Audio Engineering Society

Conference Paper

Presented at the 2022 International Conference on
Audio for Virtual and Augmented Reality
2022 August 15–17, Redmond, WA, USA

Perceptual evaluation of low-complexity diffraction models from a single edge

Joshua Mannall[1], Orchisama Das[1], Paul Calamia[2], and Enzo De Sena[1]

[1] *University of Surrey, Department of Music and Media, Guildford, United Kingdom*
[2] *Reality Labs Research at Meta*

Correspondence should be addressed to Enzo De Sena (e.desena@surrey.ac.uk)

ABSTRACT

Geometric acoustic models have a lower computational complexity than wave-based methods due to the assumption that sound propagates as rays, however this fails to consider the wave-like properties of sound such as diffraction. Historically, tthe Biot-Tolstoy-Medwin (BTM) model and the Uniform Theory of Diffraction (UTD) have been used to augment geometric acoustic models with diffraction. Computational efficiency is essential for real-time application and recently two more efficient models, the Volumetric Diffraction and Transmission (VDaT) model and an infinite impulse response filter (IIR) approximation, were proposed to approximate these solutions. A higher-order IIR filter approximation is proposed in this paper. An experiment is carried out to evaluate the perceived naturalness of these approximations compared to the more accurate analytical solutions. Stationary and moving receivers were considered in simple geometries with a single edge. The results suggest that the higher order IIR approximation is perceptually similar to the BTM model. VDaT and the low order IIR approximation were found to be less natural in some cases. While in dynamic scenes, VDaT was found to be significantly more natural than the other models. The experiment was limited in scope by the simplicity of the scenes considered, however the results suggest the models are perceptually similar. Improvements to the higher-order IIR approximation are suggested and a recommendation is made for future perceptual evaluations.

1 Introduction

In recent years, there has been a rising focus on room acoustic models tailored for applications in extended reality (XR), a term encompassing virtual reality (VR), augmented reality (AR) and mixed reality (MR). In this context, the main challenge is to simulate complex environments while keeping a computational complexity suitable for real-time operation. Wave-based methods yield the most physically accurate results, but they require significant computational resources to run [1]. Amongst the most popular models for real-time applications are delay-network based models [2, 3] and geometric acoustic (GA) models [4], which have a significantly lower computational complexity than wave-based ones. Diffraction is not inherently accounted for by GA models, but a number of methods can be used to augment GA models with diffraction. Most prominently, the Biot-Tolstoy-Medwin (BTM) model [5] is a time-domain model considered to be amongst the most physically accurate methods in the literature, and the

Uniform Theory of Diffraction (UTD) model [6] is a frequency domain model and a high frequency approximation assuming infinitely long wedges. Both models, however, carry a significant computational penalty.

Recently, a very active area of research has been the efficient modelling of diffraction for real-time GA. In 2021, Schissler *et al.* [7] proposed a dynamic method to efficiently find diffraction paths in complex geometries. In 2020, Pisha *et al.* proposed VDaT [8], an approximate model that applies diffraction to a path based on whether it, and sub paths around it, collides with the scene geometry. In 2021, Kirsch and Ewert [9] proposed a 2nd-order infinite impulse response (IIR) filter designed to approximate the frequency response of UTD while having a very low computational complexity. New diffraction models are usually validated using objective measurements [8, 9], most often in the frequency domain, while, to the best of the authors' knowledge, their perceptual effect is yet to be assessed.

The aim of this paper is to evaluate low-complexity diffraction models perceptually. Twenty subjects participated in a formal listening test comparing the following diffraction models: BTM [10, 5], UTD [6], VDaT [8], and the 2nd-order IIR model proposed by Kirsch and Ewert [9]. A similarly-designed 5th-order IIR approximation is also proposed in this paper and is included amongst the tested models.

This paper focuses on the simple and as-of-yet untested case of a single edge in free field. Participants rated the stimuli in terms of "naturalness," using a modified Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) methodology. Four static scenes were tested with the receiver in different positions around an edge. Two dynamic scenes were tested with the receiver moving towards a source occluded by a building or wall. Results of the experiment indicate that (a) the 5th-order IIR approximation is perceptually similar to BTM, (b) VDaT is perceptually similar to BTM at large bending angles, (c) the 5th-order IIR approximation is more natural than the 2nd-order IIR approximation at large bending angles, but lower near the shadow boundary, and (d) UTD is perceptually similar to BTM.

The paper is organised as follows. Section 2 outlines the five compared diffraction models. Section 3 describes the experimental methodology and setup. Section 4 presents an analysis and discussion of the results. Finally, Section 5 presents conclusions and suggests avenues for future research.

2 Compared edge diffraction models

This section gives an overview of the models compared as part of the perceptual experiments. Fig. 1 shows the cylindrical coordinates reference system used throughout the paper, and Fig. 2 shows the zones around the edge characterised by direct, reflected and diffracted components.

2.1 Biot-Tolstoy-Medwin

The BTM model is a physical model that is generally considered to be the most accurate diffraction model available [10]. The model is formulated in the time domain and is based on Huygen's principle whereby the sound field is expressed as an integral summation of secondary sources lying on the edge. In its original formulation, BTM modelled the response from a point source around an infinite rigid wedge. It has subsequently been extended to the finite-edge case and second-order diffraction [12, 13]. More recent work has focussed on analytical solutions for the source directivity functions of secondary edge sources [14], behaviour at the shadow boundary [11] and efficient edge division and integration methods [5]. Several studies have shown good agreement between the model predictions and real-world measurements [12, 13, 15, 16].

