study, because its exceedingly high number of polygons would have made the "naive" simulations impractically time-consuming.

For each environment mesh, four different ART models were prepared using four different surface segmentation approaches. These were:

α) *Largest possible patches*: all coplanar polygons sharing at least one edge were conjoined, e.g. the entire floor of each room is a single patch.

β) *Naive triangulation*: triangles exported directly from the SketchUp models were maintained as individual surface patches; very high number of long-but-narrow triangles of disparate sizes.

γ) *Target area* $4\,\mathrm{m}^2$: starting from "largest possible", large patches were split into sections of $\sim 4\,\mathrm{m}^2$.

δ) *Target area* $2\,\mathrm{m}^2$: same as the previous approach, but with a lower threshold.

In total, 28 ART models were generated and run. Table 1 reports their sizes. Note that the number of surface patches alone is not indicative of an ART model's complexity. Since the "working unit" of ART is the acoustic energy traversing each possible propagation path (from patch to patch), the "size" of the model is the number of unobstructed propagation paths. Moreover, at the heart of each iteration step in the ART simulation is the scattering matrix, which characterizes the energy transfer from each path to each adjacent path; the computation and memory requirements of running the ART model are proportional to the number of nonzero elements in this matrix. In the worst-case scenario (full visibility), the number of propagation paths scales with the square of the number of patches, and the number of nonzero matrix elements scales with its cube [6].

Because of this unfavorable complexity scaling, geometry simplification and surface segmentation approaches have severe effects on simulation runtime. For example, looking at Table 1, one can see that modeling the CR2 environment using a simplified mesh and rough segmentation (α) is almost one thousand times faster than using the original mesh and a fine segmentation (δ) – that is a ratio of ten minutes to one week. As shown later in Section 4, the simulation results of these models are nearly identical.

Table 1: For each tested mesh: number of surface patches, number of unobstructed propagation paths, number of nonzero elements in the scattering matrix. The segmentation strategies are: α) largest possible patches; β) naive triangulation; γ) target area $4\,\mathrm{m}^2$; δ) target area $2\,\mathrm{m}^2$.

Room	Mesh	Seg.	Patches	Paths	Nonzeros
CR1, closed	base mesh	α	90	1 926	60 764
		β	288	16 404	1 416 596
		γ	156	7 384	413 220
		δ	245	18 860	1 622 168
	simpl. mesh	α	54	1 158	29 820
		β	168	10 620	787 772
		γ	117	5 698	297 546
		δ	201	15 774	1 301 078
CR1, open	base mesh	α	97	2 126	70 830
		β	323	18 910	1 759 414
		γ	163	8 192	503 316
		δ	251	20 344	1 897 854
	simpl. mesh	α	54	1 186	31 252
		β	168	11 014	846 490
		γ	117	5 948	324 840
		δ	201	16 452	1 424 078
CR2	base mesh	α	155	4 206	203 748
		β	504	34 516	4 652 856
		γ	193	10 764	881 594
		δ	235	20 008	2 292 130
	simpl. mesh	α	21	246	3 248
		β	69	2 456	99 492
		γ	59	2 504	111 406
		δ	104	8 072	647 088
CR3	base mesh	α	726	56 436	8 667 592
		β	2 461	410 212	164 056 166
		γ	1 285	289 468	94 460 866
		δ	1 935	686 442	327 116 136

Fig. 1 shows the distributions of surface patch area and compactness in each of the 28 tested models. As a metric for compactness, we use the isoperimetric quotient, defined as the ratio between a patch's area and the area of the circle having the same perimeter as the patch. The value of this quotient is bound to the range (0, 1], with 1 denoting utmost compactness (a perfect circle) and values close to 0 denoting very long and narrow patches.

Note that the surface area axis is truncated: the lowest area values are not shown, for readability. All original meshes, as well as the naive triangulations of simplified meshes, feature a small number of patches with areas as small as $10^{-9}\,\mathrm{m}^2$. These outliers stretch the extrema of the illustrated distributions beyond the shown range.

