VIVE 3DSP SDK Guide
(Unity Plugin)

Release version: 0.10.0

Table of Contents

System Requirements

- Unity version 5.5 or higher
- Windows® 7 SP1, Windows® 8 or later, Windows® 10
- Intel® Core™ i5-4590 or AMD FX™ 8350 equivalent or better
- NVIDIA® GeForce® GTX 1060 or AMD™ Radeon™ RX 480, equivalent or better graphics card

Why is 3D Audio Needed in VR?

When putting on a VR headset to enjoy a scenario such as shooting at zombies or walking in a forest, people want to immerse themselves extremely in that scene. There may be awesome visuals in the scene, but it will be less immersive without accompanying audio or subpar audio. Audio is crucial to increase immersion and sense of presence in VR.

Before diving into what can be done to improve VR audio experience, the principles of how sound localization is perceived by human beings is firstly introduced.

Localization

Localization is the inherent ability of human beings to figure out where a sound is coming from based on human auditory system. By understanding the human auditory system, real-world spatialized sound is simulated in a VR environment.

These are the factors that help us with localization:

- **Lateral**
  - **Interaural time difference**: The direction of a sound source can be determined by the time difference that a sound arrives to our ears, i.e. if we hear the sound first through our right ear before our left ear, we can more or less be sure that the sound is coming from the right direction.
  - **Interaural level difference**: The direction of a sound source can also be determined by the loudness level that is awared between ears. (i.e. If the sound is louder on the right ear compared to the left ear, we can assume that the sound is coming from the right direction.)
- **Elevation and Front/Back**: In order to distinguish directions which have the same arrival time and level, direction selective cues (pinna, head, shoulder, and torso) are used as the main reliable features. These cues are used to help determine where the sound is coming from. Moreover, people can also cock the heads up or turn our heads left or right to determine the direction of the sound source more precisely.
- **Pinna**: When the sound arrives at the head, it will then be collected and

filtered by the outer ears. These filters provide cues to our brain to determine where the sound is coming from.

- **Head/Shoulder/Torso**: Sound is partially blocked or absorbed by the body before arriving at the ears. It can offer some cues as well.

Room Reverberation

If the user is inside the room, the walls and objects in that room will reflect or scatter the sound waves. The sound waves will bounce again and again, and form a particular audio perceptual experience in that room. It not only builds a realistic feeling of the room, but also improves the accuracy of the directional perception.

Distance

Another aspect of localization is our ability to gauge the distance of a sound source. Here are some of the ways we judge the distance of a sound source.

- **Loudness**: Louder sounds are usually closer while quieter sounds are usually farther. In some cases, it is hard to judge the distance by the loudness if people are not familiar to the sound. We also take into consideration the relative loudness. When sound approaches, it gets louder and vice versa.
- **Initial time delay**: This is the time between the initial sound arriving and the first early reflections. If somebody is standing right next to you in a room and talking to you, you'll hear the voice right away then the reflections of the voice a bit later. If the person is standing farther away, you'll probably hear the voice and reflections about the same time.
- **Mix of direct and indirect sounds**: A sound is judged to be near to the listener when it is louder than the background noise. Otherwise, the background noise would be easier perceived by the listener.
- **Interaural level difference**: If there is an obvious volume difference being perceived between the right and left ears, the sound is most likely to be close to the audience. If the volume difference between the two ears is negligible, the sound is probably to be far away.

Spatialization

Spatialization is an audio process which gives the listener the impression that he/she is inside a virtual 3D environment. Spatialization is a key factor for an immersive VR experience. Lots of features are accomplished to simulate the real spatial feeling in 3DSP SDK, such as:

- Head-related transfer functions (HRTF) recording and improvement
- Higher Order Ambisonic simulation of sound direction
- Room audio simulation
- Adding background noise floor
- Real-world acoustic properties of distance
- Geometric and Raycast Occlusion
- Hi-Res audio support

and many other features

VIVE 3DSP SDK

The VIVE 3DSP SDK provides an audio solution for simulating realistic sounds inside the virtual world. This SDK provides the tools you need to create engaging and immersive spatial audio for your VR apps or experiences.

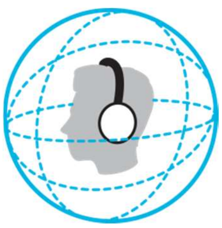
The VIVE 3DSP SDK is made for VR content developers. A host of features are designed and implemented to provide developers a rapid spatial audio solution. The audio quality and audio source localization can be further optimized if the VIVE Pro headset is used. There are some improvements made specifically for the audio system provided in the VIVE Pro headset.

Some of the unique features of the VIVE 3DSP SDK:

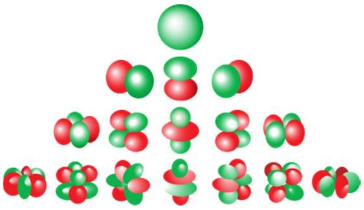
- **Higher Order Ambisonics (HOA)** that uses very low computing power.
- **Head-Related Transfer Function (HRTF)** based on refined real-world modeling (horizontally and vertically) resulting in a better algorithm that is applied to all sound filters and effects.
- **Room Audio** simulates the reflection and reverberation of a real space.
- **Hi-Res Audio Settings** source files and playback.
- **Sound Decay Model** based on real-world modelling.
- **Geometric Occlusion** doesn't use the Unity collider. The blocking area is calculated using its own algorithm.
- **Ambisonic Decoder** supports playing the AmbiX format ambisonic decoder.

The VIVE 3DSP SDK guarantees the best audio experience when developing for VIVE. However, it can be used when developing for other VR headsets as well.

Higher Order Ambisonics (HOA)



Ambisonics is the technology that uses a full-sphere surround sound technique to simulate spatial sound. It counts the sound pressure changes around the listener by spherical harmonics.



The order of the spherical harmonics is also called the order of ambisonics. The order of ambisonics has a big influence on the realistic depiction of the direction of sound. Therefore, 3rd order ambisonics is used as our spatial audio model as it generates spatial sound with better directivity and higher resolution.

The higher the order that is used, the higher computing power is required for the ambisonic process. However, our patent-pending technique that is used on our 3DSP ambisonic process requires low computing power akin to using the first order ambisonic, even though the 3rd order is used.

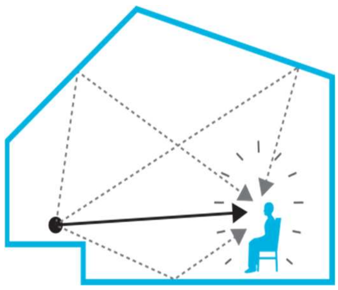
Note: Higher Order Ambisonic technology is only used for object sound spatialization and not for decoding the ambisonic file format. Ambisonic file playback support is in our future plans.

Head-Related Transfer Function (HRTF)

The head-related transfer function is the recordings of the acoustic responses in all directions. There are many factors that could change the audio experience coming from an audio source travelling to the human ears, and it is too complex to simulate by dedicated functions. Therefore, a database has been created where the acoustic response in any direction has been measured and saved. Audio response can then be simulated by using the recorded data from this database.

In the VIVE 3DSP SDK, there is an HRTF database which was measured with fine directional segments, both horizontally and vertically. This means that thousands of directions of HRTF have been measured. For HRTF directions that have not been measured, these could also be generated by the measured directions nearest to it, with the quality benefitting from the fine measurement of the HRTF database. The quality degradation caused by the environment, the measuring equipment, and other factors that affect HTRF have also been improved in our database.

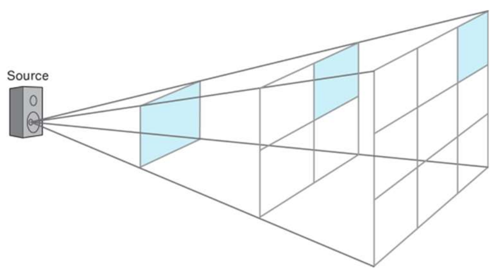
Room Audio



Room Audio is the technology that simulates room audio through the early reflection, late reverberation, background noise, and others. The early reflection is the reflected sounds within a few milliseconds, and we simulate these reflected sounds through delay.

There is also an option to choose the material for each wall (wood, carpet, concrete, and glass) in a room. Different wall materials produce different types of reflections. After the first reflection, the sound encounters more reflections and mixes with the background sound. The resulting sound could not be simulated by large amount of reflections so a well-designed formula is required for generating accurate late reverberation.

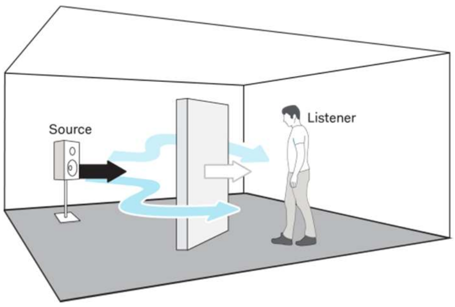
Sound Decay Model



When a sound source approaches or moves away from the user, the sound heard by the user should also change depending on the distance between the sound source and the user. The inverse square law is used for this scenario to determine the change in the sound levels.

