AURO-3D OCTOPUS CODEC

Principles behind a revolutionary codec

Rev.: 2.6, Date: 2011 May 19
Author(s):
- Bert Van Daele, R&D Engineer Auro Technologies NV (bert@auro-technologies.com)
- Wilfried Van Baelen, CEO Auro Technologies NV (wilfried@auro-technologies.com)

1 INTRODUCTION

1.1 Background

In the past few decades, sound reproduction has known several revolutions: the introduction of stereo and later surround sound, the introduction of digital audio with the CD, mp3… With the recent advent of high quality (3D) video in Digital Cinema and on BluRay Discs, several trends started to emerge in the related audio fields as well. The main five topics that started to emerge are:

1. the wish to go back to uncompressed digital audio (Linear PCM).
2. the use of high resolution digital audio formats at higher sample rates (e.g. 96kHz) and greater bit depths (e.g. 24-bit).
3. the need for a new, more immersive playback format with true 3D spatial audio capabilities
4. the need for a new, backwards-compatible delivery format that allows the simultaneous distribution of multiple playback formats in one, single stream.
5. the need for a decoder with ultra-low latency to guarantee lip-sync and avoid delays between audio- and video-streams of a live broadcast program.

As a response to these trends/wishes, Wilfried Van Baelen developed the **Auro-3D Concept:** a new **Aurophonic** audio playback system combined with technical solutions necessary to bring this new listening and production format into the market while staying within the existing standards. Together with a team of highly specialized engineers, *Galaxy Studios* and *Auro Technologies* started in 2006 with the development of,the technologies as described in the **Auro-3D Concept** document. Part of this concept is the **Auro-3D Engine** built around three main technologies:

- **Auro-3D Octopus** codec
- **Auro-Matic** upmixing algorithm
- **Auro-3D Headphone**

The Auro-3D Concept gives the perfect solution for a real 3D listening experience with the highest audio quality and full technical compatibility.

This paper will focus on the principles behind the new, revolutionary **Auro-3D Octopus** codec, which enables the new **Aurophonic** playback format and gives a solution for the trend towards uncompressed, high-resolution digital audio while remaining within the limitations set by current standards like HDMI and DCI.

1.2 Main codec features

The patented **Auro-3D Octopus** codec has the following main features:

- encoding of multi-channel audio into a **PCM-stream** with less channels (e.g. 9.1 into 5.1 or even 2.0), which can be played back by existing systems without the need for a decoder. However, only when the decoder is used, the extra channels can be retrieved in their original audio quality.
- decoding of the **original** multi-channel audio with 100% channel separation
- **virtually lossless compression** at greater compression ratios than 'true' lossless codecs (> factor 2 with inaudible distortion levels < -85dB) without the use of perceptual coding techniques
- highest **compatibility** (backward and forward) with existing playback and distribution formats.

2 TECHNICAL PRIMER: DIGITAL AUDIO

In this chapter some background information is given to clarify some of the terms used further in the paper.

When recording sound using a microphone, it is converted into an electrical signal that is analogous to the air pressure changes we can hear as sound. This analog electrical signal is continuous, which means that it has an infinite number of possible values (of voltages) that can change within infinitesimal small differences in time.

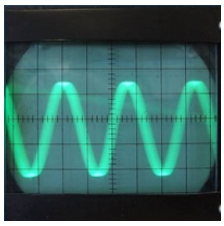


Figure 1 - Analog sinewave show on oscilloscope

In the digital world, however, sound needs to be represented by a finite number of values that come at regular time intervals. This process is called *sampling*. In an ADC or Analog-to-Digital Converter these samples are then furthermore quantized into digital numbers with a certain word length (e.g. 24 bit). A DAC or Digital-to-Analog Converter reverses this process by translating the digital numbers into analog voltages representing the sound wave.

The number of values that can be represented is not infinite anymore and is defined by the *bit depth* or *resolution* of the digital signal. Adding a low-level noise signal ('dither') to the audio signal linearizes the conversion process and thus gives infinite resolution like in the analog domain. The bit depth therefore determines the signal-to-noise ratio or dynamic range.

2.1 Sample rate

When converting sound from the analog to the digital domain, the incoming signal's value is evaluated, or sampled, at regular intervals. The time between two sample-points is determined by the *sampling frequency* or *sample rate*. A sampling frequency (f_s) of 48kHz thus means that the signal is sampled 48.000 times per second, or every 20.83μs.

According to the Nyquist theorem, the highest frequency that can be sampled is half the sampling frequency. So a sample rate of 48kHz gives a maximum audio frequency of 24kHz.

In the past decades, a trend emerged to go to higher sample rates. It is commonly known that human hearing is limited to approximately 20kHz for discrete sine waves. The sampling frequency for digital audio was initially defined with this limit in mind, hence the well-known sample rates of 44.1kHz (for audio-CD) and 48kHz (for DVD and video).

Research has shown, however, that the phase-accuracy of the human hearing is far more precise as it is a.o related to the shortest neural impulse in the brain: approx. 4-5μs. This is the period of a signal of about 200kHz, which cannot be heard as a signal (tone) on itself, but can be detected as phase differences. According to the Nyquist theorem, such high frequencies (or rather these small wavelengths and phase differences), need to be sampled at about 400kHz. Above this rate, humans cannot detect any changes anymore.

This, as well as other technical aspects (e.g. the phase accuracy of the DAC reconstruction filters), explains why higher sample rates (96, 192, 384kHz) still can be perceived to be sounding more open and natural.

2.2 Bit depth

At each sampling interval (each 20.83μs in case f_s=48kHz), the signal is represented as a number (quantization). The number of bits used determine how close the number represents to the actual signal level.

