Extending X3D Realism with Audio Graphs, Acoustic Properties and 3D Spatial Sound

Eftychia Lakka
Faculty of Computing, Engineering and Science,
University of South Wales, UK
efilakka@gmail.com

Don Brutzman
Naval Postgraduate School (NPS),
Monterey California, USA
brutzman@nps.edu

Richard Puk
Intelligraphics Inc.
puk@igraphics.com

Athanasios G. Malamos
Hellenic Mediterranean University, Crete Heraklion,
Greece
amalamos@hmu.gr

ABSTRACT

A fundamental requirement for both realistic modeling and immersive presence is spatial audio, correctly rendering the presentation of each object with aural characteristics. Auditory attributes involve perceived directions, distances, and the propagation paths from complex sound sources to each listener in a potentially complex 3D scene. Many long-running efforts in both hardware and software have improved the ability to aurally render high-fidelity spatialized sound in real time. Motivated by current progress, this work proposes integrating acoustic properties associated with geometric shapes together with 3D spatial sound in the X3D Graphics standard. This combination is possible by exploiting the structure and functionality of Web Audio API, an effective framework for processing and synthesizing audio in Web applications. The paper describes design patterns and initial implement work for this spatial composition of audio graphs and scene graphs. Both specifications are device neutral, without dependencies on specific platforms or audio hardware. Examples for evaluation lead to useful conclusions and areas for future model development.

CCS CONCEPTS

• **Information systems** → **Web applications**; • **Computing methodologies** → Virtual reality; • **Theory of computation** → Semantics and reasoning.

KEYWORDS

Spatial sound, X3D, 3D sound, 3D scene, 3D modeling, auralization, acoustic properties, virtual environments, W3C Web Audio API

ACM Reference Format:
Eftychia Lakka, Don Brutzman, Richard Puk, and Athanasios G. Malamos. 2020. Extending X3D Realism with Audio Graphs, Acoustic Properties and 3D Spatial Sound. In *The 25th International Conference on 3D Web Technology (Web3D '20), November 9–13, 2020, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3424616.3424709

Web3D '20, November 9–13, 2020, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8169-7/20/11…$15.00
https://doi.org/10.1145/3424616.3424709

1 INTRODUCTION

A growing interest in Virtual Environment (VE) technologies can be seen in a wide variety of different applications, including distance learning (e.g., serious gaming), the entertainment industry (e.g., online games, live events), architectural design, the production of art and in various training scenarios [Robotham et al. 2019]. A major part of relevant recent literature has focused on improving the realism within a composed three-dimensional (3D) scene, taking into account both graphics rendering and spatial sound propagation. Since spatial auditory realism increases immersion, users becomes more immersed in an experience when surrounded by audible 3D scene components. Realistic audio presentation also provides a better interactive experience, since audio effects are emanating from the corresponding sound source points [Sean 2017].

Sound in Web-based virtual environments (VEs) is a complex topic spanning multiple technologies and numerous domains of study. Recognizing and correctly aligning these various aspects is essential for composing multiple capabilities meaningfully, coherently and repeatably.

The main objective of the current research is to increase the realism of already existing 3D environments through enrichment with sound characteristics. In parallel, the immersive internet is moving in the same direction. However, no systematic work has been carried out to expand the field of immersive sound beyond the limits of the current technology of VEs, utilizing forthcoming Web infrastructure.

An interactive VE can be evolved with the addition of auralization effects and by taking into account the geometry of the scene. The applicable technology for the implementation of the proposed Web environment is Version 4 of the Extensible 3D (X3D) Standard, revising the spatial sound architecture originally defined in the Virtual Reality Modeling Language (VRML) [Daly and Brutzman 2007], [X3D4 Working Draft 2 [n.d.]]. This work demonstrate an innovative method to extend the X3D specification, including spatial sound attributes, matching the structure of Web Audio API [Recommendation 2020], and adding acoustic sound properties, which are involved in sound propagation [Brutzman 2019].

The remainder of this paper is organized as follows: Section 2 describes the motivation of the proposed approach. Section 3 provides the background and related works regarding the spatial sound in Web 3D. Section 4 details the methodology of the proposed X3D extension. Section 5 gives a brief overview of the testing and the performance evaluation. Finally, section 6 features the concluding remarks and pointers to future work.

2 MOTIVATION

The importance of spatial sound propagation in 3D scenes is widely recognized and seldom achieved. Improved world modeling and improved user interaction are possible through more realistic aural simulation in 3D environments. Taking into account that both Web Audio API and X3D have a fundamental role in the area of VEs, the enrichment of X3D with spatial sound features found in the Web Audio API holds great appeal. Both the X3D Graphics International Standard and the Web Audio API has been influenced by spatial design considerations in OpenAL [OpenAL 2000]. Namely, the Web Audio API has borrowed many concepts from OpenAL such as position/orientation of sources and listeners, parameters associated with sound-source audio cones, and allowing relative motion between sources and listeners. Structures found in OpenAL and Web Audio API are compatible with the original design of existing X3D sound nodes, allowing backwards compatibility with existing models and providing a major advantage for incorporating improved spatial sound in X3D.