Our SDK provides the real-world distance effect which measures the response at different distances between the sound source and listener and then applies the proper response measurement. Also, point source and line source decreasing models have been provided as well.

Occlusion



The occlusion effect is used to accurately simulate what happens to sound when it encounters an obstacle in its path. The resulting sound produced depends on the material (window, thin door, wood wall, stone wall, and curtain) and thickness of the obstacles that the sound encounters in its path.

The VIVE 3DSP SDK provides mono and binaural occlusion mode, while the binaural occlusion mode simulates a better realistic binaural sound in VR. It also provides two kinds of occlusion, Geometric Occlusion and Raycast Occlusion:

- **Geometric Occlusion:** The blocking ability of an object is calculated by analytical geometry techniques. Two shapes are currently provided: Sphere and Box. Since there is no need to use the Unity collider, you can use more resources in other physical properties.
- **Raycast Occlusion:** The blocking ability of an obstacle is calculated by pointing many rays into space and finding which ones are blocked. It can be used for any kind of obstacle shapes but uses more computing resources than Geometric Occlusion.

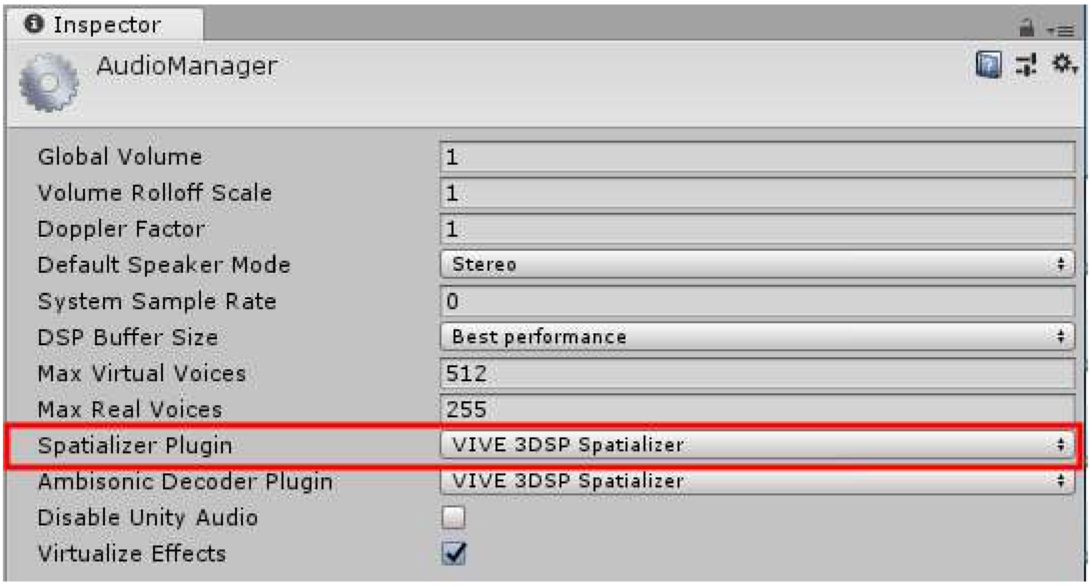
Ambisonic Decoder

Audio recordings for 360 and VR videos are now mostly achieved by sound field microphones in 4-channel B-format. These channels do not translate to a fixed representation of sound. Instead, the ambisonic decoder can interpret and reshape the sound recordings in different ways.
For Unity 2017.1 or later, the VIVE 3DSP Spatializer supports ambisonic audio. You can attach 4-channel AmbiX format (in ACN ordering and SN3D normalization) audio clips to game objects, rotate the headset, and then hear the sound field effect.

VIVE 3DSP SDK Unity Plugin

Set Spatializer Plugin

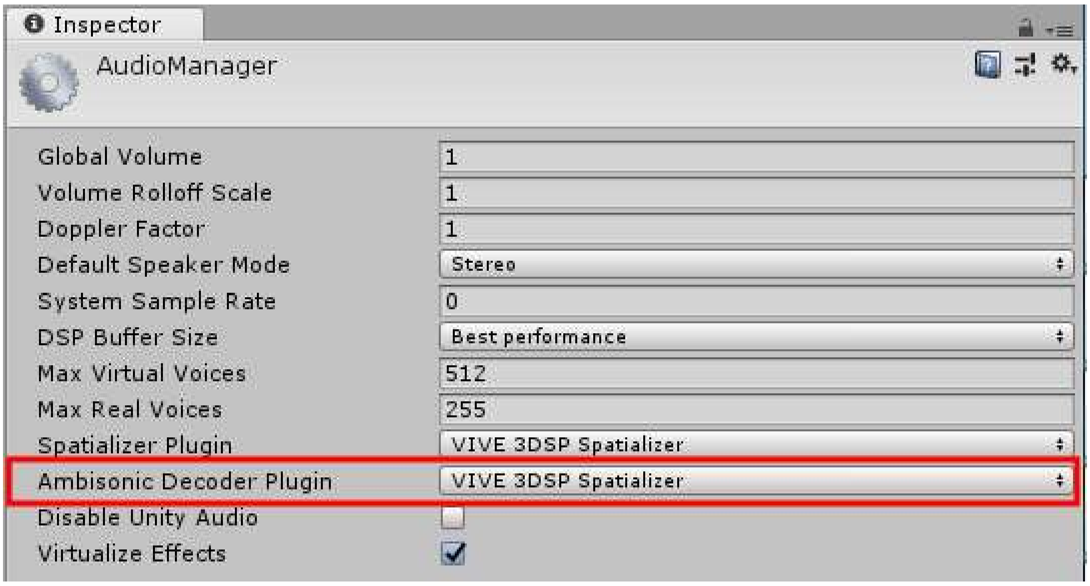
In Unity, set **Spatializer Plugin** to **VIVE 3DSP Spatializer**.



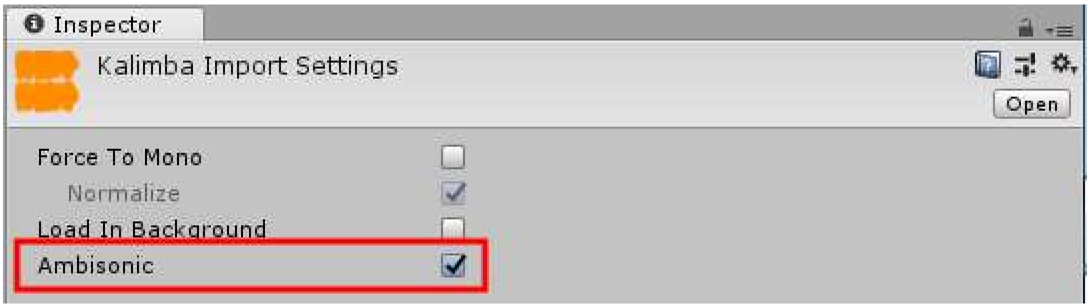
1. Click **Edit > Project Settings > Audio**.
2. Check **AudioManager** option in the **Inspector** window.
3. From **Spatializer Plugin**, select **VIVE 3DSP Spatializer**.

Set Ambisonic Decoder Plugin

In Unity, set **Ambisonic Decoder Plugin** to **VIVE 3DSP Spatializer**.(2017.1 version or later) in Audio Manager.



1. Click **Edit > Project Settings > Audio**.
2. Check **AudioManager** option in the **Inspector** window.
3. From **Ambisonic Decoder Plugin**, select **VIVE 3DSP Spatializer**.



1. Copy your AmbiX format audio file into your Unity Assets folder.
2. Find the file in the **Project** window and click it.
3. Select **Ambisonic** in the **Inspector** window and click **Apply**.

Note: Make sure **Spatialize** in Audio Source is not selected.