With 4 bits, as shown in the Figure 2, 16 different values are available to represent the signal. The quantization process determines which available digital value best approaches the actual analog value, resulting in the loss of a lot of intermediate values. This can be heard as digital noise or *quantization noise.*

The difference between the highest value that can be represented and this noise floor is called the *Dynamic Range (DR)*. This Dynamic Range is expressed in dBFS, decibels referenced to Full Scale, a logarithmic scale where each bit represents 6dB of DR.

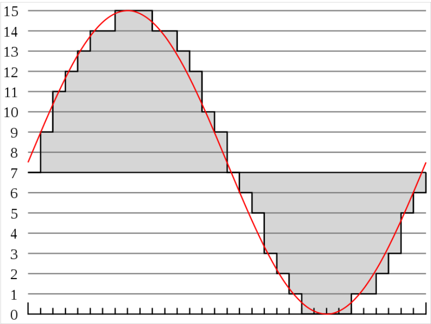


Figure 2 - Sampling and 4-bit quantization of an analog signal (red)

Bit depth	Number of steps	DR (dBFS,rms)
1	$2^1 = 1$	3
4	$2^4 = 16$	21
8	$2^8 = 256$	44
16	2^{16} = 65.536	93
18	2^{18} = 262.144	105
20	2^{20} = 1.048.576	117
24	2^{24} = 16.777.216	141

2.3 Common digital audio dynamic ranges

The following table gives a short overview of common sampling frequencies and bit depths for digital audio in linear PCM, and their respective, theoretical dynamic ranges.

Application	Sampling Rate (kHz)	Bit Depth	DR ($dBFS_{rms}$)
Audio-CD	44.1	16	93
DVD-Video	48	16	93
DVD-Audio	96-192	24	141
BD Video	48-96	24	141
Digital Cinema DCP	48-96	24	141
Pure Audio BD	96-384	24	141

All current *professional* workflows use 24-bit signals, often sampled at higher frequencies such as 96kHz or 192kHz. The main reason for using 24-bit audio is that this gives the engineers a lot more headroom to safely mix/add signals without running into problems such as clipping or excessive noise build-up due to rounding errors.

3 PRINCIPLES BEHIND THE AURO-3D OCTOPUS CODEC

3.1 24-bit Masters are overkill, but not useless

As shown in chapter 2.2, 24-bit digital audio files theoretically can provide a dynamic range of ca. 140dB[1]. Such a great dynamic range is very useful to provide the necessary headroom during production to minimize the build-up of noise during mixing as well as the introduction of distortion due to computational (rounding) errors. Sufficient proof has been given that working in 24 bits indeed gives far better audio quality than when working at lower bit depths like 16 bits as used in the audio CD format.

24-bit resolution is necessary within the production workflow, but totally overkill as a delivery format.

However, as a delivery format, 24-bit audio is actually overkill for several reasons, the main one being that it is virtually impossible to experience the full dynamic range of ca.140dB in real life. Here is a shortlist of reasons that will be detailed in the following paragraphs:

- A full dynamic range of 140dB above the noise floor will lead to severe, irreversible hearing damage.
- There are many laws limiting the maximum loudness to much lower levels (e.g. 100dB).
- There is hardly any reproduction technology capable of providing the full 140dB of dynamic range.
    - Digital-to-Analog Conversion
    - Amplification
    - Speakers

In the following paragraphs, it will be shown that a dynamic range of 100dB or even 90dB above the noise floor is a far more realistic maximum, which translates to a necessary bit depth of 18-bit or 16-bit respectively.

When using 24-bit audio as a delivery format, this thus leaves 4 to 6 bits to become available to enable the introduction of a new audio format while maintaining the highest compatibility with existing standards.

[1] *All dB-levels should be considered RMS (Root Mean Square).*
[2] *Average Dynamic Range above noise floor (25dB for quiet environments; 35dB for average environments; 50dB for noisy environments)*
[3] *Background Noise Level*

3.1.1 Hearing damage

In recent years, a dramatic increase in hearing problems has occurred, especially with young people, due to exposure to extremely high sound levels at concerts, as well as from e.g. their mp3-players. As a result, many initiatives (laws, standards, etc.) have been started all over the world to impose limitations to the maximum loudness at these venues. The average maximum level in most of these strict standards/laws is around 100-105dB(C) over a certain amount of time (several minutes). The main conclusion, however, is that this leads to maximum, short-term peak-levels of about 110-120dB.

The full dynamic range of 24-bit audio (140dB) above the average background noise can lead to permanent hearing damage

If the average noise floor in an extremely quiet listening room, such as the cinema theater, reaches the Noise Rating NR25 (Europe) or NC20 (US), a dynamic range of 95dB (equal to ca. 16bits) above the average noise floor would lead to sound pressure levels of about 120dBSPL. Higher levels create the risk for hearing damage, even at short time periods.

