Reading problems

In my previous 'sound' article I explained that some computer systems (and software) may do a better job of ripping tracks from Audio CDs than others. In this article I look in detail at the effect on the resulting files. For the sake of example I have chosen to use a specific set of audio tracks that I recorded a few years ago onto a (TDK brand) CDR. This was written onto the CDR using my Iyonix and !CDVDBurn.

I used cdparanoia to rip wave files from this CDR using two different machines. One a shuttle running Linux, using a DVD drive. The other, my wife's Iyonix with its CDROM drive. On the Iyonix, cdparanoia ripped with no reports of any errors or problem. But on the Shuttle, cdparanoia listed a string of "+++++" warnings during the second half of the ripping process. The "+" characters are cdparanoia reporting "unreported loss of streaming / other error in read". So a sign that the drive was having problems reading the disc.

In both cases I ran cdparanoia with the "-z" option. This tells the program "never accept any less than perfect data reconstruction". Hence it would re-try reading frames of data over and over again until satisfied it *has* correctly read the disc. So the appearance of "+" characters during ripping doesn't necessarily mean the result has errors. Indeed, you might expect that using "-z" means there *can't* be any errors if the rip gets to the end of the specified track(s) because cdparanoia would have gone on trying until it succeeded – or you interrupted its attempts because it was taking too long!

The two resulting files were the same size, so contained the same number of samples. That was promising. But when I used !WAVDifference to subtract one set of samples from the other I found the resulting difference Wave file *wasn't* just a series of zero-values. The files were different, so at least one of them must contain some erroneous sample values. This was despite cdparanoia showing the ":-)" smiley throughout as if happy that it was successful in overcoming the "+" problems on the way.

To investigate further I wrote a new application in the !WAV series that I have called *!WAV_Compare*. This goes though two Wave files and checks, sample by sample, to see if they agree. It then reports what it found. I could then plot the outcome.

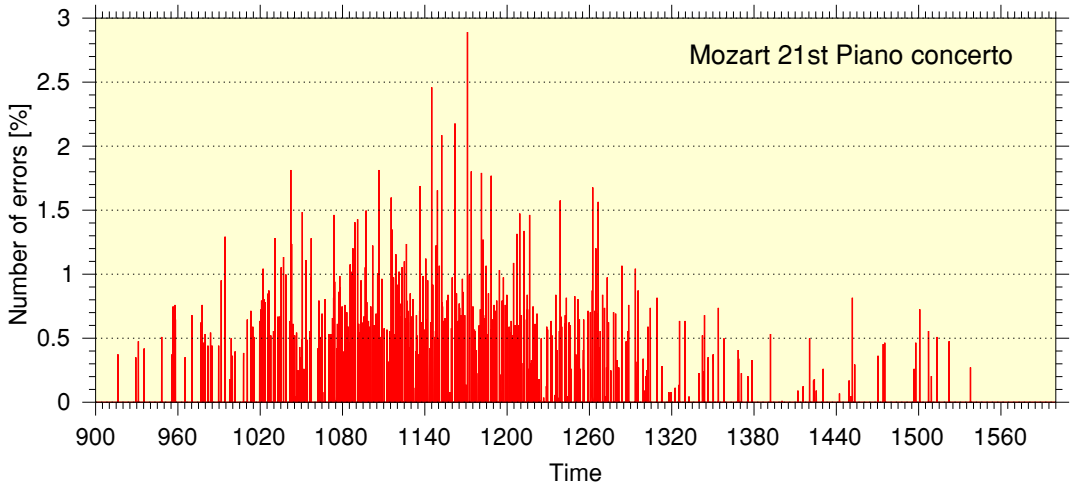

Figure 1 How many errors (differences) occur during each 0.1 sec chunk

The tracks I used are actually a copy to CDR of a performance of Mozart's 21st Piano Concerto. The analysis analyses the two rips in 100 millisecond chunks. It counts how many samples in each chunk are *not* the same in the two rips. Figure 1 plots the result as the percentage of such errors in each 100 millisecond period. Note that the horizontal axis has units of seconds measured from the start of the music.

The graph only displays the part of the music from 900 seconds onward because, before that, the two rips are exactly the same. The highest percentage value is just under 3%. For a (stereo) audio CD there will be 2 x 4410 = 8820 samples in each 100 millisecond period. So an error rate of 3% would correspond to around 260 samples being different between the two rips during that chunk.