2.2 Uniform Theory of Diffraction

In the high-frequency regime, sound wave fronts are well approximated by sound rays [17]. This approx-

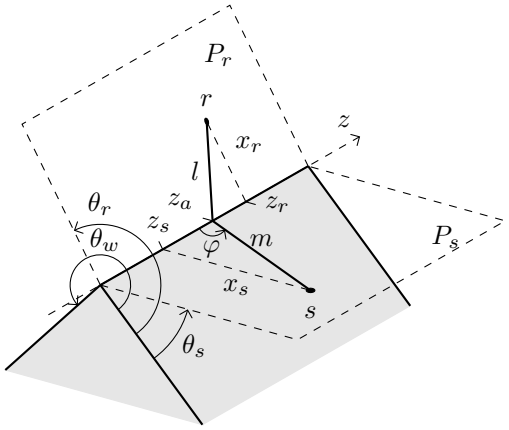


Fig. 1: Edge geometry and coordinate system. Locations are specified in cylindrical coordinates where x is the radial distance from the edge, θ is the angle between two planes and z is the distance along the edge. P_s and P_r are virtual planes that contain the edge and the source and receiver respectively. The apex point z_a is defined by the shortest path between the source and receiver via the edge. Adapted from [5].

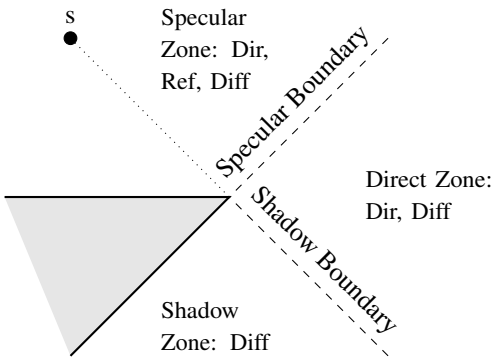


Fig. 2: Zones around an edge where direct (Dir), reflected (Ref) and diffracted (Diff) sound components will be present. Adapted from [11].

imation holds true as long as the boundary sections are much larger than the wavelength, and the surface roughness is much smaller than the wavelength [4]. The UTD model proposed in [6] is an extension of the Geometrical Theory of Diffraction (GTD) model [18]. It is a frequency-domain method that assumes infinitely long edges. While diffraction is generally present in the specular, direct and shadow zones, in order to reduce complexity, most implementations [19, 7] only consider it in the shadow zone, where it is dominant. In the implementations described in [20, 7], edge visibility maps were precomputed for more efficient diffraction path finding. Runtime processing, which included calculating the frequency response of each diffraction edge, took 0.14ms to 7.5ms per source.

2.3 Volumetric Diffraction and Transmission

The VDaT model approximates diffraction by calculating sub paths in concentric rings around the direct path between the source and receiver (or around the reflective path connecting them) [8]. The model uses the number of sub paths blocked by the scene geometry to determine the frequency response applied to the path. Fig. 3 shows an example of a path with a single ring and eight sub paths. The model was created by using BTM to calculate the average frequency response across hundreds of arbitrary planes that blocked every path in one ring and none in the next size up. They created parameters to approximate the frequency response as constant up to a frequency where it rolls off at 10dB/decade. This method ignores precise time of arrival differences and direction of arrival differences between diffraction paths. Therefore, it ignores the comb filtering effect resulting from interference between multiple diffraction paths around an obstacle. The authors of [8] argue

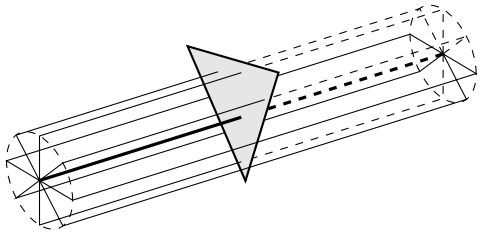


Fig. 3: Geometry of a single VDaT path. The direct path from source to receiver is in bold. Four obstructed sub paths are shown as dashed lines, while unobstructed sub paths are solid lines. The diameter of the cylinder determines the frequency being considered. Adapted from [8].

that this is perceptually irrelevant since non anechoic conditions cause an averaging effect across frequency much like VDaT. Another limitation is that objects are treated as acoustically transparent as the direction of arrival is not modelled via a specified edge. They argue that the shortest path is the sound transmitted through obstacles. Therefore, the precedence effect (that the first sound arrival tends to determine the perceived direction of arrival [21]), means accurate direction of arrival may be less perceptually important. To the best of the authors' knowledge, these arguments have not yet been validated via perceptual experiments. VDaT has been implemented in real time up to 6th order of reflection paths with runtime processing completed in 1.7 to 4.95 ms per source. The model does not require a precomputation step.