The following table gives an overview of some noises and their related average sound pressure level referenced to a good related Noise Rating.

			dB SPL	NC/NR	DR (dB SPL)[2]
Hearing Threshold			**0**		
Galaxy Studios Hall[3]			14	NC10/NR15	
Good Recording Studio[3]		Breathing	20	NC15/NR25	
Good Cinema Theater[3]		*Quiet Living Room*[3]	30	NC20/NR30	
Conversational Speech			**60**	**NR30**	**25**
Piano Practice		Electric Shaver	60	NR30	25
Business Office[3]		Piano Practice	65	NR30	30
Chamber Music		Barking Dog	75	NR30	40
Applause in Auditorium		Lawn Mower	85	NR`20	60
OSHA Required Hearing Protection in Factory			**85**		
Teleconference Room		Train at 30m	90	NR40	55
Subway	Farm Tractor		90	NR50	40
		Teenage Stereo	90	NR40	55
Sustained Exposure May Cause Hearing Loss			**90**		
Music Practice Room		Mp3-player at 5/10	94	NR30	70
French Horn	Average Factory Noise	Blender	100	NR30	75
	Diesel truck	Motorcycle	105	NR40	70
Orchestra			105	NR30	80
	Printing Press	Train	105	NR40	75
Snare Drum	Heavy Truck	Power Saw	110	NR50	75
Symphony Orchestra		Squeeky Toy to Ear	110	NR20	85
Pain Threshold			**120**		

[2] *Average Dynamic Range above noise floor (25dB for quiet environments; 35dB for average environments; 50dB for noisy environments)*
[3] *Background Noise Level*

			dB SPL	NC/NR	DR (dB SPL)[2]
Disco	Sandblasting		120	NR50	70
	Pneumatic Drill	Air Raid Siren	130	NR50	80
	Military Jet	Shotgun	140	NR50	90-110
Rock Concert	Aircraft Carrier Deck	Jet Takeoff	140	NR50	90
	Chest Wall Begins to Vibrate		**150**		
	Ear Drum Breaks Instantly		**160**		

3.1.2 Playback Calibration

In cinema theaters and recording studios for movies, the playback systems should be aligned according to the SMPTE-standard. This standard indicates that each channel should be set to reproduce a sound level of 83dB(C) SPL with pink noise at -20dBFS$_{rms}$ as input signal. This gives us two new limits to the usable dynamic range.

At the top of the dynamic range, there is the maximum level achieved with a digital full-scale signal. When the system is aligned according to the standard, the maximum level is 103dBSPL (i.e. 83dB + 20dB) per channel. Since this is the figure for a pink noise test signal, the actual maximum level for movie sound can be considered a bit higher: about 108dBSPL per channel, which theoretically should be represented with 18 bits, under the condition that there is no noise floor. Mind that if all channels are playing a full level simultaneously, the total sound pressure level can easily reach 120dB.

Again, this shows us that a dynamic range of just 108dB (18 bits) per channel is much greater than the actual usable range (of about 85dB above the very quiet background noise floor of 25dBSPL) in case the system is correctly aligned.

3.1.3 Limitations in Digital-to-Analog Conversion (DAC)

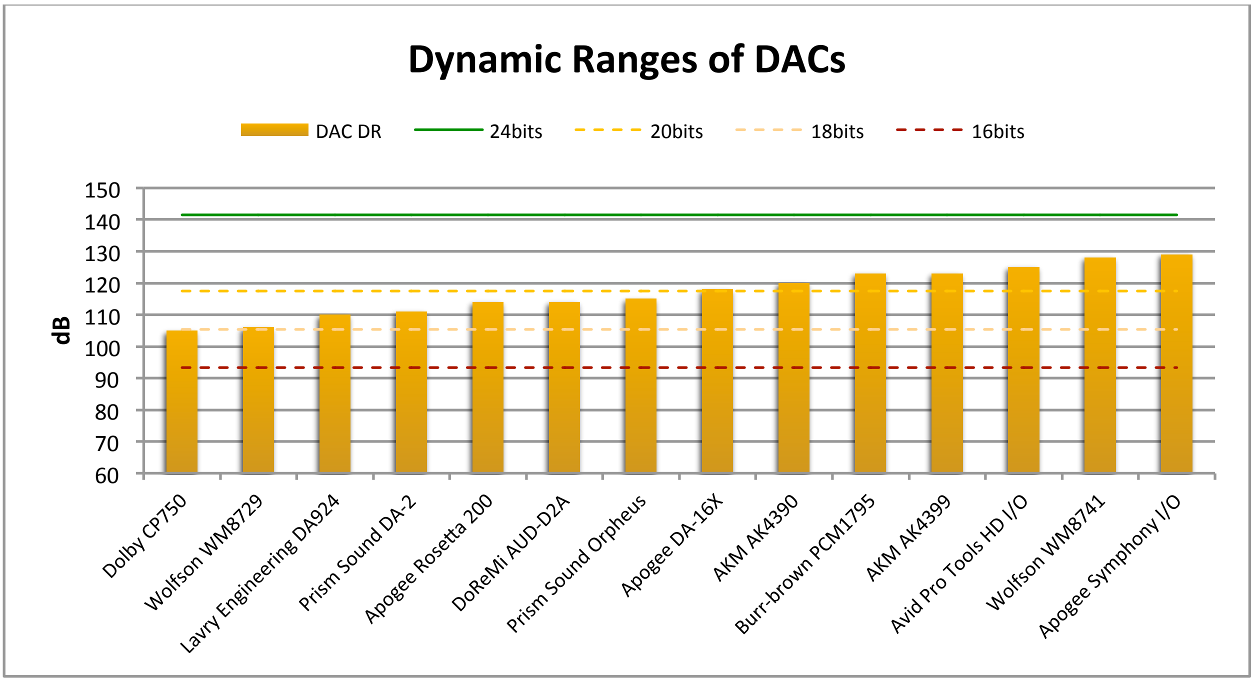
In any digital audio system, the audio needs to be converted from a digital to an analog signal at some point in the audio chain. Traditional audio systems (i.e. with analog amplification) use a DAC for this task, mostly at the end of the audio chain. While current DACs all accept 24bit audio signals, they are not capable of reproducing the full 140dB range. When looking to the specifications of even the best and most expensive commercially available DACs, the dynamic ranges hardly exceed numbers of 117dB, which is the equivalent of 20 bits.