Besides that, in this paper, an innovative solution is designed, as new audio nodes in X3D are introduced , matching with Web Audio API, but this approach offers transparent sound design to the programmer and moreover the applications are not independent of said API. This means that the specification of X3D nodes are harmonized with the corresponding Web Audio API nodes, while the 3D content management through the DOM elements can be implemented with any audio library. Except that, this work introduces the registration of new node which includes physical effects (more details in sub-section 3.1 Spatial Sound) involved in sound propagation such as surface reflection (specular, diffusion) and wave phenomena (refraction, diffraction), by taking into account the geometry of the scene increasing the perceived audio fidelity of virtual scene acoustics.

3 BACKGROUND

This section details the background and related work of the spatial sound in Web 3D scenes. This includes a review of the more recent literature in the sound propagation. Furthermore, studies are highlighted that are focused in synthesizing and processing high quality audio in Web 3D environments. Finally, spatial audio and attributes in X3D are presented in the last subsection.

3.1 Spatial Sound

In general,the meaning of spatial sound lies in the way that the listener can recognize meaningful spatial cues from a sound source, for example the direction, the distance and the spaciousness. It is referred to the rendering of spatial attributes of each auditory object, which should be conceived in the same way as they are recognized in the real world, in order to improve the realism and the sense of the immersivity in the 3D scene.

From the theoretical point of view, the spatial hearing in 3D space is inextricably linked with the capability of localizing the sound source [JianJun 2017]. This procedure requires the determination of the location of a sound source in regards to the azimuth (the horizontal plane), elevation (the vertical plane), and distance. In other words, it can be depicted as a 3D co-ordinate system, which includes an x-axis (horizontal), y-axis (vertical) and z-axis (distance), and sounds can be located anywhere in this world-space [Beig et al. 2019]. Furthermore, spatial cues for sound localization can be categorized according to polar coordinates. In fact, each coordinate is considered to have one or more dominant cues in a certain frequency range associated with a particular body component such as:(i) Azimuth and distance cues at all frequencies are associated with the head. (ii) Elevation cues at high frequencies are associated with the pinnae. (iii) Elevation cues at low frequencies are associated with torso and shoulders [Spagnol et al. 2018].

At the same time, the introduction of physical attributes, in spatial auditory, can expand it to simulate more sophisticated and complicated applications. Thus, characteristics such as surface reflection (specular, diffusion) and wave phenomena (refraction, diffraction) can be included for the formation of spatial impressions of a virtual 3D scene. Figure 1 provides an overview of the physical models of sound propagation that are considered; they are highlighted as following:

3.1.1 Specular and diffuse reflection. During the propagation of a sound wave in an enclosed space, the wave hits objects or room boundaries and its free propagation is disturbed. Moreover, during this process, at least a portion of the incident wave is thrown back, a phenomenon known as reflection. If the wavelength of the sound wave is small enough with respect to the dimensions of the reflecting object and large compared with possible irregularities of the reflecting surface, a specular reflection occurs. This phenomenon is illustrated in Figure 1 (inset a), in which the angle of reflection is equal to the angle of incidence. In contrast, if the sound wavelength is comparable to the corrugation dimensions of an irregular reflection surface, the incident sound wave is scattered in many directions. In this case, the phenomenon is called diffuse reflection and is illustrated in Figure 1 (inset b).

3.1.2 Refraction. It is the change in the propagation direction of waves when they obliquely cross the boundary between two mediums where their speed changes, as shown in Figure 1 (inset c). For transmission of a plane sound wave from air into another medium, the refraction index in equation (1) - Snell's Law is used, for calculating the geometric conditions.

$$n = \frac{c'}{c} = \frac{sin\theta'}{sin\theta}, \tag{1}$$

where c' and c the sound speed in the two media, θ the angle of incidence and θ' the angle of refraction.

3.1.3 Diffraction. The fact that a listener can hear sounds around corners and around barriers involves a diffraction model of sound. It is the spread of waves around corners, behind obstacles or around the edges of an opening as illustrated in Figure 1 (inset d). The amount of diffraction increases with wavelength, meaning that