2.4 Kirsch and Ewert's 2nd-order IIR filter

Kirsch and Ewert [9] recently proposed approximating edge diffraction from an infinite edge using a low-pass filter with a 3 dB per octave roll-off with a cut-off frequency (f_c) determined by the edge geometry. They approximated the UTD frequency response using a modified fractional order low-pass filter, found heuristically, with the transfer function

$$H(f) = \left[j \left(\frac{f}{f_c} \right)^{0.625} + 1 \right]^{-0.8}, \tag{1}$$

$$f_c = \frac{c \cdot m_w \cdot m_p}{3\pi d (1 - \cos\theta_b) \sin^2 \varphi}, \tag{2}$$

where m_p and m_w are defined as

$$m_p = 1 - 0.75 \tanh \frac{1}{2\theta_b} \sqrt{\tanh 2\theta_{min}},$$

$$m_w = \left(1 - 0.75 \left(\frac{\theta_b}{\theta_{b,max}} \right) \sqrt{\sin \frac{-\theta_w}{2}} \right)^{-1},$$

and $d = \frac{2m \cdot l}{m \cdot l}$, $\theta_b = \max[10^{-4}, \theta_r - (\theta_s + \pi)]$, $\theta_{min} = \min[\theta_s, \theta_w - (\theta_b + \theta_s + \pi)]$, and $\theta_{b,max} = \theta_w - (\theta_{min} + \pi)$. When the bending angle (θ_b) is less than zero, the receiver is in the direct zone and no filtering is applied. In other words, this model does not consider the effect of diffraction outside of the shadow zone. Kirsch and Ewert proposed approximating this filter using a combination of first-order low-pass filters (LPF) and high shelf filters (HSF) in series and parallel, the transfer functions of each are given by $H_{LPF}(z) = \frac{K + Kz^{-1}}{K+2+(K-2)z^{-1}}$ and $H_{HSF}(z) = \frac{K+2B_\pi+(K-2B_\pi)z^{-1}}{K+2+(K-2)z^{-1}}$, where B_π is the high shelf gain in dB, $K = 2\pi\frac{f_c}{f_s}$, and f_s is the sample rate. More specifically, the filter of equation (2) is approximated as a low-pass filter and a high shelf filter in series, with cut-off frequencies f_0 and f_{sh} respectively given by

$$f_0 = 1.11 f_c \left(\frac{15.6}{f_c}\right)^{0.141} \quad ; \quad f_{sh} = 209 f_c \left(\frac{15.6}{f_c}\right)^{0.827} .$$

The parameter B_π, for the high shelf filter, is taken as the difference in dB between the first-order low-pass filter and the fractional filter given in equation (1) at 20kHz. A scaling factor can be added as $A = \frac{1}{r}$ to account for attenuation over distance. This model will be referred to as IIR_{lo} henceforth.

2.5 Proposed 5th-order IIR filter

Expanding on the 2nd-order IIR approximation suggested in [9], this paper proposes a 5th-order approximation that combines three low-pass filters in parallel followed by four high shelf filters connected in series. This will be referred to as IIR_{hi} for the remainder of this paper. The cut-off frequencies of the associated filters are defined as follows:

Low-pass filter

$$f_0 = 1.11 f_c \left(\frac{15.6}{f_c}\right)^{0.141}$$
$$f_1 = 2 f_0 \,;$$
$$f_2 = 3 f_0 \,;$$

High shelf filter

$$f_{sh0} = 209 f_c \left(\frac{15.6}{f_c}\right)^{0.827} ,$$
$$f_{sh1} = 1080 \frac{f_{sh0}}{4320 - \theta_r \pi} ,$$
$$f_{sh2} = 2 f_{sh1} \,,$$
$$f_{sh3} = 1188 \frac{{f_{sh0}}^{1.3}}{4320 - \theta_r \pi} .$$

The gain of the first high shelf filter is the same as for IIR_{lo}. The gains for the second, third and fourth high shelf filters are given as

$$B_{\pi 1} = 9 \left[\frac{\theta_b}{{\theta_w}^2 - \pi} - \min\left(1, \frac{20\theta_b}{\pi}\right) \right] ,$$

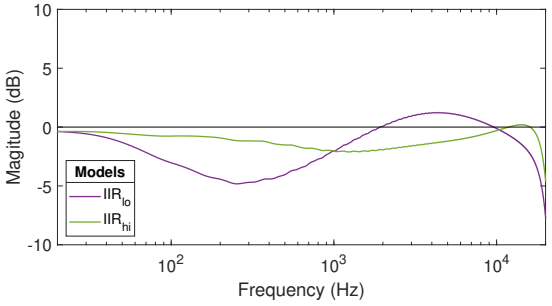


Fig. 4: Deviation by the IIR approximations from the BTM model in the frequency domain for diffraction around an edge. Geometrical parameters were: $\theta_w = 270°$, $\theta_s = 30°$, $\theta_r = 260°$, $\varphi = 90°$, $r_s = r_r = 1\text{m}$.

$$B_{\pi 2} = -9 \,; \qquad B_{\pi 3} = 15 \,,$$

when the receiver is in the shadow zone. They are zero in all other cases. The following modified scaling factor was added to account for attenuation over distance, and to adjust the level at large bending angles and small edge angles:

$$A = \frac{2.1\pi - \theta_w}{2.9 r (2.1\pi - \theta_w) + 0.1 \max[0, \theta_r - \theta_s - \frac{3\pi}{4}]^{0.2}} .$$

The parameters of the model were tuned by comparing frequency responses against those produced by BTM for receivers spaced 10 degrees apart around infinite wedges from 5 to 175 degrees. Fig. 4 reports an example of the frequency deviation of IIR_{hi} and IIR_{lo} from BTM. Fig. 5 shows examples corresponding to the static scenes described in Fig. 6a and in Section 3.1, where it can be seen that IIR_{hi} has a closer match to BTM compared to IIR_{lo}.