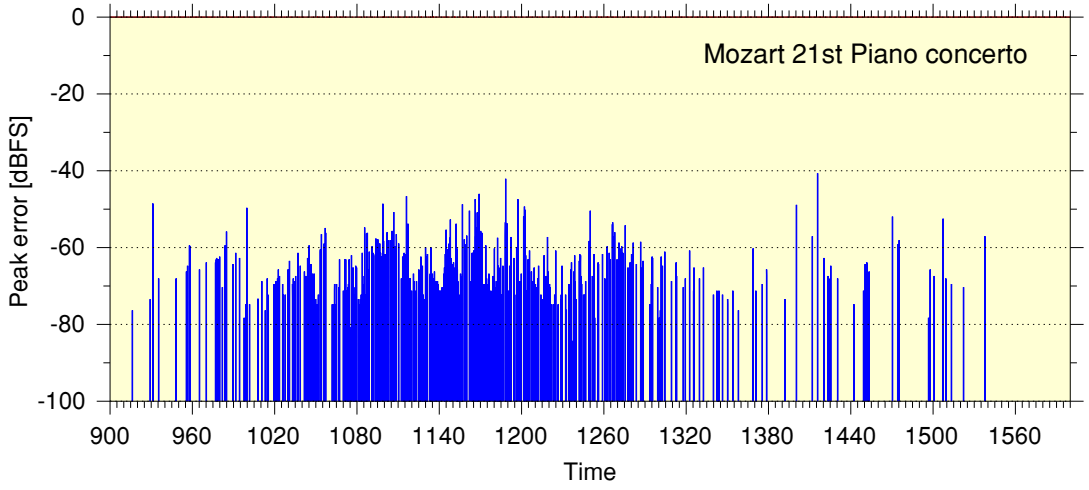

Figure 2 Difference between values in the two rips

Figure 2 plots the amount by which the values differ. To simplify things, the plot actually shows the biggest difference between rip values it could find during each 100 millisecond chunk. The difference in values is then referenced to the maximum signal level. In effect this shows the discrepancy during each chunk as if it were a brief 'noise spike' of the size indicated. For values around -60dBFS or less it seems unlikely that the effect will be noticed. However some of the errors reach around -40dBFS, which may be audible.

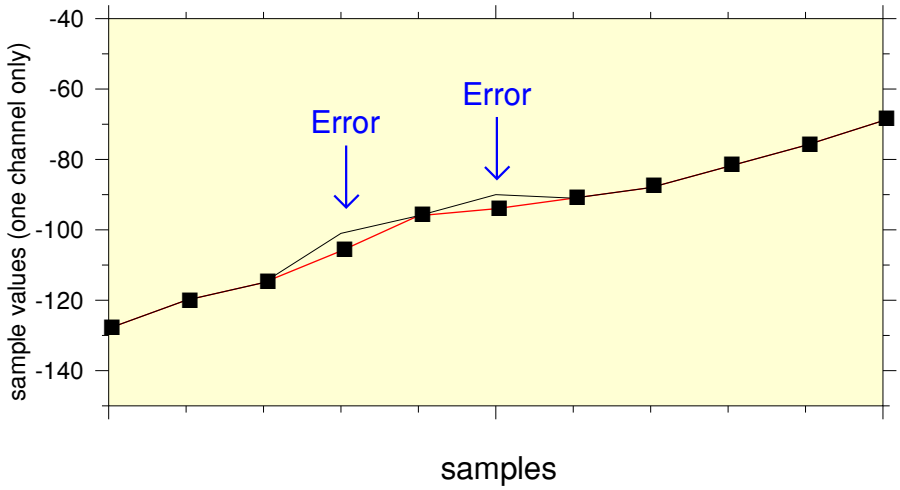

Figure 3 Detailed view of a couple of error events

Figure 3 shows what individual errors (discrepancies between the two ripped files) tend to look like when examined in detail. Here I have just plotted a few successive samples for a channel from the two files. The two lines plotted are for the two different rips of the same audio channel. I have also put squares on the plot to mark the actual sample values for one of the rips. If the two rips

were identical you would only see one line on the graph. But where the rips differ you see two lines. There are two errors apparent in this section. The sample values actually only differ by 3 or 4, which is quite small compared to the 16-bit range of audio CD values.

On the basis of the above it is tempting to conclude that the errors generally don't really matter much. Typically, the discrepancies only affect around 1% or fewer samples, and the actual change in level is small. So the two versions should sound close to indistinguishable. However there are some grounds for caution.

Here I've only looked at one example of a CD that gives different results when ripped with two different systems. So it may well be the case that other discs and ripping setups give worse problems. My impression, though, is that if cdparanoia continues to give the ":-)" smiley and rips the specified tracks despite using the "-z" setting, then the errors may be minor. More serious errors would probably cause cdparanoia to 'stall' and go on re-reading the same frames until told to quit. So it wouldn't complete the process of ripping. I've not (yet) checked, but I suspect that when the program gives "-" warnings, but continues apparently happily, then again the errors may be minor. But all that said, my analysis shows that using the "-z" setting and having the rip complete with a ":-)" throughout does *not* guarantee the result is totally free of any sample value errors. So you can't be certain that a rip is perfect.

A curio here is, what is the nature of the errors, and how do they cause such - modestly small - discrepancies in values? If the errors were random changes of an individual bit in a raw binary sample data stream then the changes produced would be just as likely to be large than small. This is because a significant bit is just as big a target as a less-significant one. Yet the errors I found all tend to produce small changes. This is strange, but it may be a result of the coding systems used for audio CD. The raw 'channel bit' data stream on the CD isn't simply the stream of sample values as LPCM. It is encoded and wrapped into a number of layers of error protection. So maybe one of the features of that system is that there is a tendency for errors to be small once the encoding, etc, are unscrambled. On the other hand, maybe the changed values are actually being produced by cdparanoia by a process of 'guesswork' or – more politely - interpolation. That is plausible, but it isn't obvious from figure 3 that this is what is happening.

Finally we are left with a more basic question. How can be we certain which of the ripped versions is 'wrong'? The only evidence we have is that the results differ for some sample values, and that one ripping process issued some warning "+" indicators. Taken at face value, the "+" warnings tell us the ripping process was struggling. But cdparanoia seemed happy that it *had* read the results OK despite this. So maybe it had, and the rip using the Iyonix is the version that actually contains errors - despite it *not* giving any warnings. Maybe cdparanoia on the Iyonix misses problems, and so doesn't correct them, when the Linux system did.

To set against that disturbing thought, I can note that for some other audio CDRs, the Iyonix system does report either "+" or "-" warnings. So the problem isn't that the Iyonix can't actually detect errors. This is encouraging and makes me feel that the Iyonix is getting better results. Leaving the Linux setup as being the one that is producing some erroneous results. Although of course, both might be wrong at times! With these questions in mind, I'll finish with another question you may like to ponder. How can we check which system may be in error, and check to see if, at times, *both* make errors without giving any warnings?

1400 Words

Jim Lesurf

15th Mar 2012