Adaptive Geometric Sound Propagation Based on A-weighting Variance Measure

Hongyang Zhou* 3130000714@zju.edu.cn
Zhong Ren* renzhong@zju.edu.cn
Kun Zhou* kunzhou@acm.org

Abstract

We introduce an A-weighting variance measurement, an objective estimation of the sound quality generated by geometric acoustic methods. Unlike the previous measurement which applies to the impulse response, our measurement establishes the relationship between the impulse response and the auralized sound that the user actually hears. We also develop interactive methods to evaluate the measurement at run time and an adaptive algorithm that balances quality and performance based on the measurement. Experiments show that our method is more efficient in a wide variety of scene geometry, input sound, reverberation, and path tracing strategies.

1. Introduction

In VR applications, the ability to simulate the sound that user hear in virtual environment with high fidelity is of great importance for the immersive experience. Over the years, different methods for simulating sound propagation in virtual environment have been proposed, which can be essentially divided into two categories: *wave-based* methods and *geometric acoustic (GA)* methods [25, 26]. In VR applications of which interaction being the key feature, GA methods are the dominating choice thanks to its performance advantage over wave-based methods.

Most GA methods work by tracing random paths that link the listener and the sound sources, from which a *impulse response (IR)* can be obtained, and then convolved with the actual sound wave of the sources to yield the sound that user hear. The number of paths, or samples, generated in the tracing process largely determines the quality of the simulation. In applications it is often desirable to determine a proper sampling rate according to a predefined quality requirement. In a recent work [10] that uses bidirectional path tracing to simulate sound propagation, a measurement based on the *signal-to-noise ratio (SNR)* of impulse response is proposed and used to control the sampling rate. This measurement, however, is computed on the impulse response instead of the sound that user actually hear. Controlling the sampling rate based on an indirect measurement give rise to undesired behavior in certain input sound – for a given impulse response the quality of the simulation result that heard by the user is obviously also dependent on the input sound wave, which is ignored in the measurement.

In this paper, we propose a new measurement for the simulation quality that directly applies on the auralized sound. We apply the widely used *A-weighting* technique [8] to accumulate variance estimation of different frequency band to yield a single-valued measurement of the quality, which can be easily used to develop adaptive sampling strategies. The main contributions of the paper include:

- A new, A-weighting based measurement that directly applies on the auralized sound for the simulation quality;
- An interactive method for estimating the above measurement at run time;
- An adaptive method based on the new measurement to balance quality and performance by dynamically determining a minimal sample budget that satisfies the predefined quality threshold.

We show that our method is more efficient in balancing quality and performance than previous method in different scene geometry, input sound wave, reverberation and path tracing strategies.

2. Background

In this section, we briefly describe geometric acoustic methods in sound propagation, and existing measures on sound quality with their limitations.

2.1. GA Sound Propagation

Many widely used methods for sound propagation are based on GA. GA methods work under the assumption that the wavelength of sound is much smaller than the objects in the scene. Under this assumption, sound propagates

*State Key Lab of CAD&CG, Zhejiang University. 866, Yuhangtang Rd. Hangzhou. China.

along straight lines, and methods used for light transportation are fit for sound propagation. Although GA methods are hard to simulate wave effects such as diffraction and interference, these methods are still more effective than wave-based methods in room acoustics with high-frequency sources where wave effects take a minor position.

Early works in GA sound propagation include image source method and ray tracing [18]. Similar to that for visual rendering, ray tracing based methods are widely used in GA methods for sound rendering [19, 21]. Paths of sound particles move through the scene and hit objects then are accumulated at the position of the listener. A bidirectional path tracing GA method was introduced in [10] using the method in visual rendering [16].

Although path-tracing algorithms in rendering the sound are similar to that in rendering an image, there are still some key differences between these two cases. The auditory system of human ears has lower spatial resolution but higher temporal resolution than the visual system. Quality measures for sounds should be focus on the temporal dimension instead of the spatial distance. Because of these difficulties, few measures are introduced to evaluate the quality of GA generated sound. [10] introduced a measure for GA algorithms based on SNR on energy response, but this measure is still not for the sound that the users would hear.

2.2. Sound Measures

Existing sound measures are mostly applied to describe the characteristics of a sound signal. Commonly used sound measures include some basic properties like sound pressure and frequency [1]. Some other measures describe how the sound propagates in an enclosed space in room acoustics, including reverberation time RT_{60}, clarity C_{50}, and early decay time EDT [5].