By design, the *target area* approaches result in patch area distributions centered around their respective targets, and the *largest possible patches* approach achieves its namesake. The *naive triangulation* approach results in the widest spread of area values, even in the simplified mesh cases, because it is wholly unconstrained; tiny triangles often appear from the triangulation of complex polygonal elements. Note that, in all non-simplified meshes, there are always at least some patches with very small areas. These are minute geometric details which

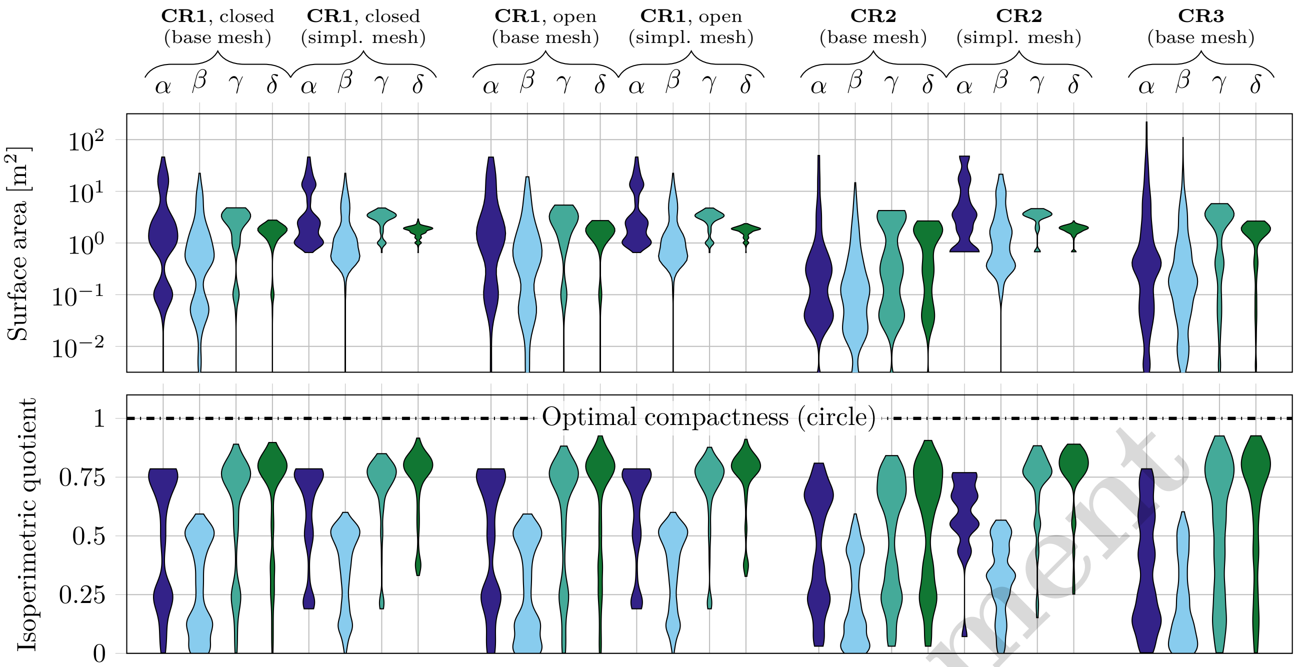

Figure 1: Distribution of the surface area and isoperimetric quotient (i.e., compactness) of surface patches in each simulated environment. These can be read like histograms: each violin shape shows the distribution of surface patch properties in a different mesh. Greek letters denote segmentation approaches as in Table 1.

cannot be merged into larger patches without altering the geometry. In particular, the CR2 environment exhibits a dominant portion of small patches under all segmentation strategies, due to a large number of small details like ceiling lights, recessed windows, and small protrusions in the corners of the room. All such details were removed in the simplified mesh.

In terms of compactness, the *naive triangulation* approach performs worst of all, as expected, because triangulation algorithms intended for graphics generally produce long, thin triangles. The *target area* approaches result in mostly compact surface patches, by design.

3. Comparison approach

The BRAS dataset provides multiple room impulse responses (RIRs)[1] for each environment, recorded with different arrangements of loudspeaker and microphone. For each loudspeaker position, five recordings are provided: one made with an omnidirectional "dodecahedron" loudspeaker, and four made with Genelec loudspeakers in different axial orientations. Since the ART simulations considered an ideal omnidirectional point source, the dodecahedron recordings were taken as the primary reference – but the Genelec recordings were also considered for comparison.