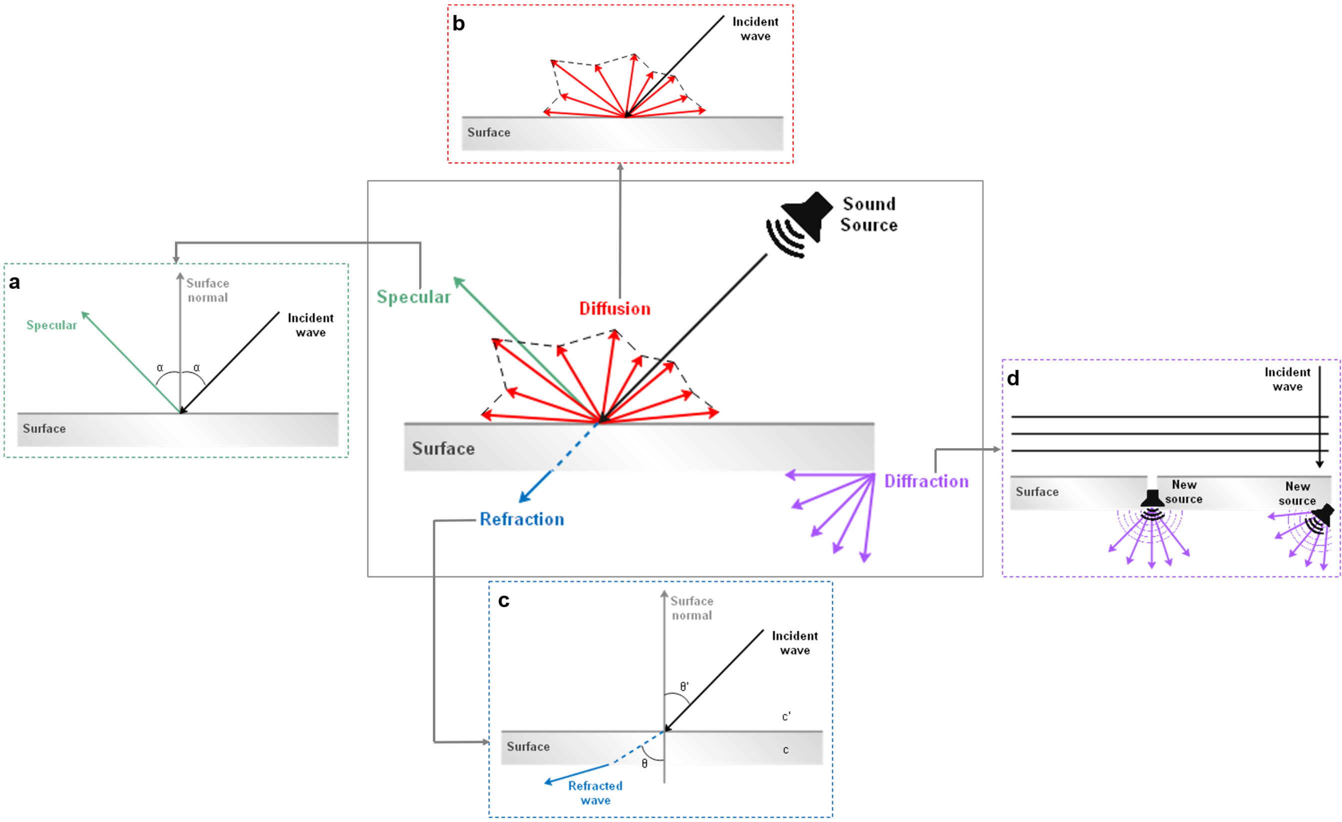

Figure 1: Overview of Sound Propagation Phenomena (Specular and Diffuse Reflection, Refraction, Diffraction)

sound waves with lower frequencies, and thus with greater wavelengths than obstacles or openings dimensions, is spread over larger regions behind the openings or around the obstacles.

In the final analysis, the most effective algorithms for the spatial sound propagation can be classified into three main approaches, acoustic wave equation methods, geometric methods and hybrid methods. The first group is based on the numerical solution of the acoustic wave equation (2).

$$\frac{\theta^2 p}{\theta t} - c^2 \nabla^2 p = F(x, t), \tag{2}$$

where p(x,t) is a time-varying pressure field in space, c is the speed of sound and F(x,t) is the forcing term corresponding to sound sources present in the scene [Manocha and Lin 2009].

The second group, geometric methods, represent acoustic waves as rays and can accurately model the early reflections. They compute propagation paths from a sound source to the listener and overcome the main drawbacks of the previous technique, which are computationally too expensive for interactive applications, and are limited to static scenes only. Finally, the hybrid methods have been developed by combining the previous sound propagation methods to take the advantages of both of them, in order to generate realistic sound effects and succeed significantly reduced calculation times. A detailed review of these methods is beyond the scope of this paper. However, a detailed overview is available by [Lakka et al. 2018].

3.2 Web Spatial Sound

As was mentioned in the Introduction, the spatial sound has a significant role in the 3D environments, due to the fact that it can provide highly realism scenes. For that reason, widespread attention has been paid to introduce sound in Web application. The first approach was accomplished by using <bgsound> tag [Cheng-Shing et al. 2005]. However, it introduced only background music and was available only for the Internet Explorer browser. Following this, the <embed> tag was developed by Netscape, which bore a close resemblance to <bgsound>, providing basically similar functionality. After that, flash was the first cross-browser way of audio on the Web, but it required the use of plugins. Hence, the focus of later research was concentrated on the element <audio> in HTML5, which provided native support for audio playback in all modern browsers. It could avoid the plugins, but was not designed for sophisticated and ambitious developments, because it was inferior to provide filters, position or direction of sources/listeners and low-latency precise-timing model.

Consequently, the prospects of later research were to deal with the above limitations. Various approaches have been proposed for Web sound. Namely, Web Audio API is one of the most popular; it has been proposed by Mozilla Foundation and its predecessor was the Audio Data API. It is a high-level JavaScript API and is characterized as an extremely powerful tool for controlling audio in the browser. Its strong point arises from the fact that a significant number of Web audio libraries and APIs have been developed, in order to use or extend it ([Ricardo et al. 2010], [Simpson [n.d.]], [Mann 2015], [Ikeda 2012]). In parallel, the recent literature demonstrates a variety of works which use Web Audio API to accomplish the sound in Web environment such as [Nikita and Hyunkook 2019], [Queiroz and Feitosa 2019], [Comunità et al. 2019] and [Lee and Reiss 2020]. Additionally, the group of Zlatinsti [Zlatintsi et al. 2018] developed a Web-based real-time application, in order to facilitate gestural interaction between players and virtual musical instruments using Web Audio API. Thomson and Fazekas [Thompson and Fazekas 2019] presented a front-end framework for interactive Web applications built on the Web Audio API. Also, a Web-based real-time interactive blackboard that can be used in a Web-based tutoring system was implemented in [Aburajab and Salman 2019].