RT_{60} and C_{50} are measures defined on IR. They are defined by $RT_{60} = t_{-60dB} - t_0$, the time sound pressure level takes to decrease by 60dB, while $C_{50} = 10log_{10}(E_{0,50}/E_{50,\inf})$, the ratio of early sound energy in first 50ms to the late sound energy. These measures give some brief information on the shape and property of IR. However, they are not enough to evaluate the quality of path-tracing generated impulse response because the expected shape of IR does not change when the sample budget changes.

Subjective sound measures based on psychoacoustics play a different role on sound evaluation. Mean opinion score terminology [7] is widely used in the perceptual evaluation of subjective sound quality. But this method is case-specific and hard to be conducted when when the number of virtual scenes is large while the number of test subjects is limited. Another way of building subjective sound measures is to conclude generic rules with the support of physiological or cognitive experiments, such as loudness [6] and localization [9] cues.

3. Analysis

3.1. Error in path tracing based GA methods

Deterministic GA methods like image source methods are precise but not accurate since they ignore late reverberation in sound propagation. Here we discuss path-tracing based GA methods. The impulse response is generated by 1) collecting random paths propagated in the scene and 2) filtering the collected result.

We assume all the ray path in one frame is generated in a same sampling strategy without loss of generality. Collected signal $\sum e_i(t)$ in the first (collecting) step is written as $E(t)$. It is clear that a convolution filter on $E(t)$ can be independently applied on each ray path:

$$\begin{aligned} E'(t) &= f(t) * E(t) = f(t) * \frac{1}{N}\sum_{i=1}^{N} e_i(t) \\ &= \frac{1}{N}\sum_{i=1}^{N} f(t) * e_i(t) \\ &= \frac{1}{N}\sum_{i=1}^{N} e'_i(t) \end{aligned} \tag{1}$$

Here $*$ is the convolution operator, and $f(t)$ is a filter on original energy response. The impulse response we are interested in $IR(t)$ is a case of filtered energy $E'(t)$. Almost all common operators between the path tracing step and the auralization step can be represented as $f(t)$ in Monte Carlo based methods, including:

- Scaling energy with a constant factor.
- Shifting signal in the time domain.
- Band-pass or crossover filters.

This shows that the expectation of impulse response $IR(t)$ is the mean value of many filtered path impulse $e'_i(t)$. Each $e'_i(t)$ is a random function defined on time domain. Changing the scene, position of sound source and position of listener also changes the distribution of each $e'_i(t)$, hence the estimated impulse response becomes different. Increasing rays from N to N' changes the variance of estimated IR inverse proportionally:

$$\begin{aligned} var(IR') &= var(\frac{1}{N'}\sum_{i=1}^{N'} e'_i(t)) \\ &= \frac{1}{N'}var(e'(t)) = \frac{N}{N'}var(IR) \\ or: var(IR)\cdot N &= var(IR')\cdot N' \end{aligned} \tag{2}$$

This can also be written in the form of SNR if needed, because mean value will not change:

$$\begin{aligned} SNR(IR') &= \frac{mean^2(IR')}{var(IR')} \\ &= \frac{N'}{N}\frac{mean^2(IR)}{var(IR)} \\ &= \frac{N'}{N}SNR(IR) \end{aligned} \tag{3}$$

These results also holds on frequency domain because:

$$\begin{aligned} E'_{freq}(s) &= FT(f(t) * E(t)) = f_{freq}(s) \cdot FT(\frac{1}{N}\sum_{i=1}^{N} e_i(t)) \\ &= f_{freq}(s) \cdot \frac{1}{N}\sum_{i=1}^{N} e_{i,freq}(s) \\ &= \frac{1}{N}\sum_{i=1}^{N} e'_{i,freq}(s) \end{aligned} \tag{4}$$

Here $FT(f(x))$ is the Fourier transform to function $f(x)$, and the functions with subscript $freq$ are defined on frequency domain.

3.2. From IR to convoluted sound

In auralization step of sound propagation, IR is convoluted with input signal to get the sound human user finally hear. In this calculation, variance on IR becomes error on convoluted sound. For a sound clip, it is hard to analyse convolution error in time domain. But the relationship between two errors is more clear in frequency domain:

$$\begin{aligned} var(T_{freq}) &= var(S_{freq} \cdot IR_{freq}) \\ &= S^2_{freq} var(IR_{freq}) \end{aligned} \tag{5}$$

Where T is the output sound and S is the input signal. This implies that the variance of output sound on the frequency domain is determined by the frequency amplitude of input sound and the variance of IR. In sound propagation, S_{freq} is the pre-defined input sound signal and always keeps constant. So the variance of output signal $var(T_{freq})$ only changes with the variance of IR $var(IR_{freq})$, which is further determined by the change of sample count N.