All dodecahedron recordings exhibit an energy trough in the 2 kHz octave band, disagreeing with the Genelec recordings as well as the simulation results. We hypothesize this is due to destructive interference between the "mid" and "high" loudspeaker units, which have nearly opposite phase responses at the crossover frequency. For all comparisons, the dodecahedron recordings' energy in the affected band was multiplied by a factor of four, which consistently brought them in agreement with other recordings' frequency profiles. The Genelec recordings are included in the comparisons to illustrate this, among other reasons.

Fig. 2 illustrates the recorded and simulated energy for one loudspeaker-microphone configuration of CR1. The top plot shows the energy responses directly, while the bottom plot shows their difference from the dodecahedron recording.

While ART simulations directly produce smoothly-decaying energy responses, the recorded RIRs are in the pressure domain – their energy counterparts are obtained through squaring and applying a 100 ms smoothing window (running average). Note that, despite the smoothing, recorded energy responses naturally appear "noisy", more so in lower frequency bands. Moreover, all recordings exhibit a noise floor, whereas simulations do not. The noise floor level varies for different frequencies, loudspeakers, etc. Lastly, some of the more reverberant recordings exhibit a truncation "roll-off" artifact before the noise floor is reached. The cause of this appears to be that the recordings were truncated *before* carrying out the sine-sweep deconvolution, such that the deconvolution "smeared" the truncation across frequencies.

Due to the combination of noise floors and truncation artifacts, a comparison of conventional energy decay curves (EDCs) would be unsound: the incongruous late energy behavior would skew the backwards integration. Performing backwards integration after truncating

[1] binaural room impulse responses (BRIRs) are also provided in the dataset, but not used in this study.