3 Experiment

A listening experiment was carried out to evaluate the naturalness of the models described in Section 2.

3.1 Stimuli

The stimuli for each model were generated as follows:

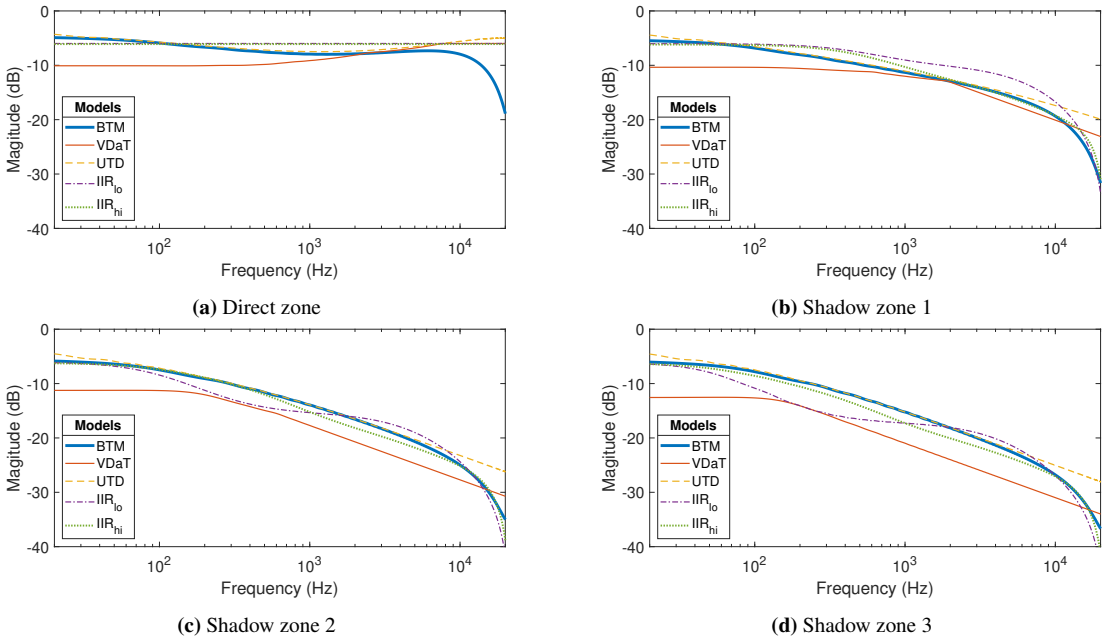
BTM: Using the EDtoolbox [22] which is based on the implementations described in [14] and [11].

UTD: Python code in [8] ported to Matlab.

VDaT: Python code in [8] ported to Matlab.

IIR_{lo}: Second-order IIR filter as described in [9]. Implemented in Matlab.

IIR_{hi}: Same as above using the filter structure described in Section 2.5.



(a) Direct zone

(b) Shadow zone 1

(c) Shadow zone 2

(d) Shadow zone 3

Fig. 5: Frequency responses of the diffraction models for the static scenes described in Fig. 6a and in Section 3.1.

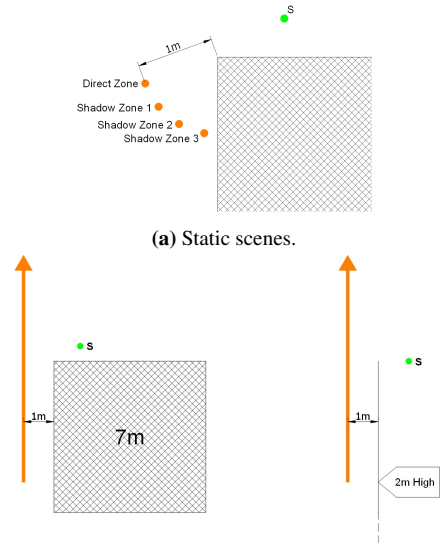
For BTM, the length of the impulse response (IR) is dependent on the length of the edge. An IR length of 512 was chosen for VDaT and the IIR models. Since UTD models the phase shift as sound propagates along the direct and diffraction paths, the IR length for UTD was extended to 2048 samples to account for the initial time delay and to prevent wrapping in the time domain. For the VDaT and IIR models, which do not model phase, an initial time delay was added based on the geometric path length. A 256-sample delay was added to the other models to allow for direct comparison in the time domain with the VDaT model.

A sample rate (f_s) of 44,100 Hz was used across all the models and the stimuli were obtained by convolving the respective IRs with the programme material. Fractional delay finite impulse response (FIR) filters [23, 24] and overlap-add with the Hanning window was used to interpolate the time delay as the receiver moved.

The aim of the experiment was to determine the naturalness of the models as a function of the bending angle. Static scenes were chosen to compare how the models perform in different zones around the edge. Dynamic scenes were chosen to compare how the models behave as the receiver moves, since this is important for XR and gaming applications.

For static scenes, a single edge with $\theta_w = 270°$, $\theta_s = 30°$ and a length of 7 m was chosen to approximate the side of a building. Four evenly spaced receiver positions at 200°, 220°, 240° and 260° were selected so that one lies in the direct zone and the others lie in the shadow zone with increasing bending angles. This is summarised in Fig. 6a. The source and the receivers were placed at $x_s = x_r$ =1 m and 1.6 m above the floor. A floor plane was added in order to prevent VDaT subpaths from traveling below the wedge.