However, in practical, result sound clips on each frame are not independent. An overlap exists between two sound clips of the length of L_{IR} or ($L_{IR} - 1$ in discrete form). But it is shown that the error of overlap is controlled on the aspect of IR energy distribution:

$$\begin{aligned} E_{clip} &= E_{tail,prev} + E_{head} \\ &= \int_{L_S}^{L_S+L_{IR}} (S_{prev} * IR)^2 dt + \int_{0}^{L_S} (S * IR)^2 dt \\ &= \int_{0}^{L_S+L_{IR}} (S * IR)^2 dt \\ &- \int_{L_s}^{L_S+L_{IR}} ((S * IR)^2 - (S_{prev} * IR)^2) dt \end{aligned} \tag{6}$$

Meanwhile,

$$\begin{aligned} &\left| \int_{L_s}^{L_S+L_{IR}} ((S * IR)^2 - (S_{prev} * IR)^2) dt \right| \\ &\leq \left| \int_{L_s}^{L_S+L_{IR}} (S * IR)^2 dt \right| \\ &+ \left| \int_{L_s}^{L_S+L_{IR}} (S_{prev} * IR)^2 dt \right| \end{aligned} \tag{7}$$

Here "clip" means the output energy on a time clip of length L_{clip}, "head" means the energy of the beginning of signal with length L_S, "tail" means the energy of ending of signal with length L_{IR}, and "prev" is the last clip or block of the input sound. This formula shows that the error from overlap can be ignored if L_{IR} is far smaller than L_S or the reverberation of IR is relatively small that IR can be approximately equivalent to a shorter one. From Parseval's theorem, the energy on the time domain is equal to that on the frequency domain. This means the overlap error on frequency can also be ignored. Experiments are described in the following chapter confirming that overlap error does not take an important part in common cases.

3.3. Get a single variance value

The variances analyzed in previous sections are functions on time or frequency domain. We use the A-weighting technique described in [8] to obtain a single variance value for further application. A-weighting curve is shown in Fig. 1, describing the relative loudness of different frequencies perceived by human ears. To make different frequencies to be equal-loudness, the variance on the frequency domain is multiplied with the square of the weights from the A-weighting curve, which is the variance of a-weighed equal-loudness sound pressure. Then the average of weighted variances is the single variance value:

$$\bar{v} = \frac{1}{S}\sum_{i=1}^{S} a_i^2 var_i \tag{8}$$

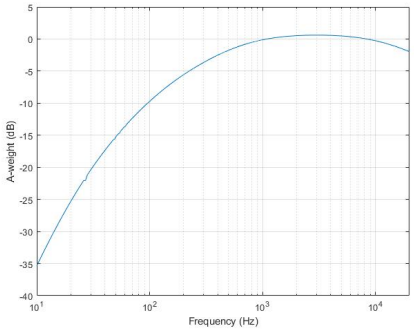
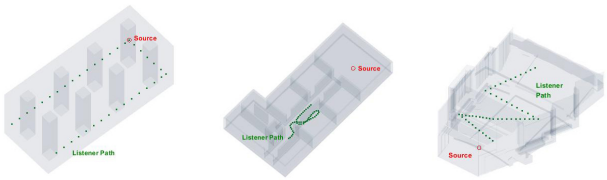
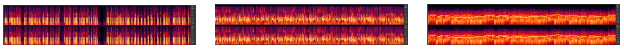


Figure 1. The A-weighting curve [8]. Sound signals of different frequencies are considered being equal-loudness if their sound pressure level multiplying the weight (or adding in log scale) are equal.



(a) pillar (b) roomset (c) elmia

Figure 2. Scenes used in our experiments and benchmarks. A sound source is located in a fixed position. The listener moves along a path, creating dynamic impulse responses and sound propagation results. (a) is a rectangular room with eight pillars creating complicated occlusion. This scene is from [11]. (b) is a realistic building floor with a set of connected rooms. (c) is the model of Elmia Concert Hall. These two scenes are from [10].



(a) male (b) female (c) music

Figure 3. Spectrograms of input sound signals used in our experiments and benchmarks. (a) is a male voice talking about "infinity" in English. (b) is a female voice in Chinese. (c) is a music clip of Canon in D Major by J. S. Bach.

Where S is frequency band count, or length of frequency domain, a_i is A-weight at frequency i, and var_i is the i-th variance on frequency domain.