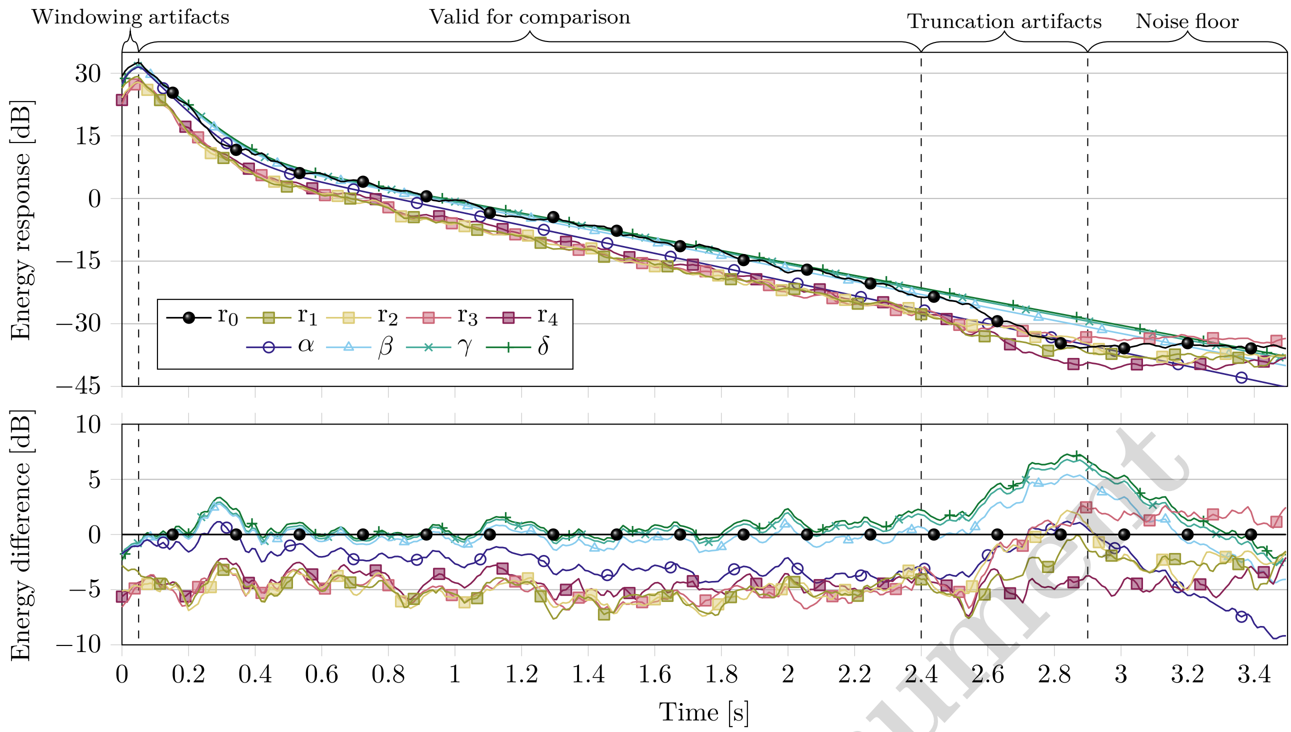

Figure 2: Recorded and modeled energy in the 1 kHz octave band for the CR1 environment (open door), with loudspeaker LS2 and microphone MP4. The dodecahedron references is denoted r_0, Genelec references are denoted $r_1 - r_4$. Greek letters denote segmentation approaches as in Table 1; simulations shown here used the original (un-simplified) mesh. The top plot shows energy responses directly, the bottom plot shows their difference from r_0.

all responses to the valid ranges would also be uninformative and/or misleading, because the late part of the EDC would present a roll-off towards $-\infty$ dB, throwing off the comparison results. As such, the results presented in the next section are based on direct comparisons of the energy responses as shown in Fig. 2 rather than on EDCs. The results were produced by comparing simulation results against reference (dodecahedron) recordings, strictly within the valid range: from 50 ms to the onset of truncation artifacts or noise floor, whichever comes first. The first 50 ms were excluded from comparisons for two reasons. First, they contain edge artifacts due to convolution with the 100 ms smoothing window. Second, this is generally considered as the threshold for early reflections; our comparison is restricted to late reverberation for simplicity.

4. Comparison results

Fig. 3 shows the results of the comparisons described in the previous section. Plots are divided by octave band, from 125 Hz to 4 kHz. These plots are to be interpreted as statistical distribution of the energy difference between each model (or Genelec recording) and the primary reference, i.e., the dodecahedron recordings – in other words, a histogram of the values within the valid range of Fig. 2's bottom plot. The distributions of energy difference levels are aggregated across all loudspeaker and microphone positions, such that each violin shape in Fig. 3 shows the range and distribution of "errors" for a given approach, in a given environment. For example, consider the case of the CR1 environment (open door) in the 1 kHz octave band – this is the case which was shown in Fig. 2. In this example, the Genelec recordings exhibit an overall energy offset of ∼5 dB below the Genelec recording, but they have the same rate of decay – this appears in Fig. 3 as narrow violins centered around −5 dB. The simulation results shown in Fig. 2 are for the original (un-simplified) mesh. The "largest patches possible" approach (α) has the correct starting energy, but over-estimates the decay rate – this appears in Fig. 3 as a wider violin with the upper tip around 0 dB, and the lower tip around −4 dB. The other segmentation approaches all have the right energy and decay rate – this appears in Fig. 3 as narrow violins centered around 0 dB.

4.1. Overall observations

In the lower octave bands, energy differences show wider spreads in all cases (for both simulations and recordings) due to energy fluctuations in the band-passed reference. In the higher octave bands, the difference margins tighten for reference recordings, while simulations spread towards an excess of energy due to an underestimation of air absorption effects. This is a known, inevitable issue of energy-based acoustic models, which must discretize the air absorption coefficient to one value per frequency band [7]. In our simulation, the chosen value is calculated with respect to the lower edge of each band, therefore its effect is underestimated for the