Therefore, Web Audio API became attractive for the spatial audio in Web environments, as it is open source and be supported from the browsers. Moreover, it provides multi-channel audio integrating with Web Real-Time Communications (WebRTC). The high-level sound abilities as filters, delay lines, amplifiers, spatial effects (such as panning) are offered. Furthermore, the audio routing graph structure using nodes gives it the advantage of compositionality. Taken together all the above, Web Audio API is currently considered as the most effective Web audio processing framework in comparison with the other proposals.

3.3 X3D Spatial Sound

The current literature shows that insufficient efforts have been done in order to incorporate the spatial sound in X3D. Particularly, two abstract nodes are included in the X3D, the first is about the sound description (X3DSoundNode) and the other one for the sound source (X3DSoundSourceNode). In fact, Sound node is designed for the description of the X3D scene sounds and AudioClip node specifies audio data that can be referenced by Sound nodes. For the audio spatialization, only a boolean field has already introduced to specify if the sound is perceived as being directionally located relative to the viewer. The first interesting approach proposed by [Bouras et al. 2008], in which a Session Initiation Protocol (SIP) 3D spatial audio server was developed using X3DSoundNode interface, in order to support 3D spatial audio. Another valuable research, on this issue, provided an extension of the X3D standard to represent the sound process and activation profile model for providing a rich audio–graphic description in X3D [Ding et al. 2011], however, it was not been generally accepted in order to be embedded in the X3D specification. Finally, our previous work has been offered satisfactory results but X3DOM [X3DOM 2009] oriented, which is a framework that declares 3D content including X3D elements as part of any HTML5 DOM tree.

Obviously, the implementation of spatial sound in X3D is still lacking, even though the spatial sound should be an integral part of an immersive 3D application. In order to overcome this drawback, this work presents an innovative solution of the spatial sound in X3D framework that based on the structure of Web Audio API. The next section highlights the motivation of this work in detail.

4 EXTEND X3D WITH SPATIAL SOUND NODES

This section is devoted to the overview of the proposed design. The main idea is to enhance the spatial representation of a 3D sound with effects to provide a more realistic and immersive auditory environment. In fact, the aim is to provide a procedure in order to define the audio flow from scene source (often an audio file) to its destination (often speakers) through sound effects. For that reason, the proposed approach is constituted by the corresponding three categories (see Table 1). It is worthwhile noting that the new SpatialSound node can be fundamental for realistic and interactive Web3D sound applications as it is responsible for the spatialized sound. The MicrophoneSource is an important new node added for the common use case to replace an AudioClip using a connected microphone or capture input from a physical microphone. Hence, this section includes in the first place a subsection with the registration of the new spatial sound nodes matching with the Web Audio API; the following subsection describes the new X3D node with properties influencing sound propagation include surface-related physical phenomena. The last subsection overviews the additional considerations.

4.1 Design and Naming Conventions

A typical workflow for Web Audio API includes creating a graph from three basic categories of nodes: source, effects and destination. An audio application can be complex, containing many nodes between the sources and destinations, in order to produce advanced synthesis or analysis. The design proposed here builds audio graphs by similarly connecting these three types of nodes.

X3D nodes are defined in an object-oriented fashion, and X3D naming conventions follow a specific design pattern. Names for concrete instantiable nodes are kept simple and functional (e.g., ImageTexture) with variations expressed as first term (e.g., PixelTexture, MovieTexture). Abstract node types begin with "X3D" and end with "Node" (e.g., X3DTexture-Node). Since many W3C Audio API classes end with "Node" as well, distinct names are needed for clarity even when semantics are identical. Name correspondences are shown in Table 1.

4.1.1 Source Type Nodes. A set of new nodes were registered in the X3D, in order to define the audio stream sources in our proposal. To enumerate, in the abstract node X3DSoundSourceNode, which is used to derive node types that can emit audio data, new nodes were added:

(1) AudioBufferSource represents a memory-resident audio asset. Its functionality is a combination of Web Audio API AudioBuffer , AudioBufferSourceNode nodes. It includes buffer field which is a data block holding the audio sample data, duration, lenght and sampleRate of the Pulse Code Modulation (PCM) audio data; numberOfChannels field for the number of discrete audio channels, detune to modulate the speed at which is rendered the audio stream and playbackRate represents the speed at which to render the audio stream. Finally, the fields loop, loopEnd, loopStart are used for the loop of playback.

Table 1: Our Implementation for the Registration of Spatial Sound nodes in X3D matching with the corresponding Web Audio API nodes

Category	X3D new Node	Web Audio API
source	AudioBufferSource	AudioBuffer , AudioBufferSourceNode
	OscillatorSource	OscillatorNode
	StreamAudioSource	MediaStreamAudioSourceNode
	MicrophoneSource	-
effects	BiquadFilter	BiquadFilterNode
	Convolver	ConvolverNode
	Delay	DelayNode
	DynamicsCompressor	DynamicsCompressorNode
	Gain/integrated field	GainNode
	WaveShaper	WaveShaperNode
	PeriodicWave	PeriodicWave
	Analyser	AnalyserNode
	ChannelSplitter	ChannelSplitterNode
	ChannelMerger	ChannelMergerNode
	SpatialSound	PannerNode
destination	AudioDestination	AudioDestinationNode
	StreamAudioDestination	MediaStreamAudioDestinationNode
	ListenerPoint	Audiolistener

(2) OscillatorSource represents an audio source generating a periodic waveform, providing a constant tone. It is composed by detune field, which offsets the frequency by the given amount, frequency of the periodic waveform; type field for the shape of the periodic waveform ("sine": a sine wave, "square": a square wave of duty period 0.5, "sawtooth": a sawtooth wave, "triangle": a triangle wave, "custom": a custom periodic wave).
(3) StreamAudioSource operates as an audio source whose media could be from a microphone or from a remote peer on a WebRTC call. It has mediaStream field which is responsible for a memory-resident audio asset (for one-shot sounds and other short audio clips).
(4) MicrophoneSource captures input from a built-in (physical) microphone. The mediaDeviceID field that is included in this node, is a unique identifier for the represented device and isActive attribute is a boolean value that returns true if the device is active, or false otherwise.