Audio Source Settings

Set an object as the audio source

a. Do one of the following:

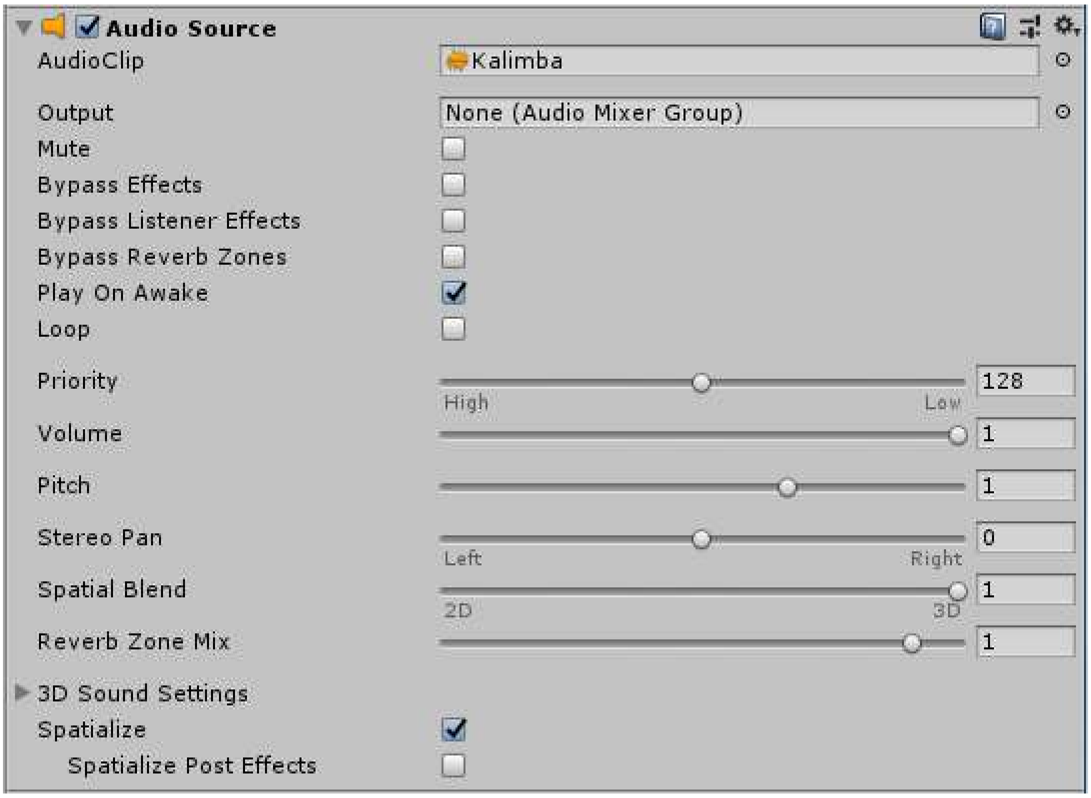
    a. Add the VIVE 3DSP Audio Source component attached to a

Unity audio source object. To add the Audio Source component,click **Add Component > VIVE > 3DSP_AudioSource.**

b. Add the VIVE 3DSP Audio Source component to an object that contains a Unity Audio Source component.

c. Add the Unity Audio Source component. Select **Spatialize** and set **Spatial Blend** to 1. This allows the source to have basic 3D spatialization features. Other advanced properties are provided in the VIVE 3DSP Audio Source component.

b. Add an **Audio Clip** to the Unity Audio Source.

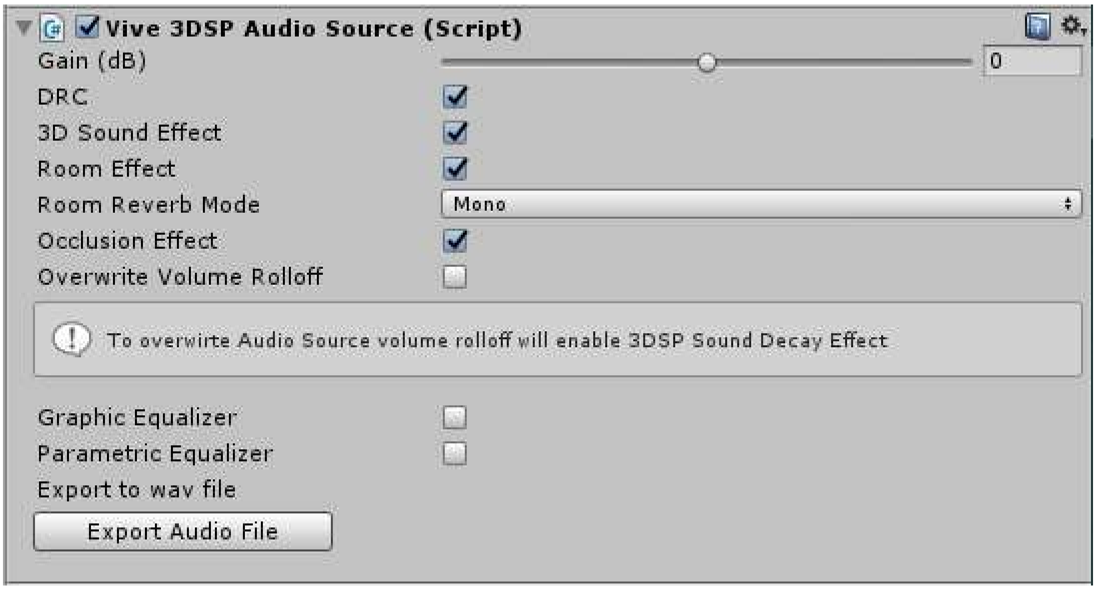


- **Priority**: The audio sources assigned with the lowest values (and audibility) are virtualized first. Priority is an integer between **0** [Highest priority] and **255** [Lowest priority].
- **Volume**: The volume level controls the level of sound coming from an AudioClip. The highest volume level is 1 and the lowest is 0 where no sound is heard.
- **Pitch**: Sets the frequency of the sound. Use this to speed up or slow down the sound. The value is in the range of **[-3.0, 3.0]**.
- **Spatial Blend**: Set how much this particular audio source is affected by 3D spatialization effect (e.g. attenuation, Doppler effect, etc). [**0.0** (Minimal 3D)

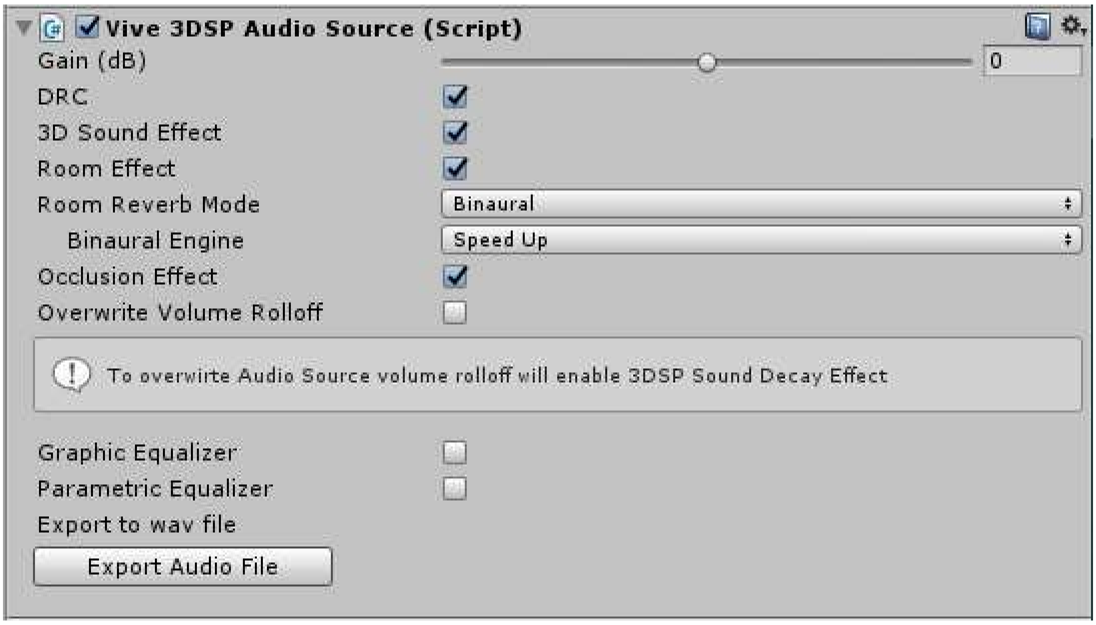
to **1.0** (Full 3D)].

- **Doppler Level**: Specify how much the pitch is changed based on the relative velocity between Audio Listener and Audio Source.