The aim for the dynamic scenes was to design scenarios the participants could relate to. Two scenarios were designed, shown in Fig. 6b and Fig. 6c. The first scenario involves a 5x5x7 m^3 cube with a source at $x_s =$ 1 m and $\theta_s = 30°$, which is intended to replicate the experience of walking around the corner of a building. The second scenario involves a 2 m high plane with a source at $x_s = 1$ m and $\theta_s = 90°$, which is intended to replicate the experience of walking around a free-standing wall. In both cases the receiver travels parallel to the wall or building, from the shadow zone to the direct zone. For all models, diffraction was considered up to 1st-order (i.e. including only diffraction paths that travel via up to a single edge). This was motivated by the fact that, compared to (BTM-computed) 3rd-order diffraction, the maximum error across all four static scenes was just 0.6 dB. In the dynamic wall scene this



(a) Static scenes.

(b) Dynamic (building) **(c)** Dynamic (wall)

Fig. 6: Diagram showing the layout for static and dynamic scenes with a source (S) and receiver positions or path.

was less than 1 dB, except at the comb filtering notches where it reached 3 dB.

The programme material was chosen to represent scenarios the subjects would be familiar with. Three anechoic items were sourced from the Zip Archive [25]: street performers, factory noise and male speech. This choice was supported by a pilot listening test where participants commented that having programme material representative of real-world experiences was useful. Specifically, the music stimulus was noted to be the most relatable. It was hypothesised that the transition across the shadow boundary would be perceptually important. So, for dynamic scenes, care was taken to ensure that the programme material was not silent at the moment of transition.

3.2 Experimental Method

The experimental method was based on the MUSHRA method as standardised in the ITU-R BS.1534-3 recommendation [26] with modifications similar to those made in [27]. No reference was provided to subjects because in the applications envisioned here, physical accuracy was considered less important than convincing rendering in terms of naturalness. The definition of naturalness was adapted from the definition given in [28] and provided to participants as *"The degree to which the stimuli conform to your experience of a source around a corner."* A 1 kHz high-pass filter was used as an anchor instead of the standard 1.5 kHz low-pass filter as it was not degraded enough to be consistently ranked lowest, as shown during a pilot experiment.

The participants were asked to judge the five models described in section 3.1. The grading phase consisted of 36 pages in a randomised order and each test page consisted of a diagram representing the scene, the original anechoic audio and the ranking scale with sliders randomly assigned to each model or the anchor. Every page was repeated twice to check for consistency. Participants completed a familiarisation phase, where they could listen to each of the unprocessed items of programme material and random selections of the static and dynamic scenes, and two training pages to familiarise themselves with the user controls.

3.3 Equipment and subjects

The listening tests were conducted in Edit Rooms at the University of Surrey with stimuli reproduction over Beyerdynamic DT 770 PRO headphones. Participants were able to adjust the listening level during the familiarisation phase. In total 20 participants completed the listening test. All participants had experience with critical listening and did not report any hearing impairment.

4 Results

4.1 Preliminary data analysis

A Kolmogorov–Smirnov test found that 34 out of 90 stimuli violated the assumption of normality required for an Analysis of variance (ANOVA) test. Investigation of the data using histograms and normality curves revealed that many of the stimuli groups exhibited significant non-normal distributions. Mendonca [29] found that MUSHRA tests often violate the interval scale, normality, equal variances and independence assumptions made in an ANOVA test which can lead to Type-I errors. Hence non-parametric tests were used for the remainder of the paper.

The mean naturalness of the models split by scene is plotted in Fig. 7. The ratings show that naturalness varies between different scenes and models. The results for BTM, IIR_{hi} and UTD have little variation among the four static scenes. The results for IIR_{lo} show a downward trend in naturalness as θ_r increases, whereas the results for VDaT show an upward trend. These observations motivated dividing the analysis by type of scene: dynamic scenes (Wall and Building), static scenes near the shadow boundary (Static Direct Zone and Shadow Zone 1) and static scenes in the shadow zone (Static Shadow Zone 2-3).

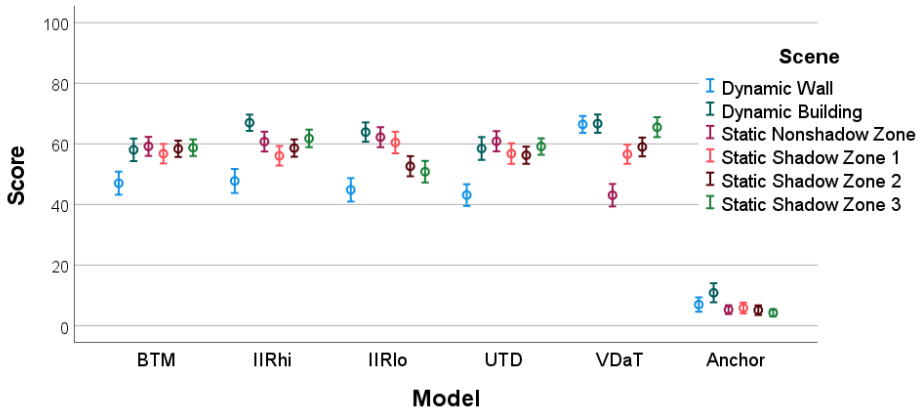


Fig. 7: Mean naturalness as a function of diffraction model for each scene. Error bars indicate the 95% confidence intervals.