4.1.2 Effects Type Nodes. For this category, three new abstract nodes were registered in X3D. Namely, the first is the X3DSoundProcessingNode which is the base for all sound processing nodes, which are used to enhance audio with filtering, delaying, changing gain, etc and is the parent from nodes:

(1) BiquadFilter represents different kinds of filters, tone control devices, and graphic equalizers. The detune field is a detune value, for the frequency, the frequency field is the frequency at which the said node operates, the gain field is the amplitude gain of the filter, the Q field is Quality Factor (Q) of the filter and the last field, type, determines the filters that can be used to achieve certain kinds of effects ("lowpass": makes sounds more muffled, "highpass": makes sounds more tinny, "bandpass": cuts off lows and highs (e.g., telephone filter), "lowshelf": affects the amount of bass in a sound, "highshelf": affects the amount of treble in a sound, "peaking": affects the amount of midrange in a sound).
(2) Convolver performs a linear convolution on the audio stream. It has the buffer field to represent a memory-resident audio asset and the normalize field which controls whether the impulse response from the buffer is scaled by an equal-power normalization when the buffer attribute is set, or not.
(3) Delay causes a time delay between the arrival of input data and subsequent propagation to the output using the delayTime field.
(4) DynamicsCompressor implements a dynamics compression effect, lowering the volume of the loudest parts of the signal and raises the volume of the softest parts. It includes the attack field which is the amount of time (in seconds) to reduce the gain by 10dB, the knee field which contains a decibel value representing the range above the threshold where the curve smoothly transitions to the compressed portion. Additionally, the ratio field represents the amount of change, in dB, needed in the input for a 1 dB change in the output, the reduction field represents the amount of gain reduction currently applied by the compressor to the signal, the release field represents the amount of time (in seconds) to increase the gain by 10dB and the threshold field represents the decibel value above which the compression will start taking effect.
(5) Gain represents a change in volume and includes the corresponding field (gain) to control the applied amount of gain.
(6) WaveShaperdescribes a non-linear distorter. Its curve field is an array of floats numbers describing the distortion to apply,

the oversample field specifies what type of oversampling (if any) should be used when applying the shaping curve.

(7) PeriodicWave defines a periodic waveform that can be used to shape the output of an OscillatorSource node.

Secondly, the X3DSoundAnalysisNode is registered that is the base for Analyser node, which receives real-time generated data, without any change from the input to output sound information. Through this process, the audio visualization of the sound source is produced. The Analyser provides real-time frequency and time-domain analysis information, without any change to the input. It is composed by the fftSize field for the size of the FFT (Fast Fourier Transform) [1] to be used to determine the frequency domain, the frequencyBinCount field which generally equates to the number of data values that play with for the visualization. Moreover, the minDecibels field describes the minimum power value in the scaling range for the FFT analysis data, the maxDecibels field represents the maximum power value in the scaling range for the FFT analysis data and the smoothingTimeConstant field is used to represent the averaging constant with the last analysis frame.

Lastly, the X3DSoundChannelNode abstract node is introduced as the base for nodes that handle of channels in an audio stream, allowing them to be split and merged. In detail, it is consisted of ChannelSplitter and ChannelMerger node. The first separates the different channels of an audio source into a set of mono outputs and the second the opposite.

Except for the above abstract nodes and their children nodes, SpatialSound node belongs in the effects category. It is the most fundamental node for the sound spatialization, as it represents a processing node which positions, emits and spatializes an audio stream in 3D space. The spatialization is in relation to the ListenerPoint (see next subsection Destination Type Nodes). It is register under the abstract node X3DSoundNode. Its functionality is harmonized with the fact that the audio is located at a point in the local coordinate system and emits sound in an elliptical pattern (defined by two ellipsoids). So, the ellipsoids are oriented in a direction specified by the direction field. The distanceModel field specifies which algorithm to use for sound attenuation, corresponding to distance between an audio source and a listener (linear, inverse, exponential). The intensity field adjusts the loudness of the sound, the location field determines the location of the sound emitter in the local coordinate system, the maxDistance field is the maximum distance where sound is renderable between source and listener, after which no reduction in sound volume occurs, the enableHRTF field specifies whether to enable Head Related Transfer Function (HRTF) auralization. Furthermore, the fields coneInnerAngle, coneOuterAngle and coneOuterGain are related directional cone model of sound (see Figure 2, Figure 3). Additionally, referenceDistance field is a reference distance for reducing volume as source moves further from the listener and the rollOffFactor describes how quickly the volume is reduced as source moves away from listener. The X3D node declaration is following:

[1] https://en.wikipedia.org/wiki/Fast_Fourier_transform

```
SpatialSound : X3DSoundNode {
 SFFloat   [in,out]  coneInnerAngle    6.2832  [0, 2π]
 SFFloat   [in,out]  coneOuterAngle    6.2832  [0, 2π]
 SFFloat   [in,out]  coneOuterGain     0       (-∞, ∞)
 SFVec3f   [in,out]  direction         0 0 1   (-∞, ∞)
 SFString  [in,out]  distanceModel     "INVERSE"
 SFBool    [in,out]  enableHRTF        FALSE
 SFFloat   [in,out]  intensity         1       [0, 1]
 SFVec3f   [in,out]  location          0 0 0   (-∞, ∞)
 SFFloat   [in,out]  maxDistance       10000   [0, ∞)
 SFNode    [in,out]  metadata          NULL
 SFFloat   [in,out]  referenceDistance 1       [0, ∞)
 SFFloat   [in,out]  rolloffFactor     1       [0, ∞)}
```

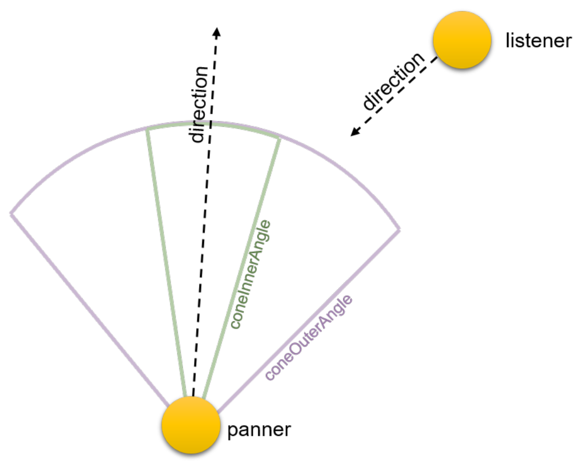

Figure 2: Cone Model of Sound: a diagram of panner and the listener in 2D space

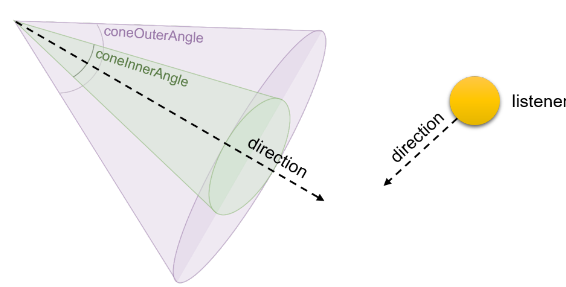

Figure 3: Cone Model of Sound: a diagram of panner and the listener in 3D space

Existing stereo X3D sound attenuation is dependent on a panning factor dependent on viewer gaze direction as shown in Figure 4 and this planar gain-reduction relationship depends on listener orientation, pertaining to relative direction of current viewer and also any additional ListenerPoint nodes.

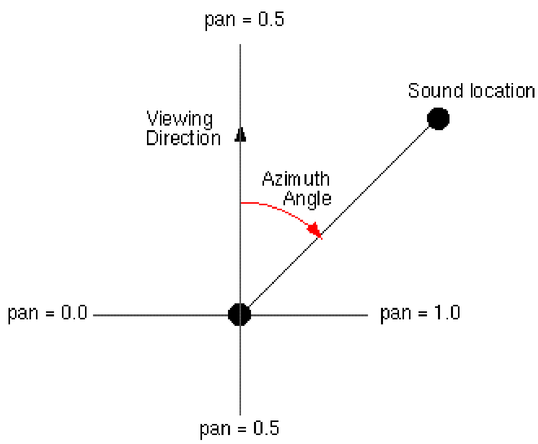

Figure 4: Stereo panning - Reprinted from: https://www.web3d.org/documents/specifications/19775-1/V3.3/Part01/components/sound.html

4.1.3 Destination Type Nodes. In this category of new registered nodes, one abstract node is proposed, X3DSoundDestinationNode that includes the AudioDestination (with field maxChannelCount that decribes maximum number of channels) and StreamAudioDestination (with field stream that represents a memory-resident audio asset) node and is the base for all sound destination nodes, which represent the final audio destination and are what the user will ultimately hear. They can often considered as audio output devices which are connected to speakers. All rendered audio to be heard will be routed to these "terminal" nodes. Besides the previous, ListenerPoint represents the position and orientation of the person listening to the audio scene. It provides single or multiple sound channels as output. Multiple ListenerPoint nodes can be active for sound processing, but only one can be bound as the active listening point for the user. Its interauralDistance field is used for binaural recording and the trackCurrentView field if it is TRUE matches the position and orientation to the user's current view.