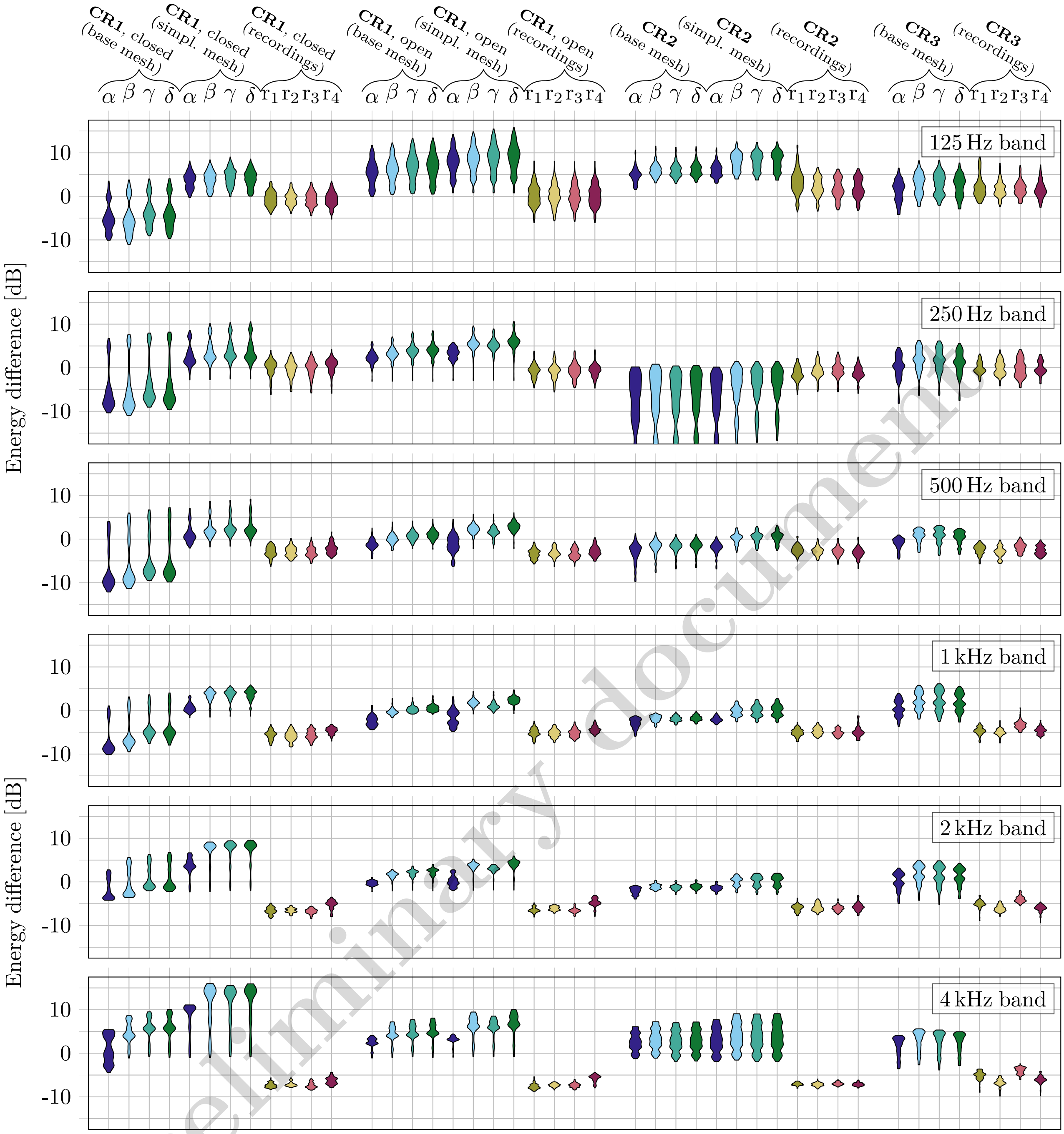

Figure 3: Comparison results in octave bands from 125 Hz to 4 kHz. Greek letters denote segmentation approaches as in Table 1, while $r_1 - r_4$ denote recordings from Genelec loudspeakers in different orientations.

higher frequencies in the same band. The issue aggravates with rising frequency; comparison results in the 8 and 16 kHz bands are dominated by the effect of this underestimation, and are left out for reasons of space.

4.2. Simulation result observations

The main takeaway from the comparison of simulation results is that the difference between most segmentation methods is negligible. The "largest possible patches" approach is the only one that stands out, more so in the case of simplified meshes. This is most likely due to the simplified meshes feature extremely large surfaces which span almost the full environment. In all cases, this approach is the most approximate, and generally provides less reliable predictions (wider energy difference spreads) as should be expected [2], [3]. The "naive triangulation" approach does not perform significantly worse than the others, despite the large proportion of non-compact patches. This is the case in all tested environments, in all frequency bands, strongly suggesting that surface patch uniformity and compactness have no significant effect on ART simulation results. The two “target area” approaches also show negligible difference between them, despite the number of patches being roughly doubled from one to the other (which severely impacts runtime).

In the CR2 environment, as well as the CR1 case with the open door, mesh simplification compromises results only slightly. In the CR1 case with the mostly closed door, simplification even seems to improve results, both in terms of overall energy levels (violin centroid) and decay rate reliability (violin range). Considering that these results were achieved with a manual, naive remeshing, we believe that an automated acoustics-aware algorithm would prevent any disruption of predictive capability.