Even the best DA-converters' dynamic range is limited to the equivalent of ca. 20 bits.

The table below shows the specifications of some converters, which are considered amongst the best available.

	Brand/Type	DAC DR (dB_{rms})
Professional Audio	**Avid Pro Tools HD I/O**	125
	Apogee DA-16X	118
	Apogee Symphony I/O	129
	Apogee Rosetta 200	114
	Prism Sound DA-2	111
	Prism Sound Orpheus	115
	Lavry Engineering DA924	110
Digital Cinema	**Dolby CP750**	105
	DoReMi AUD-D2A	114
Components (IC)	**Burr-brown PCM1795**	123
	Wolfson WM8741	128
	Wolfson WM8729	106
	AKM AK4399	123
	AKM AK4390	120

The last five examples are actually the converter-ICs, which then have to be implemented in a playback device. This means that the indicated figures are actually the best-case results measured in ideal lab-conditions. Depending on the actual implementation in the product, this figure is often not even reached, due to problems in the PCB layout, power supply, etc., except in the most expensive devices.



This thus leads to the conclusion that there currently are no converters available capable of truly delivering the full dynamic range available in a 24-bit signal, or not even in a 22-bit signal for that matter.

Only a few, very expensive, DA-converters, mostly from the professional world, are capable of reproducing a dynamic range of higher than the equivalent of 20 bit. The most common devices used in cinema theaters, have a dynamic range of max. 19 bit, e.g. DoReMi (19 bit) and Dolby (18 bit).

3.1.4 Limitations in playback systems

Next to the limitations observed in DA-converters, there are even bigger limitations in the dynamic range that should be reproduced by the average playback system in cinema theaters or by home theater systems.

As a result of the calibration, the playback system should be able to reproduce sound levels of up to $120dB_{SPL}$. Taking into account that even an exceptional speaker has an efficiency of approximately 90dB/W/m, this leads to the requirement that the amplification system should be capable of providing more than 2000W per channel. And this is also only possible in case the speakers don't break.

The next limitation is the fact that such a system should not have any audible system noise, such as hiss or hum, when there is no signal playing. In other words: the system noise should be lower than the background noise at all times. While amplifiers and speakers can be found with sufficient power handling capacities to produce the 120dB sound pressure levels, it is certainly a lot harder to find systems that do this while also lacking any audible self-noise.

It can therefore be concluded that most playback systems actually do not need to reproduce a dynamic range that is higher than 105dB, which is the equivalent of maximum 18-bits.

It should be noted that in mobile applications and even low-to-mid-end consumer electronics, these dynamic ranges are even more compromised due to component price constraints, as well as technical limitations such as low battery voltages, low power consumption, etc. In such systems, the dynamic range is limited to below 100dB, or the equivalent of 16 bits. This means that the dynamic range of more than 18 bits simply cannot be reproduced.

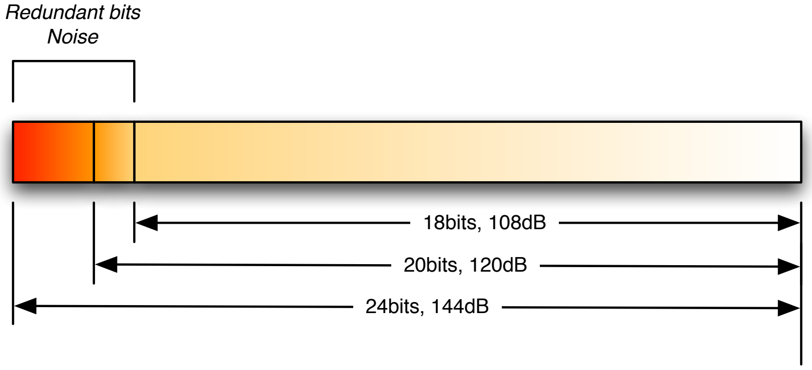
3.1.5 Conclusion

In the chapters above it has been shown that the useable dynamic range in an audio system is defined by:

- hearing damage considerations (maximum level above background noise floor)
- DA-converter limitations
- playback system limitations.

Taking all these elements into consideration, brings us to the conclusion that, although a word length of 24 bit is very useful in the professional production workflow, a delivery format using 24 bit and thus providing 140dB dynamic range is not needed. In fact, 20 bit or even 18 bit, providing a dynamic range of 117dB and 105dB respectively, is more than sufficient.

Currently accepted high-resolution delivery formats are all based on 24 bit, mainly to maintain compatibility with the professional production workflows and avoid extra conversion steps to lower resolutions. However, as explained above, such a delivery format has about 4 to 6 redundant bits that carry little information, mostly noise.



The **Auro-3D Octopus** codec can therefore use the redundant bits to carry extra information to encode extra audio channels, as well as non-audio metadata.

The Auro-3D Octopus codec makes the redundant bits of a 24-bit delivery format (master) useful again.

To keep the original audio quality at its maximum level, the codec will analyze the audio signal every 1ms and assign fewer bits to the non-audio data in case the original audio is low in level. In other words, if the audio level in a 1ms-block is low, only two bits will be used for the encoding-data in this block, while if the level is high enough in other blocks, more bits (up to 4) are used. This guarantees that the maximum possible audio quality is maintained at all times, even though it is widely accepted that the difference between 20-bit and 24-bit audio is inaudible.