The single-value weighed variance v still holds the relationship with ray count N, like variance values on the frequency domain. This value is more useful than variance function on time or frequency domain because it can be viewed as an optimization target for GA algorithms. This is going to be described in the next section.

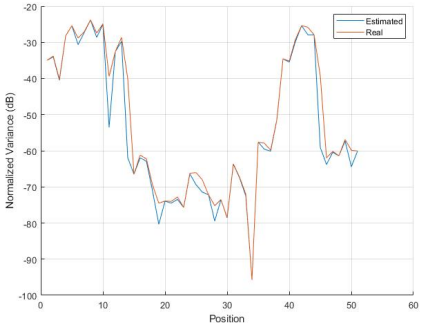


Figure 4. Estimaed and real A-weighting variance at each point on the path.

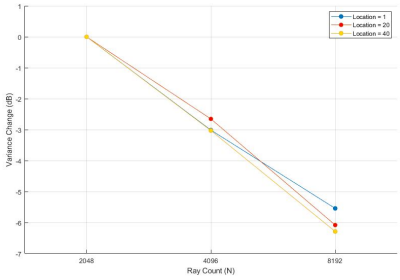


Figure 5. Variance changes with sample count N on Location 1, 20 and 40. $N = 2048$ is used as a standard sample count. From this experiment, a $3dB$ variance reduction is observed each time the sample count N is doubled.

3.4. Verification

3.4.1 Background

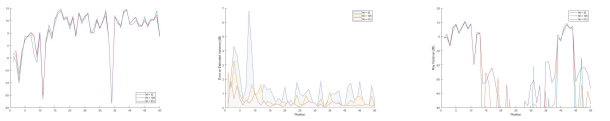
We use a simplified "roomset" scene to show several assumptions and conclusions above. This scene is a building floor with several rooms connected to a corridor. A sound source is placed in the largest hall while the listener moves through the corridor and several rooms (Fig. 2b).

3.4.2 Variance estimation

At each point on the path, we have estimated A-weighting variance using the variance of IR and input sound signal. The result fits good with calculated variance from output signal (Fig. 4), which implies A-weighting variance successfully builds up the relationship between IR and convoluted sound.

3.4.3 Variance when Ray Count Changes

From the analysis above, doubling the sample count reduces the variance by half. In decibel scale, this reduction is represented as a $10log_{10}2 \approx 3.0dB$ variance reduction. We take three points on the listener path $(1st, 20th, 40th)$ and three sample counts $(N = 2048, 4096, 8192)$ to verify the



(a) (b) (c)

Figure 6. Results of test path method. (a) shows the estimated ray variance in pillar scene when test path count $N_t = 32, 128, 512$. (b) shows that the error of variance in pillar scene is lower than 1dB in most time even the test ray count is only 32. (c) gives an example of unstable variance estimation in the roomset scene where some listener positions are highly occluded and the rays are hard to reach.

variance change. Our experiment result supports the $3dB$ rule from analysis, which is shown in 5.

4. Adaptive Sound Propagation

A-weighting variance is a measure of sound quality and can be calculated by variance on IR. Always using the same sample count in a dynamic environment, such as a scene with moving sound sources or listeners, makes sound quality unnecessarily high or unacceptably low at different points. This can be reflected by the drastic oscillation of the A-weighting variance value. This phenomenon requires a sound propagation method that modifies sample count adaptively, while A-weighting variance can be used as a reference for adjusting sample count. But the problem is that calculating the variance requires calculated IR, which does not exist until the sound propagation step finishes. The "test path" method is proposed here to solve this chicken-egg problem.

4.1. Test path method

The test path method in sound propagation is proposed in [11] for estimating the mean free path in room acoustics. We have found that the test path method can estimate more other properties on IR, including variance. From our experiment result, collecting frequency variances into a single value helps to overcome the lack of data count and get a reasonable variance value in some cases. In other cases, the variance estimation is not precise but the errors on sample count estimation and sound propagation are limited.

In path-based GA methods, the energy provided by each ray path, or each sample, is under an independent and identical distribution (IID). The variance of collected energy is the variance of the mean value of all paths, $\frac{1}{N} var_{single}$, where var_{single} is the variance of a single sample. The main idea of the test path method is to estimate var_{single} with few rays and then obtain a target ray budget with target variance in the sound propagation step.