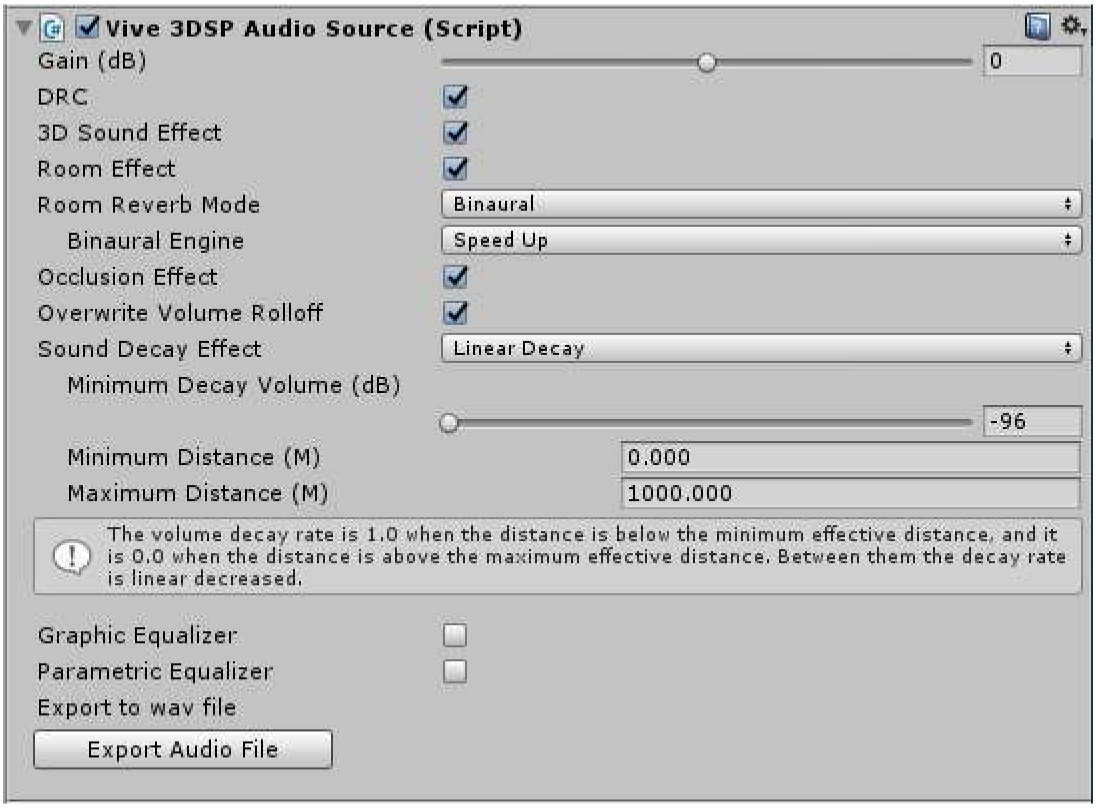
The VIVE 3DSP Audio Source component offers more options for the VIVE 3DSP spatializer:



- **Gain**: The volume increase or decrease of the audio clip, which is in the range of **[-24.0, 24.0] in dB**.
- **DRC**: Select to use the dynamic compression function to prevent sound clipping.
- **3D Sound Effect**: Turn on or off the spatialization effect. When it turns off, the audio clip will no longer has spatialization and sound decay properties.



- **Room Effect**: Turn on or off the room effect when both the current audio source and audio listener are in the same room. When it turns off, the audio clip will not be processed with room reverberation properties.
- **Room Reverb Mode**: The room reverb mode contains two options: Mono and Binaural. Mono means that the room effect engine returns mono reverberant output for both ears. Binaural means the room reverb effect engine returns reverberant spatial audio. When the Binaural mode is selected, the binaural engine option would appear and it has two options:
  a. **Normal**: It is the theoretically accurate method to calculate the binaural output. Theoretically accuracy requires high computing power but the output will be better.
  b. **Speed Up**: Compared to **Normal** mode, the numerical simplification method is used and the required computing power is much lower.
- **Occlusion Effect:** Turn on or off the occlusion effect. When it turns off, the audio source will not be blocked by pre-defined occlusion objects and gets it original sound.



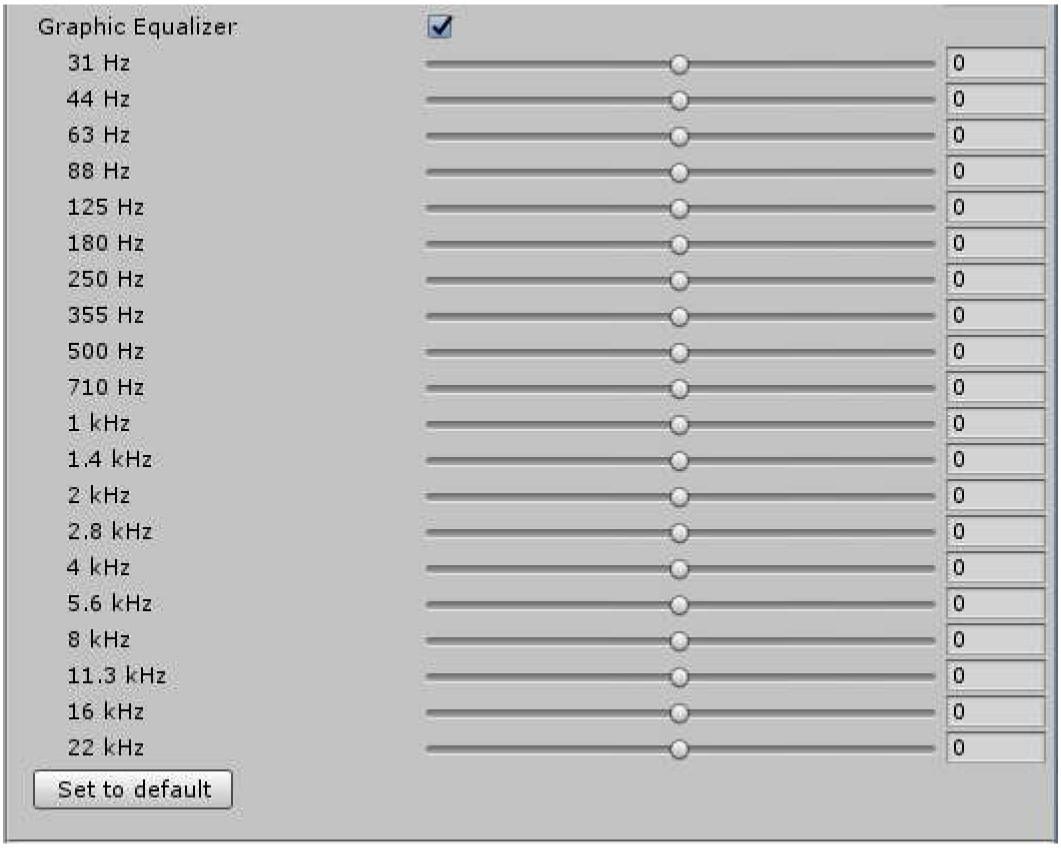
- **Overwrite Volume Rolloff**: Select to disable the Unity built-in volume rolloff effect and use the VIVE 3DSP Sound Decay Effect. Choose a **Sound Decay Effect**:
  a. **Real World Decay**: Sound decay effect based on real-world measurement database.
  b. **Point Source Decay**: Sound decay effect based on inverse-square law. It behaves as the point source where the sound energy decreased ratio is based on the propagation area (the square of the propagation distance).
  c. **Line Source Decay**: Sound decay effect based on inverse law. It behaves as the line source where the sound energy decreased ratio is based on the propagation distance.
  d. **Linear Decay**: Customized sound decay effect. The volume decay is linear to the distance between the source and listener.
     - **Minimum Decay Volume (dB)**: Keeps the minimum volume decrease caused by the sound decay effect.
     - **Minimum Distance (M)**: When the distance is lower than the Minimum Effective Distance, the volume decay rate is 1.0.
     - **Maximum Distance (M)**: The volume decay rate will be based on the **Minimum Decay Volume (dB)** if the distance is higher

than the Maximum Effective Distance.

e. **No Decay**: No sound decay effect applied.

Notes:

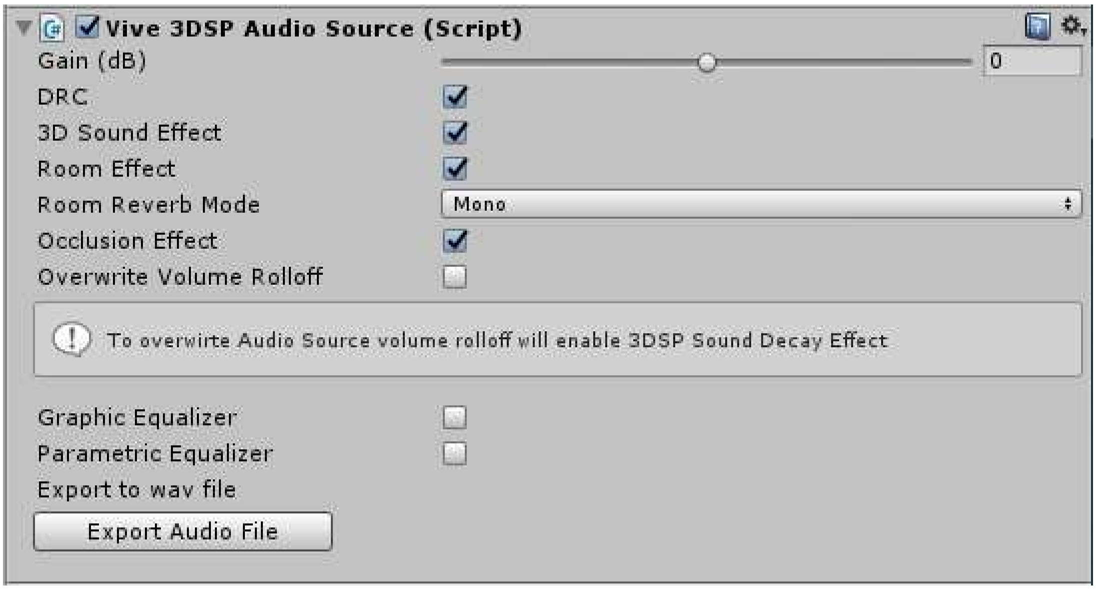
- When **Overwrite Volume Rolloff** is selected, the default Unity distance curve will be set at a constant volume regardless of the distance.
- When **Sound Decay Effect** is set to **Point Source Decay** or **Line Source Decay**, the **Minimum Decay Volume (dB)** value can also be set. It keeps the minimum volume decrease caused by the sound decay effect.
- To use the default Unity distance curve settings, clear the **Overwrite Volume Rolloff** option.