3.2 Mixing and un-mixing of audio-channels

3.2.1 Unique capabilities

The patented **Auro-3D Octopus** codec is unique in its capability to **mix** different audio-channels into one channel and '**unmix**' the original separate channels in the decoder at their original audio quality and with 100% channel separation. What this means is that the codec can take different (up to 3) playback formats and mix/encode them into a PCM-stream with a lower number of channels and un-mix/decode the correct playback format for the used playback system, without the use of dematrixing techniques as used by most other codecs.

Even more important is the fact that this mix can be **artistically controlled** by the producer/mastering engineer, giving the highest artistic freedom to produce different mixes for e.g. the stereo-, surround sound- and Aurophonic versions of the same material while using only one carrier format in one PCM-stream, which can be listened to as a (controlled) mix of the different source formats.

With the Auro-3D Octopus codec multiple playback formats can be artistically mixed into one PCM-stream with less channels and unmixed again to the original formats.

3.2.2 Some applications

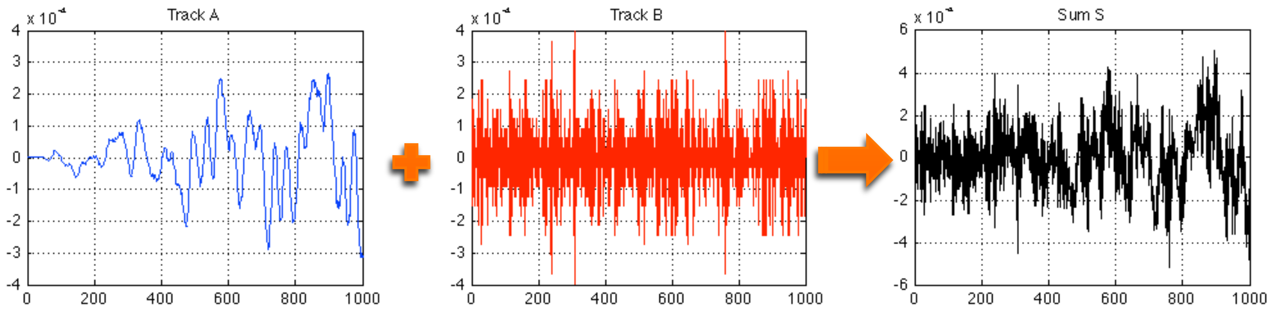
- an 9.1 Aurophonic and a 5.1 surround sound mix merged into one 5.1 Auro-encoded PCM-stream (for BD).
- an 11.1 Aurophonic and a 5.1 surround sound mix merged into one 5.1 Auro-encoded PCM-stream (for Digital Cinema DCP).
- an 13.1 Aurophonic and a 7.1 surround sound mix merged into one 5.1 Auro-encoded PCM-stream (for Digital Cinema DCP).
- a 5.1 surround sound mix and 2.0 stereo mix merged into one 2-ch PCM-stream (for broadcast)
- a 7.1 and 5.1 surround sound mix merged into one Auro-encoded 5.1 PCM-stream (for BD and Digital Cinema DCP).
- an 9.1 Aurophonic mix and a 5.1 surround sound mix both mixed into one 2-channel (stereo) PCM-stream (for broadcast).
- …

3.2.3 Mathematical magic?

The **Auro-3D Octopus**' mixing and un-mixing capabilities are at the heart of the codec. The following illustration shows why this is so special and unique.

A digital audio-signal is a long series of numbers (=sample values), where each number represents the value of the audio signal at regular intervals. Mixing two digital audio-signals is in fact the simple addition of these numbers of each original signal, creating a new signal with the summed values in the process. So, if signals A and B are added, the resulting sum S is made as follows.

$$S_n = A_n + B_n$$

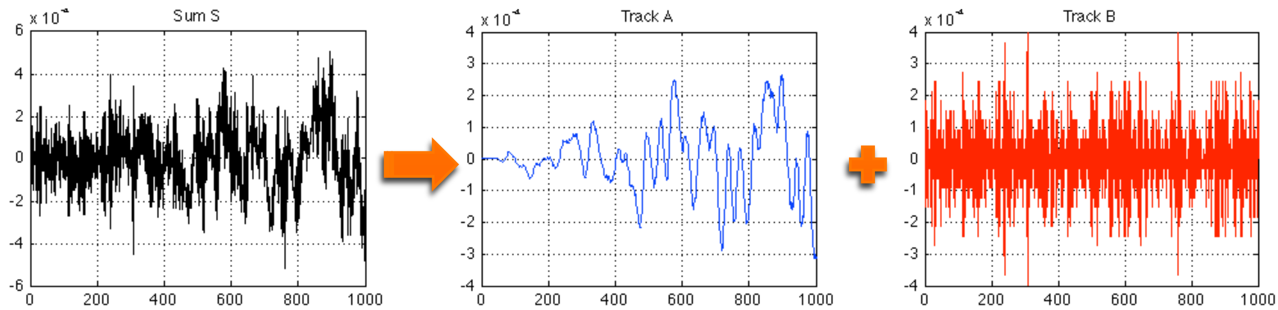


For the following example, we will use integer sample values (=whole numbers) within the 24-bit range between -8388607 and +8388608.

So, imagine the following sum was given at a certain moment (example with 24bit signal represented as integer values):

$$A_1 = 1287456$$
$$B_1 = 490566$$
$$S_1 = A_1 + B_1 = 1287456 + 490566 = 1778022$$

Normally, it is impossible to know that S_1 is actually the result of a sum, let alone what numbers this sum was made of. When the sum S_n is made by the **Auro-3D Octopus** codec, however, some extra information is stored within the lower bits of the PCM-stream that enables the decoder to reconstruct the values A_n and B_n. These values approach the original values A_n and B_n (where the sum S_n originated from) with a precision such that the error E_n is lower than the errors introduced by most DA-converters, making the decoded version perceptually indistinguishable from the original (see chapter 3.1).