From our experiment, a reasonable variance can be obtained with only 32 rays in open or slightly occluded scenes (Fig. 6). In most listener positions of pillar scene, the estimated 32-ray "single ray" variances var_{single} is close to the results using more (128 or 512) rays. Compared to reference single ray variance (obtained using 2048 rays), more test path results in a lower estimation error, but the 32-ray case is good enough (less than 5% of the estimation error of variance). Two "spikes" appear on Fig. 6b because the energy in these positions are indirect and plain path tracer suffers a relatively larger error. This is similar to the "failure" but acceptable case in next paragraph.

A "failure" case can be observed in the "roomset" scene where variance estimation gets incorrect results on some highly occluded listener positions (20th to 35th and after 45th). However, this error is not so severe in the sound propagation step. Here are two reasons why this error is acceptable: 1) this experiment only uses a plain path tracing algorithm, which is hard to handle this case. The result would be better if the test path method is applied on advanced algorithms, and 2) test path can be used as the lower bound of sample count, and the highly occluded listener positions can obtain a reasonable propagation result having this lower bound when their energy and variance is far lower than other positions, which is shown in the benchmark results below.

The test path method works with very low extra time cost. Compared to thousands or more samples per frame in sound propagation, 32 rays of test paths only increase overhead by less than 1%. The other part of the extra time cost in this method when evaluating A-weighting variance is the fourier transform. Due to the complexity of FFT is $O(nlogn)$, where n is the length of data, this part of cost is constant to total sample count N. The real time cost of fourier transform would be even lower because many points in the test path generated IR is 0. The storage overhead of test path method would be higher than plain path tracer because it stores N_t times of impulse responses. But the storage cost would not increase by N infinitely. If we assume the length of impulse response is 100K, the storage cost of a single impulse or frequency response would be less than 0.2M. 32 test path would cause a maximum of 6.4M memory cost. This means the storage overhead from test path method is acceptable in modern computers.

Meanwhile, this test path method is not available for all of the variables or statistics in GA based sound propagation, yet many statistics can be evaluated using our method. For example, estimating mean value instead of variance brings up a greater (more than one order of magnitude) error that the estimated mean does not good enough for further measure calculation. SNR estimation also cannot be achieved by test paths because the estimation of variance is accurate but the estimation of mean is poor.

4.2. Adaptive algorithm framework description

Here we describe the framework of adaptive sound propagation (Algorithm 1).

Algorithm 1 Adaptive sound propagation

Require: $Scene, N_t, var_{target}$
Ensure: IR
  for $i = 1$ to $N_t - 1$ **do**
    $IR_{test,i} \leftarrow path\_tracing(Scene, sample = 1)$
  end for
  $var_{sample} \leftarrow var(FT(IR_{test,i}))$
  $N \leftarrow var_{sample}/var_{target}$
  if $N < N_t$ **then**
    $IR = collect(IR_{test,i})$
  else
    $IR = path\_tracing(Scene, sample = N)$
  end if

The adaptive method runs every time the setting of the scene changes. For example, when the listener moves along a path, a new sample count N and scene IR is calculated at each point of the path. The adaptive method runs as follows: First, a batch of N_t test paths are emitted to estimate sample variance var_{sample}. Then the estimated variance is compared to a user-defined target variance var_{target} to get ray sample count N. If N is smaller than N_t, impulses that are already generated can be used as the final impulse response. Otherwise, the main sound propagation step is called with N samples. This method only requires that each ray in GA sound propagation is IID so that their result IR and variance can be added or linearly scaled. How the $path\_tracing$ function works in practice, like using importance sampling or bi-directional path tracing, and index-irrelative operations like HRTF, does not affect the correctness of this algorithm.

4.3. Benchmark

This subsection shows the result of our experiments on the adaptive sound propagation framework. We compare the performance of this algorithm with different scene geometry, input sound, and reverberation. Target normalized variance of this algorithm is set to $-20dB$ in most cases and $-10dB$ in "Elmia" and "music" case 7, and the result variance is calculated from output sound clips, by rendering the sound on each point 100 times. After the adaptive step, a constant-sample GA propagation procedure is conducted as a reference. The total sample count of adaptive and original sound propagation methods are identical. All the paths are emitted, reflected, and collected using a plain backward path tracer, and a bidirectional path tracer in the last benchmark, with no optimization technique like diffuse cache or path cache. The experiments are conducted in Matlab on a computer with Intel i7-6700HQ 2.60GHz CPU and 16.0GB memory.