Looking at all simulation results as a whole, we can notice a few weaknesses of ART independent of meshing and segmentation approaches. In the case of the CR1 environment with the door mostly closed, none of the simulations perform well. We believe that, because the opening is so small, a bespoke segmentation approach would be required to model the energy transfer reliably. Future works may investigate this aspect. In the CR2 room, all simulations in the 250 Hz band are severely inaccurate. This comes as no surprise, since the room was included in the dataset due to its “challenging low frequency behavior” [4] – which an energy-based model like ART cannot simulate. In CR3, simulations overestimate reverberation times for bands 1 kHz and up, possibly due to a stronger influence of air absorption in this larger space.

4.3. Genelec reference observations

The Genelec recordings are in agreement with each other and with the dodecahedron reference, with a few exceptions. The most prominent difference is that the Genelec recordings’ total energy diminishes (compared to the dodecahedron counterparts) with increasing frequency. We hypothesize that this is due to the input/output chains being calibrated to a *free field* sound pressure *in front of* the loudspeakers; since the Genelecs have significant directivity at higher frequencies, they radiated less energy overall, leading to lower *diffuse field* sound pressure. In some of the environments, there are also small disagreements among the four Genelec recordings. In CR1, one Genelec reference stands out from the other three, showing slightly higher energy at high frequencies. The orientation in question faces directly towards the open doorway, radiating more energy towards the microphone(s). On the other hand, in CR2, one Genelec reference stands out for its slightly higher energy in the 125 Hz band. This time, the orientation in question seems to provide greater stimulation to a low-frequency room mode. Finally, in CR3, the Genelec references show a more marked distinction and wider spreads – the energy responses in this complex environment exhibit large positional and directional differences well into late reverberation, up to 150 ms into the response.

5. Conclusions

We have presented a preliminary study regarding different aspects of surface segmentation in the context of acoustic radiance transfer (ART) simulations. Comparisons against measured impulse responses indicate that neither uniformity nor compactness of the surface patches have meaningful influence on simulation results. While the tested environments are few, they cover a wide range of real-world conditions both in terms of acoustics (absorption, diffusion, positional variability) and in terms of input mesh properties (polygon count, level of detail, range of patch sizes). We have also showed that the simplification of environment geometry, even conducted relatively naively, has minimal detrimental effects on modeling accuracy – while reducing runtime by orders of magnitude. The latter observation underscores the importance of developing an algorithm for automated, acoustics-aware mesh simplification; the former observation shows that such an algorithm would not need to comply with strict requirements for patch design.

References

[1] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, “The room acoustic rendering equation,” *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1624–1635, 2007.

[2] E. Nosal, M. Hodgson, and I. Ashdown, “Investigation of the validity of radiosity for sound-field prediction in cubic rooms,” *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3505–3514, 2004.

[3] R. T. Muehleisen and C. W. Beamer, “Steady state acoustic radiosity for the prediction of sound pressure levels in enclosures with diffuse surfaces,” *Noise Control Engineering Journal*, vol. 57, no. 3, pp. 244–262, 2009.

[4] L. Aspöck, M. Vorländer, F. Brinkmann, D. Ackermann, and S. Weinzierl, *BRAS – benchmark for room acoustical simulation*, 2020. DOI: 10.14279/depositonce-6726.3

[5] S. Siltanen, T. Lokki, L. Savioja, and C. Lynge Christensen, “Geometry reduction in room acoustics modeling,” *Acta Acustica united with Acustica*, vol. 94, no. 3, pp. 410–418, 2008.

[6] M. Scerbo, S. J. Schlecht, R. Ali, L. Savioja, and E. De Sena, “Modeling nonuniform energy decay through the modal decomposition of acoustic radiance transfer (MoD-ART),” *IEEE Trans. Audio, Speech, Language Proc.*, vol. 33, pp. 3363–3376, 2025.

[7] R. H. C. Wenmaekers, C. C. J. M. Hak, and M. C. J. Hornikx, “The effective air absorption coefficient for predicting reverberation time in full octave bands,” *J. Acoust. Soc. Amer.*, vol. 136, no. 6, pp. 3063–3071, 2014.