Before the mixing/addition, it is even possible to apply a gain to one or both signals to create an artistic downmix. Such a gain is actually a multiplication of the sample values. So the original summing-formula becomes:

$$S_n = \alpha \cdot A_n + \beta \cdot B_n$$

With the encoder-information stored within the PCM-stream, the decoder tis then capable of restoring A'_n and B'_n with the same precision as described above, 'un-doing' the downmix-gains in the process.

But that's not where the codec's capabilities end. The **Auro-3D Octopus** codec can even perform this same process with **three** original signals and with the same audio-quality (albeit with a slightly higher noise floor)!

3.3 Staying in the PCM-domain

3.3.1 Advantages

An important feature of the **Auro-3D Octopus** codec is that the encoded file can still be listened to, just as regular audio, without any audible compromise. This is possible since the remaining bits in each sample after encoding, still represent regular audio data as linear PCM, without any compromise to its original audio quality. So, anyone that does not have a decoder can still listen to the audio, as a controlled mix of the encoded channels at the same quality as uncompressed PCM-audio. If a decoder is available, it will automatically detect the presence of the **Auro-3D** encoded material, and output the original signals.

The Auro-3D Octopus codec completely stays within the PCM-domain, guaranteeing the highest compatibility and audio quality.

The codec accomplishes this by using the 2-4 lowermost bits of each sample to encode the additional channels, depending on what the encoder really needs. This number is variable within the audio stream, as the encoder evaluates the original dynamic range of each block of 512 samples (i.e 1ms at 48kHz) and adapts its data set accordingly, making sure that the decoded waveforms give exactly the same audio quality as the source.

Due to the fact that the **Auro-3D Octopus** codec stays in the uncompressed PCM-domain, it is far more relaxing to the brain and body than other codecs based on perceptual coding while giving similar bandwidth reductions.

3.3.2 Other considerations

In case the encoded PCM-stream is further compressed using a lossless codec (e.g. DTS Master Audio HD), the decoded audio can be still be further decoded by the **Auro-3D Octopus** codec.

Just like all digital codec schemes, the **Auro-3D Octopus** codec is sensitive for volume changes made to the encoded PCM-stream. But while in all other systems this renders the audio-data completely useless, an Auro-encoded PCM stream can still be listened to as high-quality audio in the carriers format. E.g., if volume changes are made to a 2-channel Auro-encoded PCM stream that contains a surround-mix, the 5.1-format will be lost, but not the PCM stereo-version.

Using a special tool, however, it is even possible to safely perform volume changes to the complete Auro-encoded PCM-stream, without the need for any intermediate decoding or re-encoding, keeping the audio at its highest quality and keeping all included formats intact.

3.4 No perceptual coding or masking techniques

In contrast to most common, *lossy* codecs (mp3, AAC in iTunes, Dolby AC-3…), the **Auro-3D Octopus** codec is not based on any psycho-acoustical principals such as masking or other perceptual coding techniques, but rather uses advanced mathematical principles working on the digital data, much like *lossless* coding algorithms (FLAC, Apple Lossless, Dolby TrueHD…).

The **Auro-3D Octopus** codec, however, does not produce bit-identical copies of the original samples after decoding, but the differences between the original waveforms and their decoded counterparts (or in other words, the artifacts or side-effects) are extremely small, so that they are next to inaudible for the human ear (see chapter 3.5.1). This thus brings the audio quality of the **Auro-3D Octopus** codec to the same level as true lossless codecs, exceeding the quality of lossy codecs based on perceptual coding. For this reason, the **Auro-3D Octopus** codec can be considered a **virtually lossless audio codec**.

However, the main strengths of the **Auro-3D Octopus** codec over truly lossless codecs are:

- compatibility
    - the audio remains in the PCM-domain and can therefore be listened to without any decoder
    - multiple playback formats (stereo, Surround, Aurophonic sound) can be merged (artistically mixed) into **one** single PCM-stream, the selection is done at the decoder

    - encoded streams remain within the constraints of existing standards (HDMI, BD: max. 8 discrete channels, while 10 are needed for 9.1 Aurophonic audio)
- higher compression ratios
    - from 50% up to 88% (factor 2 up to 9); Note: true lossless codecs only reach 40% reduction (a factor 1.6) on average for movies and only 20% reduction for pop-music.

3.5 Virtually lossless

The **Auro-3D Octopus** codec is considered to be virtually lossless, as the used compression techniques only introduce noise-like artifacts that are extremely low in level and thus easily masked by the decoded audio signal..

3.5.1 Inaudible side-effects

Every existing audio device, analog and digital, has unwanted effects (distortion) on the signal to some degree. Many studies have been done in the past to quantify the audibility of different kinds of distortion products. As a general rule, it can be said that distortion becomes inaudible when its level remains at least 60dB below the original audio (cfr. Report…).

Compared to many other codecs, like mp3, AAC, AC3…, the **Auro-3D Octopus** codec excels in this respect as the artifacts, i.e. the differences between the original signal and the version after decoding, remain far below this threshold. In most cases, the artifacts are below -85dBFS[4] compared to the original signal, often even below -90dBFS. Although this is indeed not bit-by-bit identical, it is clear that to the human ear the decoded audio signal sounds virtually identical to the original audio.

3.5.2 Artifacts sound character

While the level of the artifacts is already extremely low, another reason why the original audio and its decoded counterpart are virtually indistinguishable from each other is the fact that the spectrum of the artifacts is actually similar to white noise. Other, lossy codecs produce a distortion spectrum that is highly correlated to the source, while also containing artifacts such as pre-echoes from the various filter-banks causing time-smearing. As such they have a very intrinsic sound that can cause fatigue to the listener.