4.3.1 Sample Count Optimization

Studies on cognitive psychology point out that thresholds of "resolutions" widely exist on human perception, including the *just noticeable difference (JND)* in sound clarity and speed [22, 23]. In the sound propagation method, a threshold variance can be assigned as the maximal target value to make sure that each of the propagated sound clip has a lower or equal variance compared to the target. In plain pase-based GA methods, the sample count is determined only by the highest variance to keep all the variance lower than the target, which causes unnecessary samples on other positions of the listener. In our adaptive method, the necessary sample count is predicted at each position of the listener, which makes the most of the calculation resources. The sample count in the original method can be calculated as var_{max}/var_t to make sure the whole variance curve drops below the target, which could be reductant. Table. 1 shows the significant reduction of required sample counts to reach the same maximal variance limit in adaptive path tracing than in the original method in different scenes.

Scene	Samp (req.)	Samp (opt.)	Reduction
Pillar	15789	4411	72.1%
Roomset	6266	703	88.8%
Elmia	42720	8419	80.3%

Table 1. Sample Count Optimization

4.3.2 Scene Geometry

We tested sound propagation of the same input signal (male voice) in multiple scenes: Pillar (Fig. 2a), Roomset (Fig. 2b) and Elmia (Fig. 2c) using same total sample budget of original and adaptive methods. The results are shown in Fig. 7a, 7b and 7c. In Pillar, adaptive sound propagation has a more stable variance, while the original method results in a high variance in some listener positions. In the roomset scene, the adaptive variance becomes lower than the target in the middle few positions of the path. This is because N calculated in the test path step is smaller than N_t so the test paths are used as the final result, which has a smaller variance than expectation. Since small variance represents a more accurate sound signal, this result meets the demand of limiting large errors in sound propagation. Elmia scene encounters a bit larger error on balancing the variance. This is caused by the complicated geometry of that scene and the only 32 test paths are not enough to get a precise estimation result as other scenes. Still, the result in Elmia is acceptable and the unnecessary samples are reduced.

4.3.3 Input Sound

Three sound clips, the default male voice clip in English (Fig. 3a), a clip of female voice in Chinese (Fig. 3b), and a clip of Canon music in piano (Fig. 3c), are applied as input sound signal in this experiment. All the other conditions, including scene geometry (pillar), source and listener position, and reverberation are fixed. The results are shown in Fig. 7a, 7d and 7e. The case of the female voice has a similar performance to the default male voice. An error is observed on optimized variance at the beginning of the music case. This is caused by the very narrow frequency band (clear sound of piano keys) of the input signal at those points.

4.3.4 Reverberation

To test our adaptive method in different reverberation, we set sound velocity to $s_1 = 343m/s$ and $s_2 = 0.1s_1 = 34.3m/s$ respectively. The setting that sound travels in $1/10$ velocity is equivalent to the reverberation magnified by 10 times. The results are shown in 7a and 7f. The high error at the beginning of the long reverb case is caused by the low sound velocity. Some of the sound paths are still propagating to the listener in the beginning seconds, which makes the variance harder to estimate. Reverberation caused by large scenes also suffers from this problem. In the later seconds, the adapted variance is flattened, meaning the input sound finally reaches the listener and does not bring up more error generally.

4.3.5 Using IR-related measure

We test the SNR measure from [10] to calculate N_t in our adaptive framework. The SNR measure is defined on IR and do not control the auralized result. In the sample count estimation step, the maximum value of SNR on the time-domain energy response is used as a target. Experiment result is shown in Fig. 7g. Two conclusions can be obtained from theoretical analysis and experiment result: 1) Controlling measures on IR cannot also control measures on auralized sound, causing the result unstable and the maximal variance is not reduced; 2) SNR measure helps flatten the beginning variances because the instability of these variances is mainly caused by the change of scene geometry, but on the following positions of the listener, geometry change and input signal change both play a role in variance change, which is hard for SNR measure to handle. Both of these conclusions support our A-weighting measure.