- **Graphic Equalizer**: The graphic equalizer is a convenient tool for tuning audio. In VIVE 3DSP Audio Source, the graphic equalizer has 20 bands and each band's level bar ranges from **-12dB** to **12dB**. Click **Set to default** to set all level bars to **0dB**.

- **Parametric Equalizer**: Set your desired boost/cut frequency positions and tune your desired gain and bandwidth. The first and last filters, **Low Shelving Filter** and **High Shelving Filter**, are two special filters that are effective at frequencies lower or higher than the user-defined frequency. The others are called **Peaking filters**, which boost/cut the frequency curve based on the user-defined frequency. There are three parameters that can be defined:
    a. **Frequency**: The center frequency for the **Peaking Filter** and the start frequency for the **Low/High Shelving Filter**. It ranges from 20 Hz to 20 kHz.
    b. **Gain**: The amount in dB which the frequency curve is raised or brought down. It ranges from -12 dB to 12 dB.
    c. **Q**: Controls the sharpness of the frequency curve of the **Peaking**

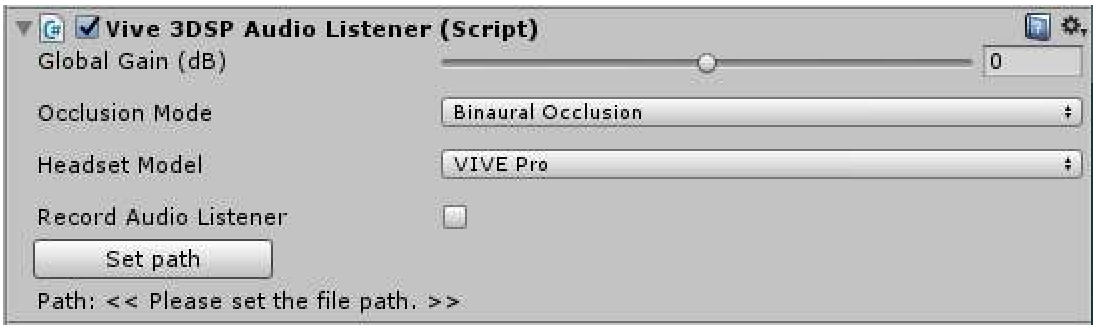
Filter. The larger the **Q** value, the sharper the frequency curve. It ranges from 0.2 to 5.



- **Export Audio File:** Click to export a specific audio source attached with the Vive 3DSP effects as an audio file. The Vive 3DSP effects include the Gain, Graphic Equalizer, and Parametric Equalizer.
  Note: Currently, the audio file can only be exported in the WAV audio format.

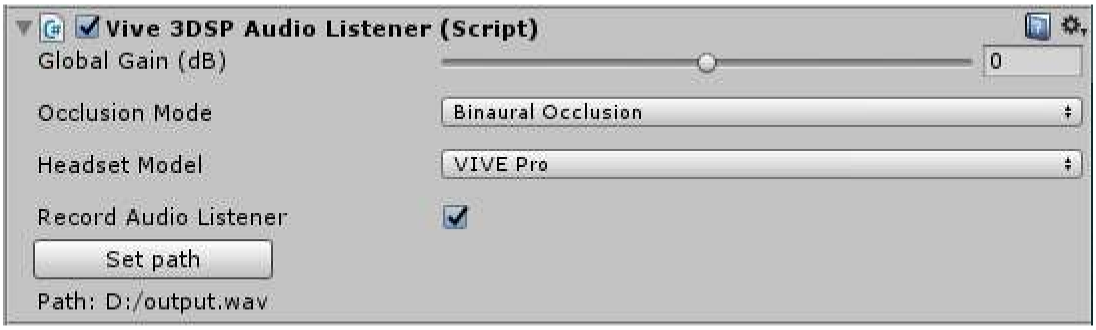
Audio Listener Settings

The audio listener is the content user—i.e., the main camera in Unity. Therefore, to have spatial audio effect, the VIVE 3DSP Audio Listener component need to be added to the main camera. To add the Audio Listener component, click **Add Component > VIVE > 3DSP_AudioListener.**



- **Global Gain**: Set the total volume changes of the audio from VIVE 3DSP Audio Listener. It is ranged from **-24 dB** to **24 dB**.
- **Occlusion Mode:** See **Occlusion Mode** in **Audio Occlusion Setting.**

- **Headset Model:** Select the target headset model. VIVE 3DSP will then tune the sound for that particular model to produce the best audio experience.
  a. **Generic**: No specific audio localization optimization done. It is suitable for general devices and headphones.
  b. **VIVE Pro**: Audio localization is tuned and optimized for the built-in headphones of the VIVE Pro.

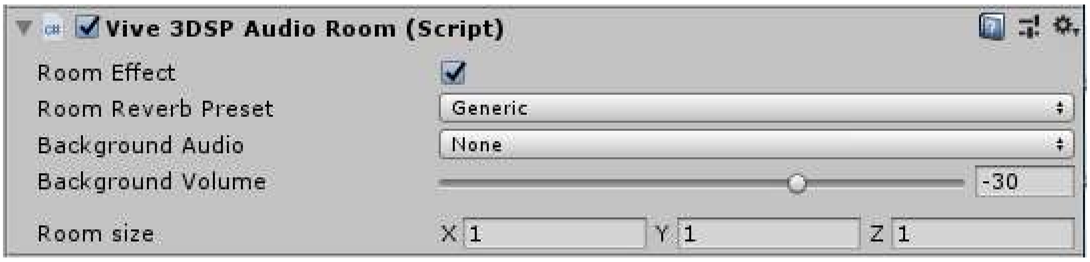


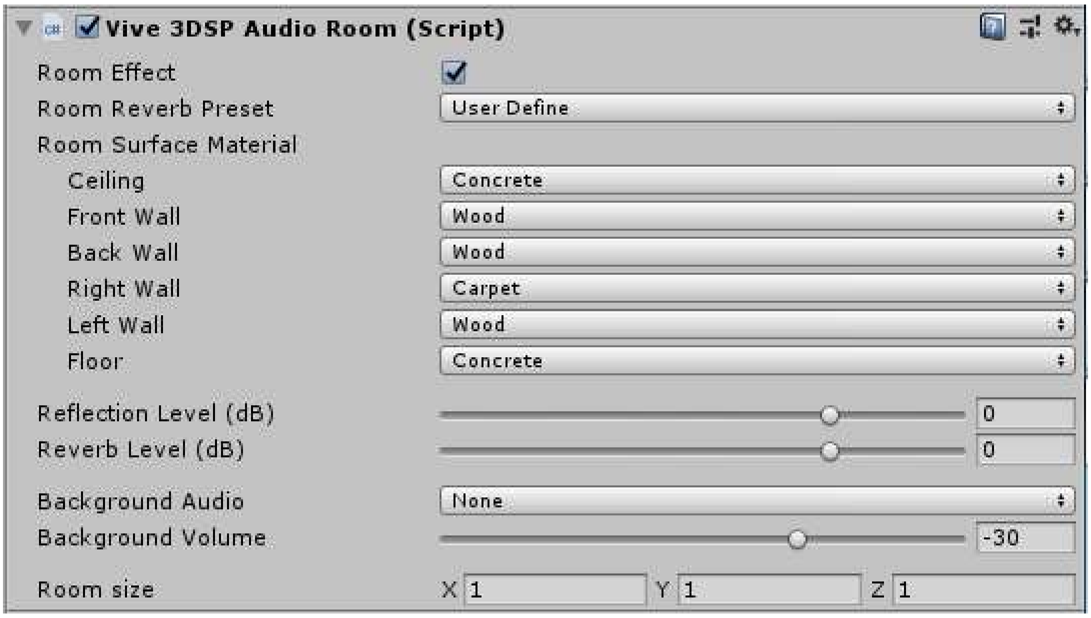
- **Record Audio Listener:** Select to record whatever the listener is hearing in the scene and export it as an audio file. Click **Set path** to set the location where to save the exported audio file. When you clear the **Record Audio Listener** option or stop Unity Play Mode, an audio file will be exported in the WAV audio format. Currently, the audio file can only be exported in the WAV audio format.
  Important: You must set the path and be in Unity Play Mode for **Record audio listener** to work.