The artifacts produced by the **Auro-3D Octopus** codec, however, are more *noise-like* as well as less *signal-dependent* and thus act in the same way as a dithering algorithm where added noise (i.e. distortion) actually improves the sound.

[4] Distortion figures in this revision of the document are provisional. A table with more detailed measurements for different codec modes and compression ratios will be published in later revisions of this document.

4 OTHER FEATURES OF THE CODEC

4.1 Fastest audio decoder ever

An important feature of the **Auro-3D Octopus** codec is its extremely low decoding latency of only ca. 3 samples (ca. 60μs @ 48kHz). This not only ensures perfect lip-sync capabilities, but also easy video-editing with Auro-encoded PCM-streams since the audio remains within the same video-frame before and after decoding, removing the need for offset video editing tools.

4.2 Multiple generations of en-/decoding

The **Auro-3D Octopus** codec allows for multiple sequential encoding and decoding passes without any changes to the audio data. This means that the audio quality will not be influenced by multiple generations of en-/decoding.

4.3 Metadata

The information stored in the lower audio bits is not only used for the encoding of the extra audio information. In fact, a special data channel is foreseen to carry extra, non-audio information known as metadata. This metadata information has several purposes:

- information of the embedded audio streams (channel count and identification)
- information about the reproduction of the audio, such as Dynamic Range Reduction, etc…
- information about the artistic downmixes
- parameters for the upmixing algorithm at the decoder side (for more information, see Auro-Matic White Paper)
- information about the Prgram Loudness and Loudness Range as defined by ITU-R BS.1770 and EBU E 128. (P-Loud)
- information about the use of the X-curve
- …

5 APPLICATIONS

While the **Auro-3D Octopus** codec was developed with Aurophonic multi-channel audio in mind, the codec has the potential for many other uses as well. Here is an overview of its possible applications.

5.1 Aurophonic multi-channel audio (3D → 2D)

Aurophonic multi-channel audio (sound in 3D) can have any number from 10 (9.1) to 24 (22.2) discrete audio channels. However, current standards for the distribution and transportation of multi-channel audio are limited to 8 (HDMI/BD) discrete channels.

In its *Auro-3D* mode, the **Auro-3D Octopus** codec can encode the extra height channels of any *Aurophonic* playback format, into a PCM stream that fits within the constraints of the currently accepted standards (e.g. 11.1 Aurophony into a 6-channel (5.1) PCM-stream, sent over HDMI).

5.2 Surround Sound (2D → 1D)

Just like *Aurophonic* multi-channel audio needs to remain within the boundaries set by existing standards, standard surround sound (5.1, 7.1) faces similar boundaries. In many cases the surround sound needs to be transported over a 2-channel interface, sometimes multiplexed with a separate stereo mix.

Also for such cases, the **Auro-3D Octopus** codec provides a perfect solution: 5.1 Surround can be dynamically and artistically mixed into a stereo PCM-stream. The decoder can then recreate the original 5.1 mix in its original audio quality. This is known as the *Auro-2D* mode of the codec.

5.3 StereoMagic (3D → 2D → 1D)

This special mode even makes it possible to encode an Aurophonic 9.1 mix artistically controlled (dynamically mixed) into a stereo PCM-stream, while at the same time including the artistically controlled 5.1 surround-mix, which was created from the Aurophonic mix. This means that the 2-channel StereoMagic PCM-track becomes the carrier for 3 different artistic mixes, or the equivalent of 18 tracks in total:

- artistic stereo mix
- artistic 5.1 surround mix
- artistic 9.1 Aurophonic mix.

5.4 Reduction of number of audio streams

At the same time, the need exists to present the same title in different formats simultaneously (e.g. Aurophonic 11.1, 5.1 surround sound and stereo). With uncompressed audio and currently available compression technologies, this rapidly adds up to very high track counts (20 or more).

With the **Auro-3D Octopus** codec, however, this is not the case. Up to three different playback formats (e.g. 2.0, 5.1 and 9.1), each representing a different artistic mix, can be encoded into one single, 6-channel PCM-stream or one single stereo PCM-stream (see StereoMagic). This thus eliminates the need for storing and transporting multiple audio streams of the same program on the same medium (e.g. Blu-Ray Disc) while remaining within the constraints of currently accepted standards (max. 8 channels of audio per stream).

5.5 Multi-channel audio in less channels

Using the **Auro-3D Octopus** codec, a high number of audio tracks can be reduced to a lower number to reduce bandwidth. E.g., a 24-channel audio multi-track can be encoded into an 8-channel PCM-stream. Similarly, 18 tracks of PCM audio can be encoded into a 5.1 PCM-stream.

5.6 Bandwidth reduction/compression

The **Auro-3D Octopus** codec is not only capable of encoding different channels into a lower number of channels. The patented techniques in the codec also allow for bandwidth reduction in stereo or mono files. Bandwidth reduction factors of 2 to 9 can be achieved, while still maintaining as close possible the original high-quality audio after decoding. One of the extra advantages of this codec is that the compressed PCM-files can still be played by any media player, e.g. for fast browsing through a large collection of files, as they are still in the standard PCM-format, but at a lower sample rate.

This mode of the codec is known as the *Auro-Cx* mode.

6 ANNEX 1 – AURO-3D SYSTEM

6.1 System Overview

The **Auro-3D Octopus** codec is only one part of the **Auro-3D System**.

This system consists of the **Aurophonic** sound playback system with height speakers and the **Auro-3D Engine**.