Model	Scene (Mean Rank)		
	Dynamic	Static, Near shadow boundary	Static, Shadow zone
BTM	538.76	606.89	604.31
IIR_{hi}	622.85	614.66	639.76
IIR_{lo}	561.97	676.03	485.06
UTD	509.44	632.52	583.78
VDaT	769.48	472.39	689.59
Asymp. Sig.	<0.001	<0.001	<0.001

Table 1: Kruskal-Wallis tests showing statistically significant differences between models for three types of scenes. The mean naturalness ratings of each scene type are ranked from 1 (lowest) to 1200 (highest) and the cells show the mean ranking of results corresponding to each model.

4.2 Kruskal-Wallis and Friedman tests

Kruskal-Wallis tests reveal that a statistically significant difference is observed between the compared models for all three types of scenes (see Table 1). Friedman tests were also carried out as recommended in [29], with stepwise step-down comparisons to reveal where the differences are observed between groups. The results are shown in Table 2.

4.3 Static scenes

Table 2 shows that no statistically significant differences are observed between BTM and IIR_{hi}. This suggests that the approximate models may be used in applications with tight computational constraints without significant loss of perceptual quality. It should be noted, however, that in more complex cases than the single edge case, particularly with multiple small edges, the BTM model could be perceived as more natural since IIR_{hi} overestimates the low-frequency response.

In the shadow zone, there is a statistically significant difference between IIR_{lo} and the other models. IIR_{lo} is ranked lower, which shows its limitations at large bending angles and highlights the benefits of IIR_{hi}. Fig. 5d suggests that this is because IIR_{lo} over-emphasises high frequencies in the shadow zone. In the shadow zone, no statistically significant difference is observed between BTM, IIR_{hi} and VDaT.

IIR_{lo} is ranked highest near the shadow boundary. This is likely because IIR_{lo} does not attenuate the high frequencies as much as the other models at small bending angles, as shown in Fig. 5b. This is a somewhat unexpected result. A possible cause is that the top-down view affected participants' judgments on the scene geometry as the lack of direct sound may not have been clear at small behind angles. VDaT was ranked lowest near the shadow boundary. Fig. 5a suggests this is because it underestimates the low-frequency response.

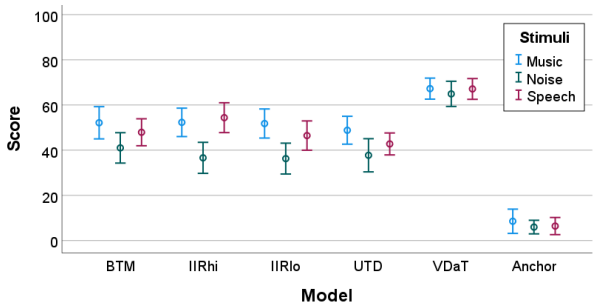
Previous studies have found that in some cases diffraction outside the shadow zone is audible [30]. However, the results here do not show any statistically significant differences between IIR_{hi}, which does not model diffraction in the direct zone, and BTM and UTD, which do. Torres [31] conjectures that in a more complex scene with many and smaller edges, diffraction in the direct and specular zones would be more audible. Testing this hypothesis is left for future work.

	Model	Subsets 1	2	3
Dynamic	UTD	2.629		
	BTM	2.719	2.719	
	IIR_{lo}	2.796	2.796	
	IIR_{hi}		3.094	
	VDaT			3.763
Test Statistic		0.325	5.819	.
Sig. (2-sided test)		0.850	0.055	.
Adjusted Sig. (2-sided test)		0.958	0.089	.
Static, Near shadow boundary	VDaT	2.263		
	BTM		3.040	
	IIR_{hi}		3.079	
	UTD		3.160	
	IIR_{lo}			3.458
Test Statistic		.	1.094	.
Sig. (2-sided test)		.	0.579	.
Adjusted Sig. (2-sided test)		.	0.763	.
Static, Shadow Zone	IIR_{lo}	2.421		
	UTD		2.888	
	BTM		3.081	3.081
	IIR_{hi}			3.288
	VDaT			3.323
Test Statistic		.	1.350	3.381
Sig. (2-sided test)		.	0.245	0.184
Adjusted Sig. (2-sided test)		.	0.505	0.288

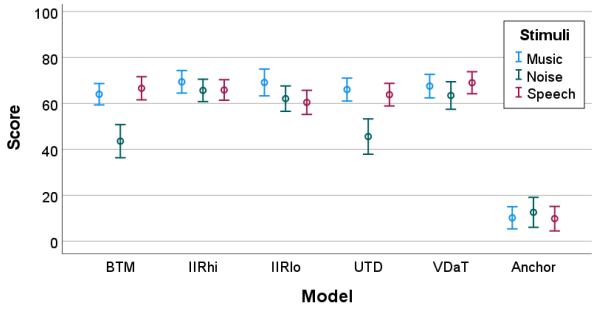
Table 2: Friedman test results showing statistically significant differences between models for three criteria: dynamic scenes, static scenes near the shadow boundary and static scenes within the shadow zone. Homogeneous subsets are based on asymptotic significances. The significance level is 0.05. The Sig. value is the significance between models within subsets and models in separate subsets have a significance level < 0.05. Cells show the average rank of models.