4.2 Extend X3D with AcousticalProperties node

Acoustic effects including surface reflection (specular, diffuse), wave phenomena (refraction, diffraction) and acoustic coefficients of physical materials must be taken into account for proper realism of sound propagation in a virtual environment. For that reason, AcousticProperties has been defined to describe coefficients related to said physical propagation phenomena of sound for various materials. These coefficient values are expected to fully account for physical and structural characteristics of the associated geometry such as width, height, thickness, shape, softness and/or hardness, and density variations. As an illustration, the absorption field specifies the sound absorption coefficient of a surface which is the ratio of the sound intensity absorbed or otherwise not reflected by a specific surface that of the initial sound intensity. This characteristic depends on the nature and thickness of the material. Sound energy is partially absorbed when it encounters fibrous or porous materials, panels that have some flexibility, volumes of air that resonate, and openings in room boundaries (e.g., doorways). Moreover, the absorption of sound by a particular shape depends on the angle of incidence and frequency of the sound wave. The diffuse field describes the diffuse coefficient of sound reflection. The refraction field describes the sound refraction coefficient of a medium, which determines the change in propagation direction of a sound wave when it obliquely crosses the boundary between two mediums where its speed is different. These relationships are described by Snell's Law. Finally, the specular field describes the specular coefficient of sound reflection. Below is the X3D declaration of the node:

```
AcousticProperties : X3DAppearanceChildNode {
 SFFloat   [in,out]  absorption    0     [0, 1]
 SFFloat   [in,out]  diffuse       0     [0, 1]
 SFNode    [in,out]  metadata      NULL
 SFFloat   [in,out]  refraction    0     [0, 1]
 SFFloat   [in,out]  specular      0     [0, 1] }
```

4.3 Additional Considerations

Note that diffraction sources are not explicitly defined. Computational sound geometry engines are expected to handle sound propagation around corners, behind occluding objects, etc. Complex geometric openings may also be modeled by an audio chain including ListenerPoint and SpatialSound to emulate sophisticated diffraction propagation paths.

Computational sound propagation is similar to ray tracing in 3D graphics and can be computationally expensive. Past work has demonstrated that far simpler geometry can provide useful acoustic response (e.g., simple plane instead of cluttered wall). If a simplified geometry alternative from Collision proxy field is available, it is used preferentially by collision-detection algorithms for sound propagation, rather than descendant children of the Collision node. Such geometric simplifications can often reduce computational costs significantly without reduction in perceived audio fidelity of 3D scene acoustics. Based on ongoing work with examples, it is possible that a separate Collision field for acousticProxy field may be appropriate.

Important work in progress is determining the best way to connect X3D nodes making up an audio graph node. Current work is examining whether parent/child SFNode/MFNode relationships are sufficient. Alternatively ROUTE connections might be employed, but at the cost of further detail and possible introduction of error. Creating a large set of example audio graphs that capture the full richness of Web Audio API will reveal whether remaining design issues are resolved satisfactorily.

5 EVALUATION - X3D EXAMPLES

In order to verify the validity of our design, the new registered nodes and the proposed structure were applied to several experiments – examples that were already implemented in previous work

[Stamoulias et al. 2015] continue to be updated. The first example evaluates the attenuation of one sound source, in a 3D scene. The spatial nature of the sound is expressed by the movement when that the user approaching nearby to the sound source then the volume is increased and accordingly when removed therefrom is reduced. Hence, taking to account our new approach, this example can be represented in X3D code (see Example 1):

```
<Transform>
   <SpatialSound
       direction='0 0 1' coneInnerAngle='6.2832'
       coneOuterAngle='6.2832' coneOuterGain='0'
       distanceModel='inverse' enableHRTF=TRUE
       referenceDistance='1'
       rolloffFactor='1'/>
   <AudioClip loop='true' url="sound/africa/beat.mp3"/>
   <ListenerPoint trackCurrentView = true/>
</Transform>
```

The second example assesses the capability of the audio channels split including a simple sound source which can be moved right and left. Depending on the position of the sound source, the user can hear the produced sound from the corresponding output speaker. Accordingly, there is a source that can be passed through a SpatialSound node for the spatialization of the input audio. In the same way, the corresponding X3D code with new registered nodes is following (Example 2).

```
<Transform>
   <SpatialSound
       direction='0 0 1' coneInnerAngle='6.2832'
       coneOuterAngle='6.2832' coneOuterGain='0'
       distanceModel='inverse' enableHRTF=TRUE
       referenceDistance='1'
       rolloffFactor='1'/>
   <AudioClip loop='true' url="sound/africa/beat.mp3"/>
   <ListenerPoint trackCurrentView = true/>
   <BiquadFilter
       frequency='100' detune='10.0'
       Q='10' gain='0' type='lowpass'/>
</Transform>
```

The third and forth example follow the same process as the previous with the difference that manages the filters and the panning from more than one sound sources.

6 CONCLUSION

To sum up, the purpose of our work is to suggest a structure of new nodes in order extent the X3D specification both with spatial sound attributes and with physical effects which are involved in sound propagation, such as surface reflection (specular, diffuse) and wave phenomena (refraction, diffraction). The strong point of our work is the fact that is harmonized with Web Audio API, which is the most effective framework for spatial audio in Web (3D) environments, but it does not depend solely on this, as it can be parsed through others sound libraries. Equally important is the additional of acoustic properties ensuring that the quality of 3D scenes can be increased.

The evaluation of specific examples proves that the interactive virtual scene can be composed with the use of new X3D registered nodes and the results confirm our methodology. As for future work, our approach will be enhanced with further attributes and will be applied in more sophisticated 3D scenes in order to achieve the expansion of the field of immersive sound beyond the limits of the current studies, utilizing the forthcoming virtual applications.

REFERENCES

Ata Aburajab and Adnan Salman. 2019. Interactive Blackboard for Web-based Real-time Tutoring System. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. 63–68.

Mirza Beig, Bill Kapralos, Karen Collins, and Pejman Mirza-Babaei. 2019. An Introduction to Spatial Sound Rendering in Virtual Environments and Games. *The Computer Games Journal* 8 (2019), 199–214.