The **Auro-3D Engine** is the combination of the different available technologies that enable the complete aurophonic listening experience. These are:

- ISA (Input Source Analyser), with optional Encryption/Decryption
- Octopus codec
- Auromatic upmixing algorithm
- Aurophonic headphone algorithm (*under development*)

The **ISA**-block in the **Engine** analyses the incoming PCM-signal to see whether the stream contains the extra data needed for the **Auro-3D Octopus** codec. In case this data is not detected, the **Auromatic** upmixing algorithm can be used to create an Auro-3D listening experience from any other input signal.

More details of the complete **Auro-3D Engine** and other blocks will be described in a separate document.

6.2 Auro-3D Octopus Codec Modes

As described in the previous chapters, the **Auro-3D Octopus** codec has four different operational modes:

- **Auro-3D**: This ultimate mode is used to encode *Aurophonic* multi-channel streams, ranging from 8.0 up to 15.1, into a PCM-stream with 4 to 8 channels, depending on the selected configurations.

- **Auro-2D**: This mode is used to encode *Surround Sound* multi-channel streams, ranging from 3.0 up to 7.1, into a *2-channel* PCM-stream.

- **Auro-1D**: This mode is used to encode a stereo-stream into a mono-stream, thereby effectively reducing the bandwidth to 50% (factor 2). It can also be used to combine two separate stereo mixes into one single stereo PCM stream.

- **Auro-CX/Compression**: In this mode, the encoder can reduce the bandwidth of a (stereo) PCM signal with factors from 2 up to 9 depending on the required bandwidth reduction. The resulting files are PCM-streams at lower sampling rates, that can be played by any media player and that provide very high audio quality after decoding, as close to the original as possible.

7 ANNEX 2 – GLOSSARY AND REFERENCES

7.1 Glossary

AAC	Advanced Audio Coding, also known as MPEG 4 audio compression or mp4a
AC-3	Audio Coding 3, the compression method used in Dolby Digital
AES	Audio Engineering Society
BD	Blu-ray Disc
Codec	Combination of encoder and decoder
DAC	Digital-to-Analog Converter
dB	Decibels; unit to express differences on a logarithmic scale
$dB_{SPL}(A)$	A-weighted sound pressure level
dB/W/m	dB per Watt per meter: Sound Pressure Level measured at 1m from the speaker with 1W of input signal (typically a 1kHz sine-wave)
dBFS	Decibels, referenced to digital Full Scale
dBSPL	Decibels expressing Sound Pressure Levels, referenced to 20μPa air pressure
DCI	Digital Cinema Initiative
FLAC	Free Lossless Audio Codec
HDMI	High-Definition Multimedia Interface
IC	Integrated Circuit
Lossless codecs	Codecs that provide bit-by-bit identical copies of the original audio after decoding e.g. FLAC, Dolby TrueHD, DTS Master HD
Lossy codecs	Codecs that use perceptual coding techniques based on psychoacoustics (e.g. masking) to reduce the bandwidth, dismissing some audio content in the process. e.g. mp3, AAC, ac3, DTS
MLP	Meridian Lossless Packing, the compression method used in Dolby TrueHD
mp3	MPEG 1, layer III audio compression
MPEG	Moving Picture Experts Group
PCM	Pulse Code Modulation
SMPTE	Society of Motion Pictures and Television Engineers
SPL	Sound Pressure Level

7.3 References

Beranek, Leo L. *Acoustics.* nr. December 1986. Amer Inst of Physics, 1954.

DCI. *Digital Cinema System Specification, version 1.2.* Standard, Digital Cinema Initiatives, LLC, 2008.

Dijksterhuis, A. P. *Het slimme onderbewuste - Denken met gevoel.* Amsterdam: Bert Bakker (Prometheus), 2008.

EBU. *R-128 - Loudness normalisation and permitted maximum level of audio signals.* Rec., European Broadcasting Union, 2010.

Everest, F. Alton, en Ken C. Pohlmann. *Master Handbook of Acoustics.* McGraw-Hill/TAB Electronics, 2009.

Fielder, Louis D. „Dynamic Range Requirement for Subjective Noise Free Reproduction of Music.” *AES Convention 69, May 1981, Paper Number 1772.* Audio Engineering Society, 1981.

Holman, Tomlinson. *Surround Sound - Up and Running.* Oxford: Elsevier, 2008.

ITU. *Rec. ITU-R BS.1770-1 - Algorithms to measure audio programme loudness and true-peak audio level.* Rec., International Telecommunication Union, 2006-2007.

Katz, Bob. „Bits, isn't 20-bit good enough?” *Digital Domain.* http://www.digido.com/audio-faq/b/bits-isn-t-20-bit-good-enough.html.

Olive, Sean E,. *High School Students Prefer The Sound Quality of Accurate Loudspeakers and CD Over MP3.* Harman Kardon Group, 2010.

Pras, Amandine, Rachel Zimmerman, Daniel Levitin, en Catherine Guastavino. „Subjective Evaluation of mp3 Compression for Different Musical Genres.” *AES 127th Convention, Oct 2009, Paper 7879.* New York, NY: Audio Engineering Society, 2009.

SMPTE. *RP 0200:2002 - Relative and Absolute Sound Pressure Levels for Motion-Picture Multichannel Sound Systems.* Recommendation, Society of Moving Picture and Television Engineers, 2002.

SMPTE. *ST 0323:2004 - Channel Assignments and Levels on Multichannel Audio Media.* Standard, Society of Moving Picure and Television Engineers, 2004.

Watkinson, John. *The Art of Digital Audio.* Oxford: Focal Press, 2001.