

 [Print Page](#)

Dolby, The Chipmunks And NAB2004

3/26/2004

This time we will briefly wrap up our somewhat detailed look at audio compression by covering areas that some readers have submitted questions about. We will then take a look at a technology that changes the relationship between pitch and time. Finally, I'll tell you about what I will be on the lookout for on the NAB show floor and what I will be stopping by to ask manufacturers.

NO GLUE HERE

You might think that the way to encode 5.1-channel audio is to glue six mono or three two-channel codecs together. I certainly thought this would be a reasonable approach, but as is often the case, what did I know? It turns out that in the case of Dolby Digital (AC-3), Ray Dolby suggested to his codec development team that the better and more efficient way to do things was to code multichannel audio as a complete whole. This allows the technique of high-frequency coupling that we described last month to act across multiple pairs of channels simultaneously to save even more data. Note that this technique is not employed in Dolby E where the channels are coded independently. Because the Dolby E system can actually carry unrelated programs in each channel, high-frequency coupling between channels could create some strange sounds.

Treating multichannel audio as a whole also allows channels that are not demanding as many bits for coding a given section of audio to deposit the extra into a common "bit pool" where they can be applied to other channels to improve quality. These two techniques, while somewhat useful in a two-channel coder, become critical tools in successfully balancing quality and efficiency.

Probably the best introduction to Dolby Digital (AC-3) coding was written by Mark Davis as an article for the now out-of-print July 1997 issue of "Audio Magazine." I saved a reprint of the article because I found it so readable and understandable. Further, it describes a very interesting technique borrowed from analog noise reduction that further improves system performance. Although focused on Dolby Digital (AC-3), it is still an invaluable introduction to many aspects common to most all audio coders.

TIME AND PITCH

I was very lucky to have worked with the research and engineering teams at Dolby Laboratories for several years. One of the most interesting technologies I encountered was one that is embodied in the company's Model 585 Time Scaling Processor. This unit allows the duration of a given audio selection to be stretched or condensed (i.e. time-scaled) with no change in pitch, or to be pitch-shifted with no change in duration. Although not a new idea, there are products that have been around for awhile that can do this. One product that immediately came to mind during the development of this technology was a unit manufactured by Lexicon many moons ago that interfaced to Studer analog tape recorders and allowed a user to vari-speed the machine while it automatically corrected the pitch. It was used quite a bit in commercial production to get spots to exactly 30 or 60 seconds, but it didn't really catch on too far beyond this application.

I bring up the analog tape recorder application because it is probably the easiest way to explain how time-scaling and pitch-shifting are integrally related. If you increase the playback speed of a spoken commercial (for example) on an analog tape machine, the pitch of the audio is increased and so is the speed of the spoken words. If you were then able to shift the pitch down by the same percentage that you increased the playback speed, the result would be that the pitch is the same, but the speech remains faster--you have just time-scaled the audio. A fine example of the reverse situation (pitch-shift) is from my childhood. Remember Alvin, Simon and Theodore? They were from the holiday record "Alvin and the Chipmunks" created by Ross Bagdasarian Sr. (aka David Seville). Here is how he did it: he spoke the words in normal pitch but did so s-l-o-w-l-y and recorded it at one speed on an open reel recorder. Then he played the recording back at a higher speed, and voila! The Chipmunks were born (they had high-pitched voices but spoke at a normal pace). Yup, it works (I tried it in high school once; I forget what I was really supposed to be doing), but it is (and was) tedious and will not work on anything that has been pre-recorded because the pitch will shift and so will the speed.

Since the Lexicon product, modern digital products and software plug-ins have been developed. By various methods, they are all able to change the speed of the audio and thereby the pitch. Usually it is by adding or repeating small sections to make a selection slower (then playing it out at a higher sample rate to restore the timing but enable the pitch to increase), or by removing small slices of the audio where they think it might be inaudible to make a selection's pace faster (then playing it out at a lower sample rate to restore the timing but enable the pitch to drop). Sometimes it works, and sometimes it does not. For example, how do you slice up a sustained violin note?

Unfortunately, in addition to sometimes being very audible in their actions, most all seem to have the drawback of needing to be told what kind of audio they will process. If I have to tell a unit to optimize for speech or music, it is probably going to get into some serious trouble handling a typical film soundtrack that has both speech and music and

switches back and forth without prior notice. The Holy Grail then is a technology that handles both equally well, doesn't need to be told anything about the source material, and knows where to slice or add.

Brett Crocket, the primary inventor of Dolby's time- and pitch-scaling algorithm set out to develop a process that worked independently of the type of audio presented to it. His work stems from research into Auditory Scene Analysis (ASA), or how the human auditory system (i.e., the ear to the brain) breaks apart the continuous stream of different sounds it is presented with into separate perceptual events. What was that? After much discussion with Brett, my simplified take on ASA is that it is the study of how we differentiate between the different components of what we hear: the sound of a piano, cello and two violins plus a conductor's baton being dropped will all be grouped in certain ways into so-called "events." It may mean that a group of sounds are perceived as one event, or possibly in the case of the baton a single sound equaling a single event. This is critical because unlike other processes that may simply rely on the amplitude of a signal for determining where to act, knowing how sounds will be perceived helps the Dolby algorithm make much better informed judgments about where and when to add or drop sounds. The result is extremely natural sounding audio, just faster/slower or pitched up/down. For the complex details of what is going on, I highly recommend Brett's paper from the AES 115th convention (paper 5948), which I am sure will be published in the Journal of the Audio Engineering Society soon.

Some interesting applications arise for a product such as this. One obvious one is the video transfer of 24fps film for 25fps PAL countries. Amazingly, until now this process has simply delivered films with the audio 4 percent higher in pitch. Now, with the timing being increased by playing the 24fps film back faster, the audio can be corrected back down 4 percent and will sound normal. (Well, it will sound normal to us but perhaps not normal to PAL viewers who have become accustomed to Arnold Schwarzenegger sounding a little bit squeakier.)

NAB NOTES

Every year I go hunting for something at this show. I figure you have to have at least a couple of goals or the days just turn into a blur of endless walking and collecting swag. I will of course stop by the Dolby booth and hope that the folks are showing the Model 585. I recommend that you stop by to take a listen--it simply must be heard to be believed. I will be visiting Pixel Instruments to see what they have cooked up for fixing lip-sync issues with certain video switchers. It turns out it was not an audio problem after all, but it took some audio guys to fix it--we'll reveal details in an upcoming column. I will also try to visit all the distribution encoder and decoder/IRD manufacturers to see how they are coming along with properly passing compressed audio formats like Dolby E. Companies like Aastra Digital Video and Tandberg Television have it working now, but some still do not (and you know who you are). Of course I will stop by the Wohler booth to see what they have cooking. Carl Dempsey and crew never cease to surprise and impress the broadcast industry (and me), and I look forward to seeing what they will have this year.

Also, I will join Nigel Spratling, Joe Wellman and friends at the Sigma Electronics booth where there will be several interesting things to see including some potential in-service lip-sync test and correction technologies. Please stop by for some swag, to say hello, or to get a question answered faster than e-mail.

Special thanks again to Dr. Deepen Sinha of Audio Technology and Codecs (ATC), and to Brett Crocket from Dolby who is the primary inventor of the time-scaling process we discussed this month.

For more information on Alvin, Simmon and Theodore, visit www.thechipmunks.com where you can even listen to some  hilarious samples. [Print Page](#)