4.3.6 Different Sample Strategies

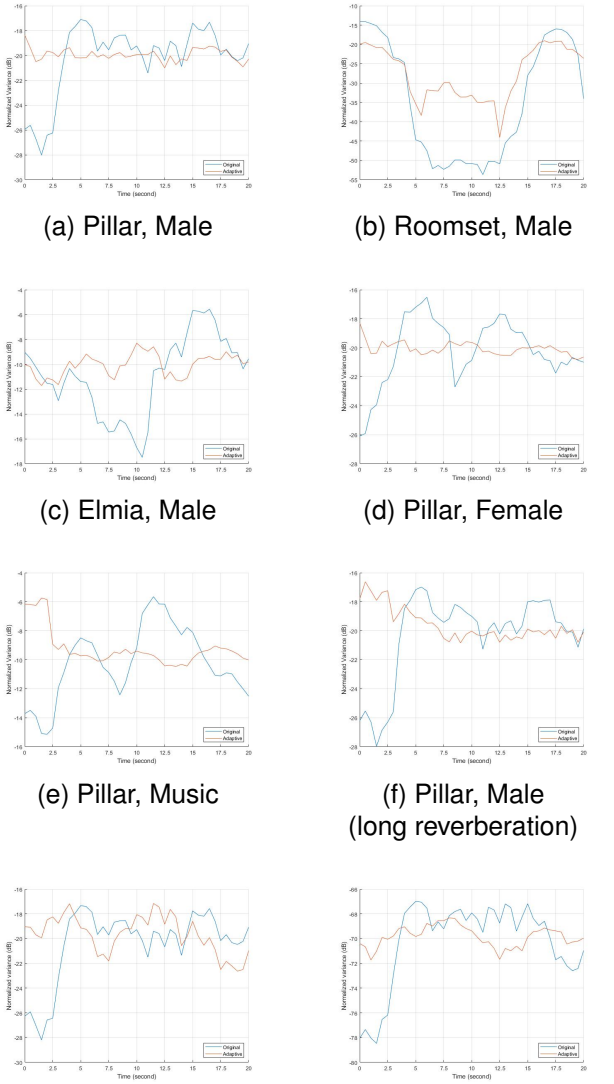
Finally, we implemented a bidirectional path tracer to replace the plain backward path tracer in previous experiments. The result is shown in Fig. 7h, showing that the bidirectional path tracer is also fit for our A-weighting variance and adaptive method. Since the A-weighting variance is calculated using output sound, it is natural to compare different algorithms with this measure. The total time cost of the BDPT experiment is similar to the first case, but the variance obtained is much lower (-70dB), which implies bidirectional path tracer performs better in the Pillar scene with obstacles and occlusion.

5. Conclusion

We have analyzed Monte Carlo path tracing in GA sound propagation and have proved that A-weighting variance is the suitable measure which is both controllable by sample count in sound propagation methods, and effective for measuring the quality of the auralized sound the users would hear. We have presented an adaptive sound propagation method to estimate the required sample count, balance the computation budget, and reduce unnecessary samples. Our method is universal enough to cooperate well with different sampling strategies or path tracers. All the benchmarks show the effectiveness of our measure and adaptive method in different scene geometry, input sound, and acoustic condition. Compared to existing measures like SNR, our measure and adaptive approach perform better on output sound signals.

A limitation of our work is the A-weighting variance not perceptual enough to describe how the users feel with the propagated sound. A-weighting variance is mainly calculated in theory but psychological experiments are required to prove the effectiveness of this measure in human perception. Another limitation is from GA methods which our work bases on. GA methods have system error on low-frequency sound, which makes A-weighting variance not suitable for an input signal of low frequency perceptually. Meanwhile, input signals with narrow frequency bands will make our method hard to evaluate a correct variance value. Finally, how to apply the test path method in non-GA or hybrid methods is still a problem. Many hybrid sound propagation frameworks have been proposed nowadays and finding a way to apply our method into these frameworks is helpful to get a better sound propagation result.

More works can further be done to improve our knowledge of sound quality from GA methods. The adaptive method would be more useful if the threshold variance is guided by measures in human cognition, meaning that psychological experiments can be conducted to find out the just noticeable A-weighting variance most users hold. Other perceptual measures can be added to our adaptive framework to get a better computation balance. On the frequency domain, a more stable method can be developed for handling input sound cases of low frequency and narrow frequency bands. Finally, our A-weighting variance, adaptive



(a) Pillar, Male (b) Roomset, Male

(c) Elmia, Male (d) Pillar, Female

(e) Pillar, Music (f) Pillar, Male (long reverberation)

(g) Pillar, Male (ERSNR controlled) (h) Pillar, Male (BDPT)

Figure 7. Benchmark of A-weighting variance of different scene, input signal and reverberation. (a) is the original setting: "pillar" scene, "male voice" input and normal reverberation. (b) and (c) replaces scene with "roomset" and "Elmia concert hall" respectively. (d) and (e) use a female voice and a piece of Canon music as input signal. (f) enhances the scene reverberation by 10 times using lower sound velocity. (g) replaces estimated A-weighting variance for testing sample count with SNR measure in [10]. (h) replaces plain path tracer with a bidirectional path tracer. The target variance in "Elmia" and "music" are assigned to -10dB instead of -20dB to reduce sample count under 10k and make sure the experiments do not take too much time. The target variance in "BDPT" is -70dB because bidirectional path tracer performs far better than the plain one.

methods, and test path technique can also be explored in other hybrid sound propagation frameworks.