Audio Room Settings

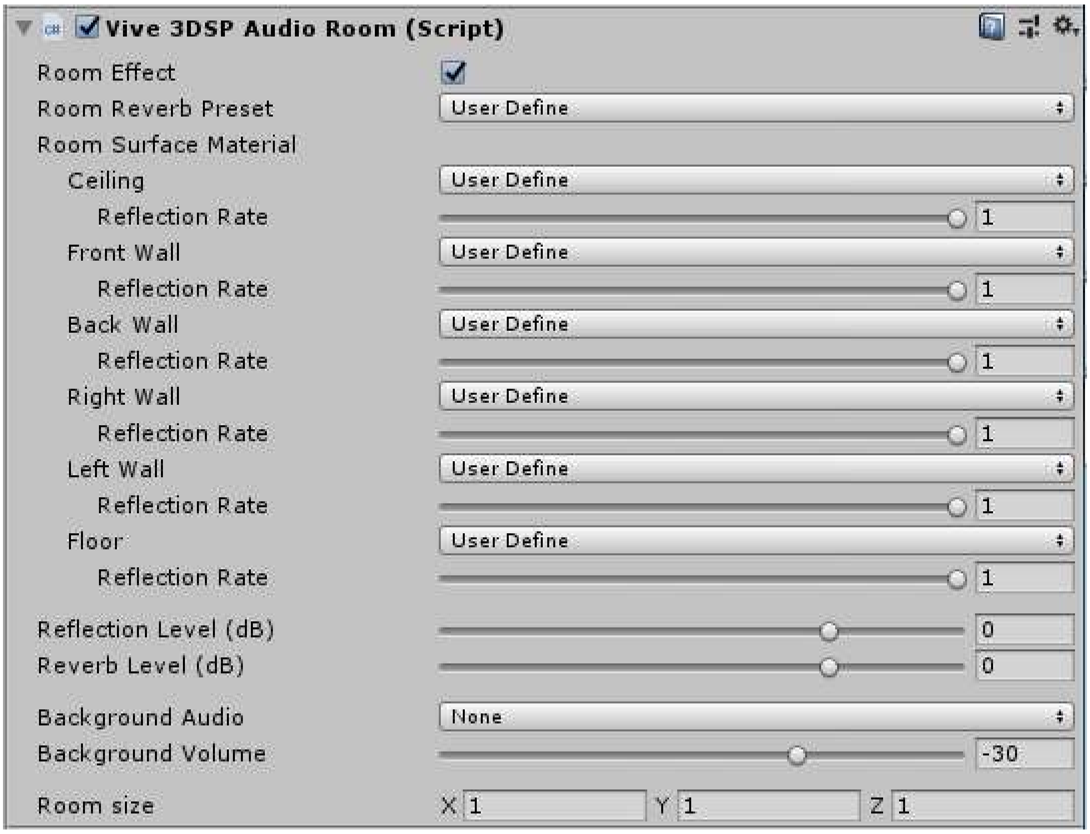
In the VIVE 3DSP Audio Room component, reverberations can be set to create specific room sound effects. All sounds inside a room are affected by the room size, wall materials, etc. How much influence these variables have on the sound depends on the reflection and reverb levels.

Several room types that are used most frequently is provided. If **User Define** is selected, a list of surface materials and sound properties that customize the surface material can be selected. To customize the surface material, click **Add Component > VIVE > 3DSP_AudioRoom.**



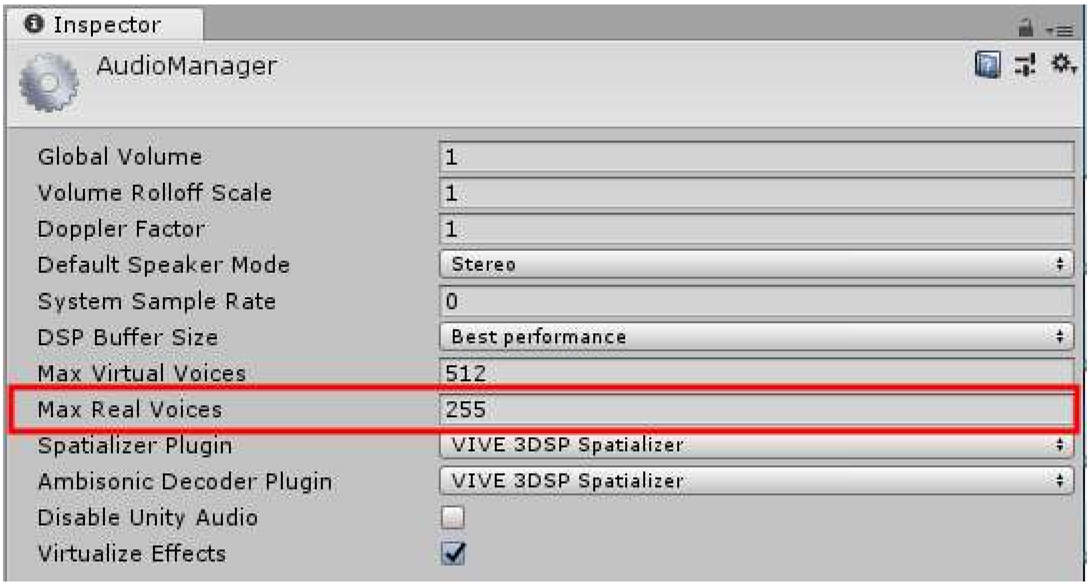


- **Room Effect**: Select to add the room reverberation effect into the object.
- **Room Surface Material**: The material of the six walls of the room would change ingredients of the reflected sound, which is the key component of room sound field. There are pre-defined material options for each wall, or to create a custom material, choose **User Define**, and set the **Reflection Rate** as the parameter of sound reflection ability of the wall.



- **Reflection Level**: It sets the volume level of the reflection, which is generated by our room effect. It ranges from -30 dB to 10 dB.
- **Reverb Level**: It sets the volume level of the reverberation for the room effect. It also ranges from -30 dB to 10 dB.
- **Background Audio**: The background audio is the ambient sound in the room. There are several default background audio types in the background audio option. Select **User Define** to use the preferred audio file for the background audio.

Note: The background audio is counted as one audio source in Unity. Some audio sources may not be played back because there is a limit in the Unity AudioManager settings. It can be set up to simultaneously play back up to 255 audio sources.



- **Background Volume**: Set the volume of the background audio.
- **Room Size**: Set the dimensions of the room in length, width, and height. The actual room size is calculated by multiply the room size with the corresponding scale in the Transform.

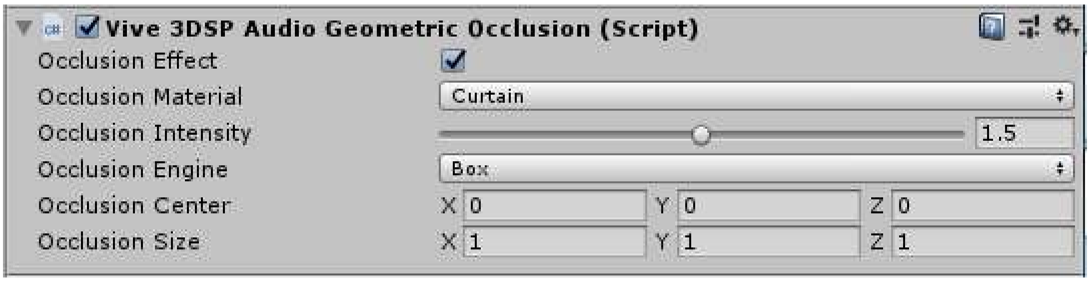
Audio Occlusion Settings

The occlusion setting should be set in **both** the audio source and audio listener. For audio source, the VIVE 3DSP Audio Occlusion component needs to be added.

There are two kinds of occlusion components in VIVE 3DSP: **Geometric Occlusion** and **Raycast Occlusion**.

