TVTechnology

Print Page

The Skinny on Audio Coding

1/21/2004

To start the new year off on a solid foundation, we will begin a multi-part series of articles on audio data reduction, a.k.a. audio compression, a.k.a. audio coding. Ever wonder how Dolby Digital (AC-3), AAC, MP3, PAC, WMA, and other schemes actually get the job done? How do they get all of that audio into such a small package and keep it sounding good? This month we will investigate the requirements for and added benefits of compressing stereo and multichannel audio signals and describe the basics that are common to most all audio coding schemes.

AUDIO COMPRESSION GOALS

Why data-rate reduce audio at all? With AES/EBU and SMPTE standards, we have ways to carry full-bandwidth audio with relative ease, and increasingly even store it all, so why bother? There are several inescapable truths, one being that transmission of audio (or anything) through the air has limits, and the other is that speed and convenience may actually count more than absolute quality sometimes. For example, a standard 20-bit, 48kHz stereo audio signal takes approximately 1.92 Mbps to carry this audio data. Do you know of any way to get that much data into your house in real time? It would be easy if that were the only stream to which you wanted to listen. What about other programs or 5.1 channel surround sound? A 5.1 channel, 20-bit, 48kHz stream of channels runs at almost 6 Mbps-I have seen quality video programs running at this data rate-it is a huge chunk of data! What does this mean? Simple-audio compression allows a balance to be struck between quality and quantity.

Today's audio compression schemes are simply amazing. The fact that a huge variety of audio, multichannel or otherwise, has the potential to be delivered to consumers with a quality nearly indistinguishable from the source is nothing short of miraculous. That being said, all systems are not created equal, and if pushed too hard or used inappropriately they will reveal their flaws.

Some compression systems are designed to be the final link to the consumer and are very efficient (i.e. low bitrate), while others are capable of being decoded and re-encoded. All systems have certain basic functionalities in common, and usually vary only in their intended use and efficiency. Don't be fooled by the hype however: A higher data rate does not always mean that a codec is less efficient, nor does it necessarily mean that it sounds better.

TIME VS. FREQUENCY DOMAIN

Let's jump right in. The first thing we have to do is get comfortable with the idea that audio signals exist in both the time and frequency domains and that the two domains are intimately, and as it happens, inversely related. An audio signal on a standard oscilloscope is being shown in the time domain, while a spectrum analyzer shows it in the frequency domain. The "mess" we can make on a 'scope screen with an audio signal actually looks fairly orderly if viewed on a spectrum analyzer and we need both views to accurately describe an audio signal.

Both signals have level as the Y-axis (vertical), with either frequency or time defining the X-axis (horizontal). The two signals are mathematically the inverse of each other and much information can be gleaned by looking at signals in both domains.

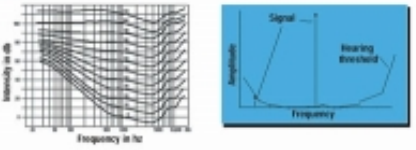
HUMAN HEARING

In the mid-1930's, H. Fletcher and W.A. Munson, two researchers at Bell Laboratories, published a study that showed that human hearing is not equally sensitive at all frequencies and importantly, that this sensitivity changes with loudness. Since then, research has improved the accuracy of the measurements, but the results have withstood the test of time. Basically, quiet low and high frequency sounds fall below the threshold but the ear is much more sensitive to the 1 kHz to 6 kHz region, regardless of loudness.

From this table, a single curve normalized to digital audio levels can be generated and is shown on the right side of Fig. 2. In this drawing, you can see two signals: a large one near the center that is clearly above the hearing threshold, and a small one near the right that is just below the hearing threshold. As we are about to see, this is very useful.

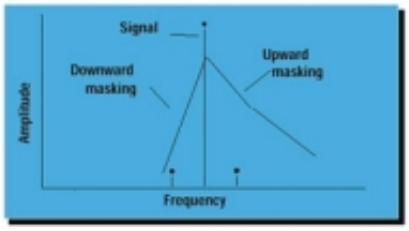
Human Hearing

(click thumbnail)
Fig. 1 A sine wave viewed in the frequency domain (left) and the time domain (right)



(click thumbnail)
Fig. 2 Fletcher-Munson hearing threshold table (left) and normalized hearing threshold (right) Drawing courtesy of Dolby Laboratories

MASKING



(click thumbnail)
Fig. 3 Frequency domain masking. Note that the two low-level frequencies near the fundamental frequencywill be completely inaudible due to the hearing threshold being raised.

Now that we can look at events in both the time and frequency domains and know that there are limits to the threshold of human hearing, some very interesting avenues open up. Think about this scenario: You are standing on a sidewalk talking to a friend when a big yellow taxi drives up next to you and leans on his horn. You continue to talk but your friend can no longer hear you. Has your voice disappeared? No, in fact it is still there and with a frequency domain analyzer (i.e. a spectrum analyzer), you would see very large peaks at the frequencies that make up the car horn, and in the valleys would be the remnants of your voice. Even though the analyzer sees this, because of Frequency Domain Masking it is inaudible to the human ear. Obviously the peaks will cover or mask your voice at those same frequencies because they are louder, but interestingly the human auditory system will also mask frequencies near the peaks. Picture a skirt around one of the frequencies-sort of like a circus tent around the main pole. The tent is the hearing threshold, and the main pole is the frequency of interest. The higher the peak goes, the wider the girth of the skirt and the more masking takes place. Fig. 3 shows this phenomenon. You should also notice that the skirt is not symmetrical; Upward Masking causes the hearing threshold to become less sensitive above the fundamental frequency, while Downward Masking, which has less of an effect, causes the hearing threshold to become less sensitive below the fundamental frequency, hence the asymmetry.

There is another type of masking that takes place when two sounds arrive at the ear in close succession called "Temporal Masking." Basically, sounds appearing after a loud sound stops will continue to be masked (called "Post-Masking") even though the loud sound has stopped. Amazingly, temporal pre-masking also exists and can actually cause sounds to be masked just before a loud sound starts. Post-masking is approximately 10 times more effective than pre-masking, and both depend on the length of the masking signal, but are important effects because highly efficient audio compression schemes rely on saving as much inaudible data as possible.

What you will notice is that underneath the hearing threshold there is a varying amount of "stuff" that is inaudible to the average listener. If it is not audible, and there is a need to fit the audio in a smaller pipe, why not ignore it? Good question-and therein lies the actual magic of audio compression: Knowing precisely what to ignore and exactly how to ignore it. It is a lengthy but interesting answer and will have to wait until next time.

The next Audio Notes will continue our fascinating peek at the guts of audio coding. We will show how coding gain is achieved, begin to investigate some additional tools used to save even more data, and we will discover what certain compression artifacts sound like and why they occur.

Special thanks to Leif Claesson for help with the drawings and to Dr. Deepen Sinha, one of the primary developers of the PAC audio codec for his expert input, clear explanations, and patience with me. If you are interested in finding out more about the psychoacoustic principles presented here, drop me a line and I can refer you to some classic (and readable) texts. **Print Page**