6. Acknowledgement

The authors are partially supported by the National Natural Science Foundation of China (NSFC) grant (No 61732016).

References

[1] Burg J, Romney J, Schwartz E. *Digital Sound & Music: Concepts, Applications, and Science[M]*. Franklin, Beedle & Associates, 2017. 2

[2] Salom I M, Mijić M M, Ćertić J D, et al. *Subjective evaluation and an objective measure of a church bell sound quality*[J]. Applied acoustics, 2014, 85: 97-105.

[3] Toole F E. *Subjective measurements of loudspeaker sound quality and listener performance*[J]. Journal of the Audio Engineering Society, 1985, 33(1/2): 2-32.

[4] ISO 3382-1:2009. *Acoustics — Measurement of the reverberation time of rooms with reference to other acoustic parameters, Part 1: Performance spaces*. ISO. 2009.

[5] ISO 3382-3:2012. *Acoustics — Measurement of the reverberation time of rooms with reference to other acoustic parameters, Part 3: Open plan offices*. ISO. 2012. 2

[6] ISO 226:2003. *Acoustics — Normal equal-loudness-level contours*. ISO. 2003. 2

[7] ITU-T P.800.1. *Methods for objective and subjective assessment of speech and video quality*. ITU. 2016. 2

[8] IEC 61672-1:2013. *Electroacoustics - Sound level meters - Part 1: Specifications*. IEC. 2013. 1, 3, 4

[9] Middlebrooks J C, Green D M. *Sound localization by human listeners*[J]. Annual review of psychology, 1991, 42(1): 135-159. 2

[10] Cao C, Ren Z, Schissler C, et al. *Interactive sound propagation with bidirectional path tracing*[J]. ACM Transactions on Graphics (TOG), 2016, 35(6): 1-11. 1, 2, 4, 7, 8

[11] Rungta A, Rewkowski N, Klatzky R, et al. *P-Reverb: Perceptual Characterization of Early and Late Reflections for Auditory Displays*[C]//2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019: 455-463. 4, 5

[12] Kuttruff H. Room acoustics[M]. Crc Press, 2016.

[13] Blauert J. Spatial hearing: the psychophysics of human sound localization[M]. MIT press, 1997.

[14] Langendijk E H A, Bronkhorst A W. Contribution of spectral cues to human sound localization[J]. The Journal of the Acoustical Society of America, 2002, 112(4): 1583-1596.

[15] Ladefoged P, McKinney N P. Loudness, sound pressure, and subglottal pressure in speech[J]. The journal of the Acoustical Society of America, 1963, 35(4): 454-460.

[16] Lafortune E P, Willems Y D. Bi-directional path tracing[J]. 1993. 2

[17] Lafortune E. Mathematical models and Monte Carlo algorithms for physically based rendering[J]. Department of Computer Science, Faculty of Engineering, Katholieke Universiteit Leuven, 1996, 20: 74-79.

[18] Vorländer M. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm[J]. The Journal of the Acoustical Society of America, 1989, 86(1): 172-178. 2

[19] Lentz T, Schröder D, Vorländer M, et al. Virtual reality system with integrated sound field simulation and reproduction[J]. EURASIP journal on advances in signal processing, 2007, 2007: 1-19. 2

[20] Schissler C, Manocha D. Interactive sound propagation and rendering for large multi-source scenes[J]. ACM Transactions on Graphics (TOG), 2016, 36(4): 1.

[21] Schissler C, Mehra R, Manocha D. High-order diffraction and diffuse reflections for interactive sound propagation in large environments[J]. ACM Transactions on Graphics (TOG), 2014, 33(4): 1-12. 2

[22] Bradley J S, Reich R, Norcross S G. A just noticeable difference in C50 for speech[J]. Applied Acoustics, 1999, 58(2): 99-108. 6

[23] Quené H. On the just noticeable difference for tempo in speech[J]. Journal of Phonetics, 2007, 35(3): 353-362. 6

[24] Kosten C W. The mean free path in room acoustics[J]. Acta Acustica united with Acustica, 1960, 10(4): 245-250.

[25] Manocha D, Lin M C. Interactive sound rendering[C]//2009 11th IEEE International Conference on Computer-Aided Design and Computer Graphics. IEEE, 2009: 19-26. 1

[26] Savioja L, Svensson U P. Overview of geometrical room acoustic modeling techniques[J]. The Journal of the Acoustical Society of America, 2015, 138(2): 708-730. 1