4.4 Dynamic scenes

In the dynamic scenes, the participants clearly preferred VDaT. Participants' feedback suggest that this may be due to the fact that VDaT is the only method that does not result in audible comb filtering effects as the receiver moves. This is because it does not correctly model the time of arrival between multiple diffraction paths. In the dynamic wall scene, comb filtering would physically be present due to the interference pattern between the path over the top of the wall and the one around the vertical edge. In the dynamic building scene, only a single edge is considered in the shadow zone. Once the receiver is in the direct zone, BTM and UTD include diffraction along with the direct path. The IIR approximations do not model diffraction in the direct zone, so, in this scenario, they consider a single path, and hence do not exhibit any comb filtering effects. Fig. 7 shows that they are consequently ranked higher.



(a) Dynamic (wall)



(b) Dynamic (building)

Fig. 8: Plot of mean naturalness as a function of edge diffraction model split by programme material for the two dynamic scenes. Error bars indicate the 95% confidence intervals.

The noise stimuli resulted in a lower mean score in all the cases where comb filtering is present as shown in Fig. 8. Because of the stimulus broadband nature, these effects are more perceivable. It is concluded that, despite comb filtering being present in real situations, it is perceived negatively by participants. It can be argued that the simplicity of the monaural scene, which neglects source directivity, reflections, higher order diffraction, late reverberation and direction of arrival, highlights the effects of comb filtering. In real world

experiences, all these factors reduce the effect of comb filtering between two prominent paths. A more complete room acoustic model might reduce the audibility of comb filtering and be perceived more favourably.

4.5 Participant feedback

After the test, participants were asked what aspects of the stimuli influenced their decisions, with the aim to ascertain a broader picture of important criteria for diffraction models. Most participants said that the frequency spectrum was an important characteristic, with dominance of low frequencies preferable when at large bending angles. For dynamic sources, participants said that the transition across the shadow boundary impacted their ratings. A few participants said that the phasing (comb filtering) effects in some stimuli had a significant negative effect on their ratings, and some participants commented that the test was conceptually challenging as they found it difficult to imagine the given geometry. However, a majority of participants reported that the choice of stimuli aided in the task which suggests that aiming to replicate common experiences is useful when conducting tests of this nature.

5 Conclusions and future work

This paper presented an evaluation of the perceived naturalness of state-of-the-art diffraction models suitable for use in real-time rendering of acoustics in XR applications. Four existing models and a proposed model for edge diffraction were compared perceptually: Biot-Tolstoy-Medwin, Uniform Theory of Diffraction, Volumetric Diffraction and Transmission, a low-order IIR filter model (IIR_{lo}) and a proposed high-order IIR filter (IIR_{hi}). The stimuli were compared in terms of perceived naturalness for the simple case of a single edge in free field using both static and dynamic scenarios. The results indicated that (a) there is no statistically significant difference in naturalness between the proposed IIR_{hi} model and the BTM model, (b) there is no statistically significant difference in naturalness between the UTD model and the BTM model, (c) there is no statistically significant difference in naturalness between the VDaT model and the BTM model at large bending angles, (d) increasing the order of the IIR approximation leads to a statistically significant increase in naturalness at large bending angles, but not near the shadow boundary, and (e) despite comb filtering being physically present in the tested dynamic scenarios, it is perceived negatively by participants. The results support the use of efficient models based on IIR-based approximations. They also support the assumptions made in the VDaT model [8], that not modelling comb filtering between diffraction paths is preferable in dynamic scenes.

While the simple scenarios considered here are more akin to outdoor conditions, future work is needed to assess the models' performances in more complex scenarios such as fully auralised room acoustics with multiple edges, reverberation, higher-order diffraction paths and binaural rendering. The computational complexity of the infinite impulse response approximations increases with the order, and a computational complexity analysis is needed to contrast it with the improvement in naturalness, and to explore optimal trade-offs. Also, considering that IIR_{lo} was shown to be the more natural at small bending angles but less natural at large bending angles, adaptive strategies could be designed to adjust the order according to the bending angle. A further promising direction is to generalise the IIR approximation models to higher order diffraction.

Acknowledgments

This work was supported in part by the Engineering and Physical Science Research Council under grant EP/V002554/1 "SCalable Room Acoustics Modeling (SCReAM)". The authors would like to thank those who participated in the listening test.

References

[1] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty Years of Artificial Reverberation," *IEEE Trans. Audio, Speech, Language Proc.*, vol. 20, no. 5, pp. 1421–1448 (2012).

[2] T. B. Atalay, Z. S. Gul, E. De Sena, Z. Cvetković, and H. Hacıhabiboğlu, "Scattering Delay Network Simulator of Coupled Volume Acoustics," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 30 (2022).

[3] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel, "More Than 50 Years of Artificial Reverberation," presented at the *Audio Eng. Soc. 60th Int. Conf. on Dereverberation and Reverberation of Audio, Music, and Speech* (2016).

[4] L. Savioja and U. P. Svensson, "Overview of Geometrical Room Acoustic Modeling Techniques," *J. Acoust. Soc. Amer.*, vol. 138, no. 2, pp. 708–730 (2015).