Christos Bouras, Thrasivoulos Tsiatsos, and Vasileios Triglianos. 2008. A SIP spatial audio server for the EVE platform. In *3rd International Conference on Computer Graphics Theory and Applications, Funchal, Madeira, Portugal*. 395–402.

Don Brutzman. April 14, 2019. *Strategies to Improve X3D v4 Sound Component*. https://www.web3d.org/sites/default/files/page/X3Dv4%20Implementations/X3dAudioExtension_v02.pdf

Lai Cheng-Shing, Shi Chang, and Huang Jun. 2005. Portable communication device and multimedia message processing method for same. US Patent App. 10/844,358.

Marco Comunità, Andrea Gerino, Veranika Lim, and Lorenzo Picinali. 2019. Web-Based Binaural Audio and Sonic Narratives for Cultural Heritage. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. http://www.aes.org/e-lib/browse.cfm?elib=20435

Leonard Daly and Don Brutzman. 2007. X3D: Extensible 3D Graphics Standard. *Elsevier Morgan Kaufmann Publishers* 24, 6 (2007), 130–135. http://x3dgraphics.com

Hui Ding, Diemo Schwarz, Christian Jacquemin, and Roland Cahen. 2011. Spatial Audio–Graphic Modeling for X3D. In *Web3D*. Paris, France, 1–1. https://hal.archives-ouvertes.fr/hal-01161293

Tomohiro Ikeda. 2012. XSound GitHub repository. https://xsound.jp/

He JianJun. 2017. *Literature Review on Spatial Audio*. Springer Singapore, Singapore, 7–37. https://doi.org/10.1007/978-981-10-1551-9_2

Eftychia Lakka, Athanasios G. Malamos, Konstantinos G. Pavlakis, and Andrew J. Ware. 2018. Spatial Sound Rendering–A Survey. *IJIMAI* 5, 3 (2018), 33–45.

Jake Ryan Rajjayabun Lee and Joshua D. Reiss. 2020. Real-Time Sound Synthesis of Audience Applause. *J. Audio Eng. Soc* 68, 4 (2020), 261–272. http://www.aes.org/e-lib/browse.cfm?elib=20732

Yotam Mann. 2015. Tone.js GitHub repository.

Dinesh Manocha and Ming C. Lin. 2009. Interactive sound rendering. In *2009 11th IEEE International Conference on Computer-Aided Design and Computer Graphics*. IEEE, 19–26.

Goddard Nikita and Lee Hyunkook. 2019. A Web-Based Tool for Microphone Array Design and Phantom Image Prediction Using the Web Audio API. In *Audio Engineering Society Convention 146*. http://www.aes.org/e-lib/browse.cfm?elib=20357

OpenAL 2000. *OpenAL*. Retrieved July 20, 2020 from http://www.openal.org/

Jordan S. Queiroz and Eduardo L. Feitosa. 2019. A Web Browser Fingerprinting Method Based on the Web Audio API. *Comput. J.* 62, 8 (01 2019), 1106–1120. https://doi.org/10.1093/comjnl/bxy146 arXiv:https://academic.oup.com/comjnl/article-pdf/62/8/1106/29162322/bxy146.pdf

W3C Candidate Recommendation. 2020. *Web Audio API*. Retrieved July 20, 2020 from https://www.w3.org/TR/2020/CR-webaudio-20200611/

Cabello Ricardo et al. 2010. Three. js. *URL: https://github. com/mrdoob/three. js* 34 (2010).

Thomas Robotham, Olli S. Rummukainen, and Emanuël A. P. Habets. 2019. Towards the Perception of Sound Source Directivity Inside Six-Degrees-of-Freedom Virtual Reality. (2019). https://doi.org/10.22032/DBT.39956

Ong Sean. 2017. *Spatial Sound*. Apress, Berkeley, CA, 141–152. https://doi.org/10.1007/978-1-4842-2769-5_7

James Simpson. [n.d.]. howler.js - JavaScript audio library for the modern web. https://howlerjs.com/

Simone Spagnol, György Wersényi, Michał Bujacz, Oana Bălan, Marcelo Herrera Martínez, Alin Moldoveanu, and Runar Unnthorsson. 2018. Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids. *Wirel. Commun. Mob. Comput.* 2018 (March 2018), 9. https://doi.org/10.1155/2018/3918284

Andreas Stamoulias, Eftychia Lakka, and Athanasios G. Malamos. 2015. "Wrapping" X3DOM around Web Audio API. (2015).

Andrew Thompson and Gyorgy Fazekas. 2019. A model-view-update framework for interactive web audio applications. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 219–222. https://doi.org/10.1145/3356590.3356623

X3D4 Working Draft 2 [n.d.]. *X3D4 Working Draft 2.* Retrieved July 20, 2020 from https://www.web3d.org/specifications/X3Dv4Draft/ISO-IEC19775-1v4-WD2/

X3DOM 2009. - *x3dom.org*. Retrieved July 20, 2020 from https://www.x3dom.org/

Athanasia Zlatintsi, Panagiotis P. Filntisis, Christos Garoufis, Antigoni Tsiami, Kosmas Kritsis, M. A. Kaliakatsos-Papakostas, Aggelos Gkiokas, Vassilis Katsouros, and Petros Maragos. 2018. A Web-Based Real-Time Kinect Application for Gestural Interaction with Virtual Musical Instruments. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion* (Wrexham, United Kingdom) *(AM'18)*. Association for Computing Machinery, New York, NY, USA, Article 12, 6 pages. https://doi.org/10.1145/3243274.3243297