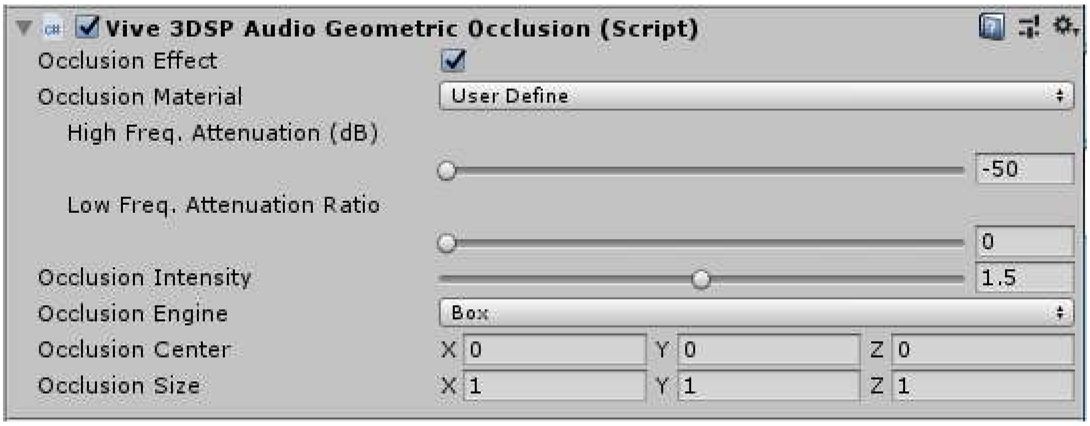
Geometric Occlusion

Click **Add Component > VIVE > 3DSP_AudioOcclusion > Geometric.**



- **Occlusion Effect**: Select to add the occlusion calculation into the object.

- **Occlusion Material**: The material of the occlusion object, which blocks the transmission path from audio source to listener, definitely cause the audio colored when the sound pass through it. There are some materials of the occlusion object in our Occlusion Material option. If all of them are not suitable, choose **User Define** to set your own effect. There are **High Freq. Attenuation (dB)** and **Low Freq. Attenuation Ratio** of the occlusion object can be set.

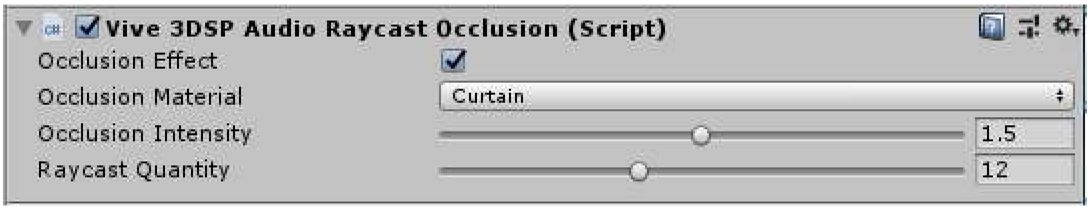


- **Occlusion Intensity**: The blocking effect of an occlusion object is also affected by the object density or the thickness, thus we define the occlusion intensity to represent those factors. The higher the value, the more occlusion effect is applied.
- **Occlusion Engine**: For the convenience, the geometry occlusion can deal with sphere and box geometric shapes. If it is set to sphere, the center and radius could be set. If it is set to box, the center and size could be set.
- **Occlusion Center**: Set the center coordinates of the object especially when importing objects from an existing file.
- **Occlusion size**: Set the blocking size in length, width, and height of the occlusion object.

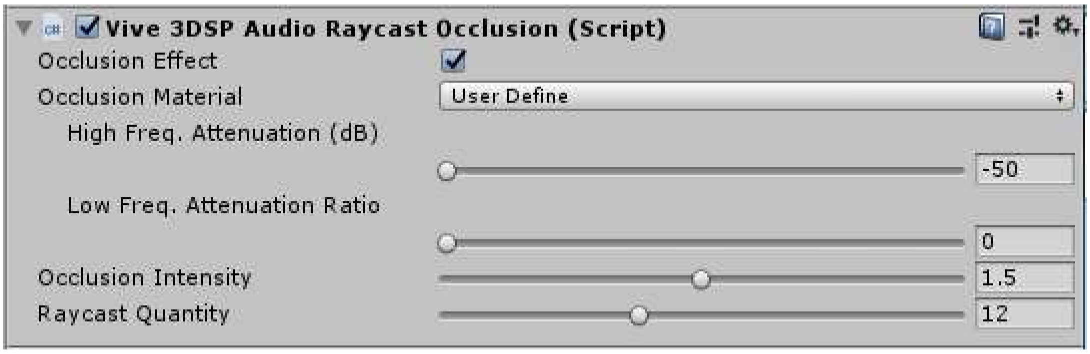
Note: There is no need to add the collider component to turn on the Geometric Occlusion effect. The Geometric Occlusion has its own built-in collider calculator.

Raycast Occlusion

Click **Add Component > VIVE > 3DSP_AudioOcclusion > Raycast.**



- **Occlusion Effect**: Select to enable the occlusion calculation into the object.
- **Occlusion Material**: Select the material of the object. Set to **User Define** so than **High Freq. Attenuation (dB)** and **Low Freq. Attenuation Ratio** of the occlusion object can be set. The default high frequency cut-off frequency is 5 kHz, and the **High Freq. Attenuation (dB)** sets the attenuation level of our filter. The **Low Freq. Attenuation Ratio** is defined as a parameter which influences both the cut-off and attenuation gain. It ranges from 0 to 1, which 0 means heavily attenuated and 1 means not attenuated

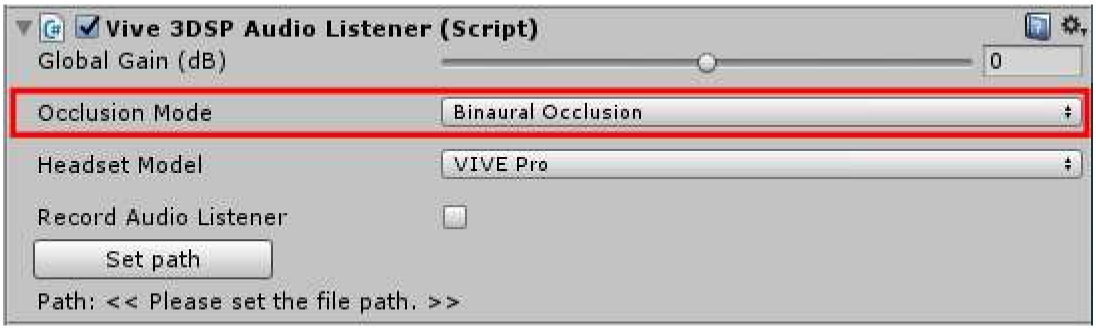


- **Occlusion Intensity**: The blocking effect of an occlusion object is also affected by the object density or the thickness, thus we define the occlusion intensity to represent those factors. The higher the value, the more occlusion effect is applied.
- **Raycast Quantity**: Set the raycast number to use in the occlusion effect. The higher the quantity, the better the occlusion effect quality. The quantity could be set from 1 to 30.

Important: Raycast Occlusion only works when the collider component is attached.

Audio Listener

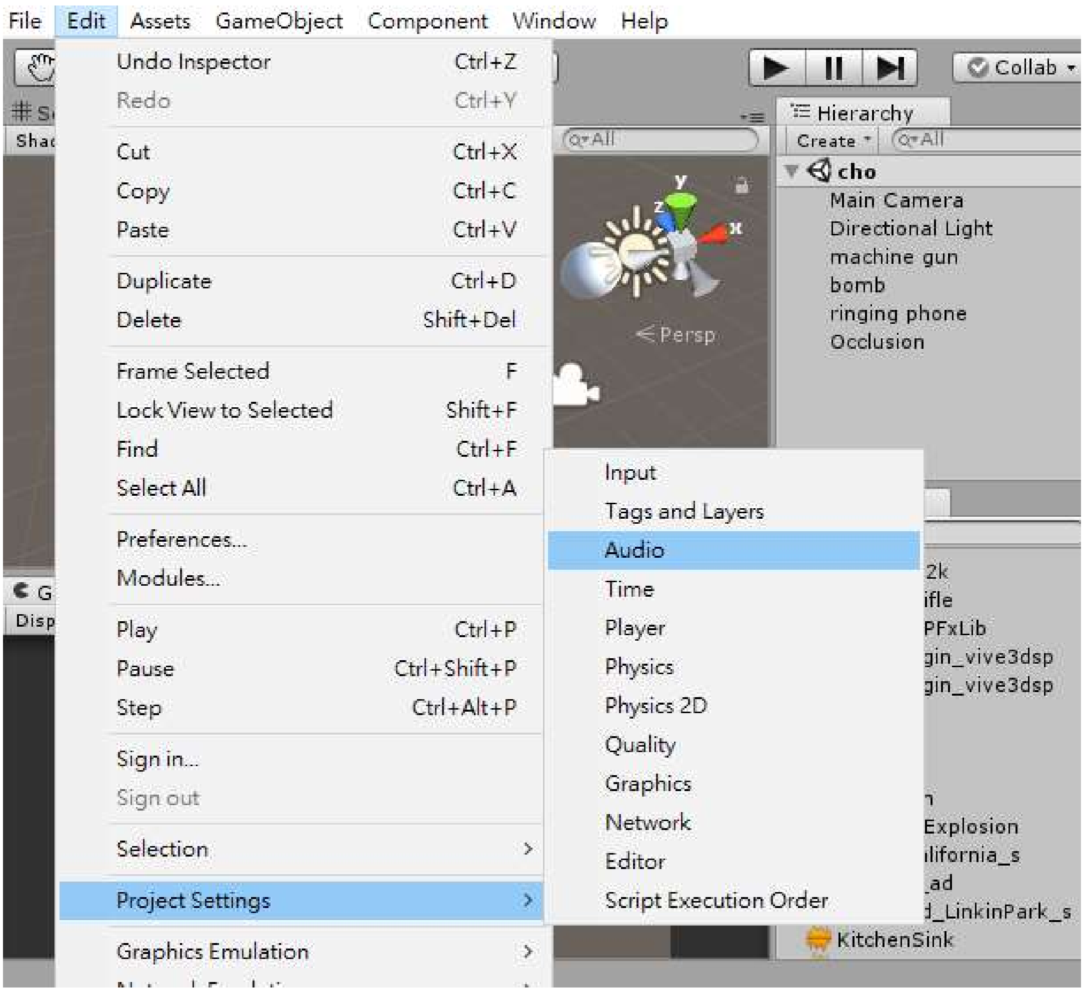
In Audio Listener, occlusion modes can be chosen.