[5] P. T. Calamia and U. P. Svensson, "Fast Time-Domain Edge-Diffraction Calculations for Interactive Acoustic Simulations," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1 (2006).

[6] R. Kouyoumjian and P. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *Proc. IEEE*, vol. 62, no. 11, pp. 1448–1461 (1974).

[7] C. Schissler, G. Mückl, and P. T. Calamia, "Fast Diffraction Pathfinding for Dynamic Sound Propagation," *ACM Trans. Graphics*, vol. 40, no. 4 (2021).

[8] L. Pisha, S. Atre, J. Burnett, and S. Yadegari, "Approximate Diffraction Modeling for Real-Time Sound Propagation Simulation," *J. Acoust. Soc. Amer.*, vol. 148, no. 4, pp. 1922–1933 (2020).

[9] C. Kirsch and S. D. Ewert, "Low-Order Filter Approximation of Diffraction for Virtual Acoustics," presented at the *IEEE Workshop Appl. of Signal Proc. to Audio and Acoustics*, pp. 341–345 (2021).

[10] M. A. Biot and I. Tolstoy, "Formulation of Wave Propagation in Infinite Media by Normal Coordinates with an Application to Diffraction," *J. Acoust. Soc. Amer.*, vol. 29, no. 3, pp. 381–391 (1957).

[11] U. P. Svensson and P. T. Calamia, "Edge-Diffraction Impulse Responses Near Specular-Zone and Shadow-Zone Boundaries," *Acta Acust. united Ac.*, vol. 92, no. 4, pp. 501–512 (2006).

[12] H. Medwin, "Shadowing by Finite Noise Barriers," *J. Acoust. Soc. Amer.*, vol. 69, no. 4, pp. 1060–1064 (1981).

[13] H. Medwin, E. Childs, and G. M. Jebsen, "Impulse Studies of Double Diffraction: A Discrete Huygens Interpretation," *J. Acoust. Soc. Amer.*, vol. 72, no. 3, pp. 1005–1013 (1982).

[14] U. P. Svensson, R. I. Fred, and J. Vanderkooy, "An Analytic Secondary Source Model of Edge Diffraction Impulse Responses," *J. Acoust. Soc. Amer.*, vol. 106, no. 5, pp. 2331–2344 (1999).

[15] T. Lokki and V. Pulkki, "Measurement and Theoretical Validation of Diffraction from a Single Edge," presented at the *Proc. Int. Congress on Acoustics*, pp. 929–932 (2004).

[16] A. Løvstad and U. P. Svensson, "Diffracted Sound Field from an Orchestra Pit," *Acoust. Sci. Technol.*, vol. 26, no. 2, pp. 237–239 (2005).

[17] H. Kuttruff, *Room Acoustics* (Taylor & Francis Group, Boca Raton, Florida, USA, 2016).

[18] J. B. Keller, "Geometrical Theory of Diffraction," *J. Optical Soc. of Amer.*, vol. 52, no. 2, pp. 116–130 (1962).

[19] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom, "Modeling Acoustics in Virtual Environments Using the Uniform Theory of Diffraction," presented at the *Proc. Conf. on Comp. Graphics and Interactive Techniques*, pp. 545–552 (2001).

[20] C. Schissler, R. Mehra, and D. Manocha, "High-Order Diffraction and Diffuse Reflections for Interactive Sound Propagation in Large Environments," *ACM Trans. Graphics*, vol. 33, no. 4 (2014).

[21] H. Wallach, E. B. Newman, and M. R. Rosenzweig, "A Precedence Effect in Sound Localization," *J. Acoust. Soc. Amer.*, vol. 21, no. 4, pp. 468–468 (1949).

[22] U. P. Svensson, "Edge Diffraction Matlab Toolbox (EDtoolbox)," Available at: https://github.com/upsvensson/Edge-diffraction-Matlab-toolbox (2022).

[23] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating Interactive Virtual Acoustic Environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705 (1999).

[24] V. Valimaki and T. I. Laakso, "Principles of fractional delay filters," presented at the *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 6, pp. 3870–3873 (2000).

[25] ODEON, "ODEON Room Acoustics Software," Available at: https://odeon.dk/ (2022).

[26] B. Series, "Method for the subjective assessment of intermediate quality level of audio systems," *ITU Radiocomm. Assembly* (2014).

[27] S. Djordjevic, H. Hacihabiboglu, Z. Cvetkovic, and E. De Sena, "Evaluation of the Perceived Naturalness of Artificial Reverberation Algorithms," presented at the *148th AES Convention* (2020).

[28] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application* (John Wiley & Sons, Inc., Chichester, England, 2006).

[29] C. Mendonça and S. Delikaris-Manias, "Statistical Tests with MUSHRA Data," presented at the *Proc. 144th Audio Eng. Soc. Convention* (2018).

[30] R. R. Torres, U. P. Svensson, and M. Kleiner, "Computation of Edge Diffraction for More Accurate Room Acoustics Auralization," *J. Acoust. Soc. Amer.*, vol. 109, no. 2, pp. 600–610 (2001).

[31] R. Torres, N. de Rycker, and M. Kleiner, "Edge Diffraction and Surface Scattering in Concert Halls: Physical and Perceptual Aspects," *J. Temporal Design in Architecture and the Environment*, vol. 4, no. 1, pp. 52–58 (2004).