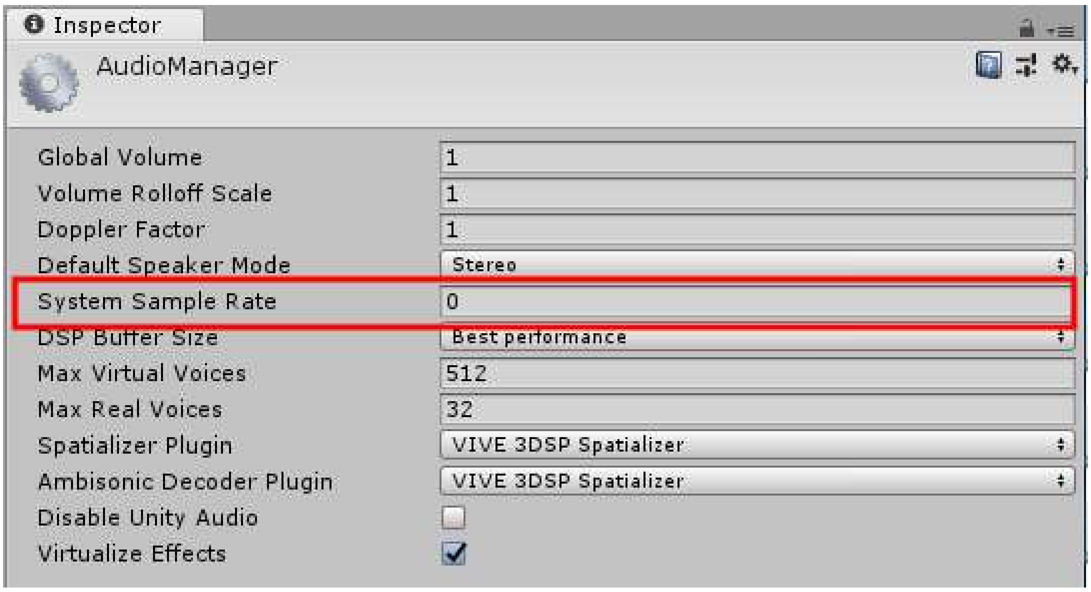


For **Occlusion Mode**, select **Mono Occlusion** to render the same occlusion effect in both ears; select **Binaural Occlusion** to render the occlusion effect separately in both ears.

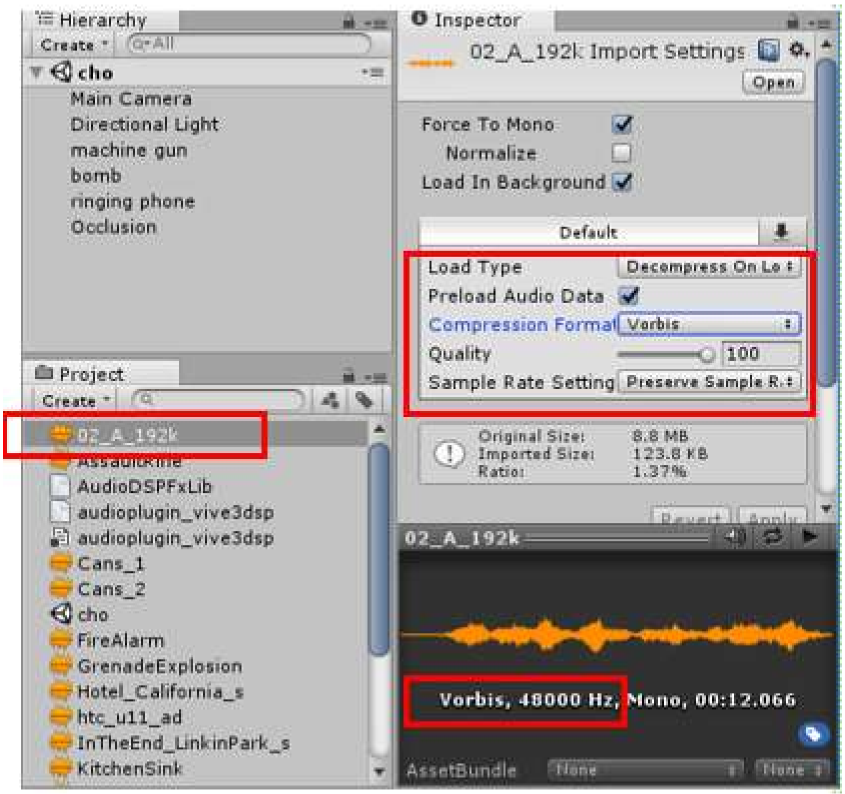
Hi-Res Audio Settings

1. For project settings, set **System Sample Rate** to **96000** (maximum sample rate of Unity).

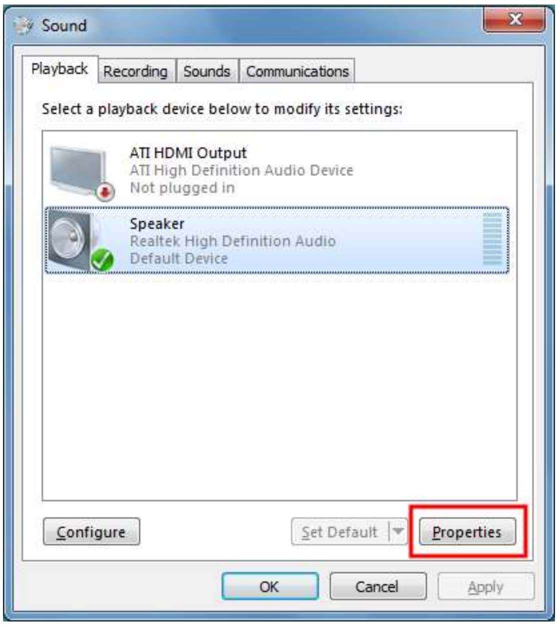


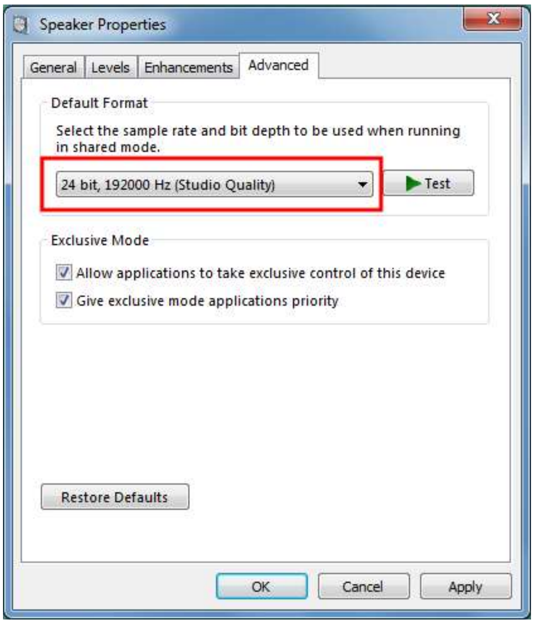


2. Modify the audio clip: Change **Compression Format** to **PCM**.



3. Modify the OS audio settings: Change sample rate to **96000 Hz** or **192000 Hz**





Troubleshooting

Camera is not responding to mouse movement

If the camera is not responding to mouse movements, clear the **Simulator** option in **Edit > Preferences > VIU Settings**.