E4896 - Music Signal Processing

Alexandros Eleftheriadis
Department of Electrical Engineering
Columbia University
eleft@ee.columbia.edu

Corrections to the textbook "Total Recording", by D. Moulton

January 13, 2004

p. 156: Typo in caption of Figure 2.80A-B: "40 Az" should read 40 Hz.

p. 157: A great way to demonstrate the nonlinearity that gives rise to intermodulation products is to construct a graph that shows the input/output relationship of the nonlinear system (it will look like the plot of $y=x^{1/2}$). You then construct two plots on this graph. The first one will use the original positive 'x' axis as amplitude axis and the negative 'y' axis as the time axis for the *input* waveform. Then we construct the corresponding *output* waveform plot by using the negative 'x' axis as time and the positive 'y' axis of the original plot as amplitude. This way we can clearly show how different amplitudes in the input get mapped very differently in the output. So when two waveforms are superimposed, the output includes the IM byproduct and not just the combination of the outputs that would correspond to each individual input.

p. 189: In the 5th line, "0.000,1 Volts" should be "0,0001 Volts". Also, in the same page, the discussion of optical transmission should mention that ADAT can carry 8 channels at a time (originally stored on S-VHS video tape), and also mention Toslink which carries essentially a stereo S/PDIF signal. (I believe that this was originally written as TOS-Link, and was related to Toshiba, but now everybody seems to be using the simplified spelling.)

p. 190: Definition of "word" in the first paragraph appears to be incorrect. A word traditionally refers to the smallest amount of data that a processor can handle. It is typically a multiple of bytes, and it does not normally include additional information. In some cases, internal hardware may include parity bits that can be used for error detection. Typically this information is not visible to the end-user or even programmer.

p. 191: The discussion of dithers should mention POW-R dither.

p. 193: In the "Oversampling" section, it should be pointed out that oversampling trades off sampling rate for word length: by increasing the sampling rate several time, we can correspondingly reduce the number of bits used to represent each sample. At the limit, with an infinite (very high) sample rate, a single bit is sufficient to represent the original signal. This is the basis of all modern Sigma-Delta oversampling A/D converters, as well as the DSD format.

p. 195: The last paragraph of the page, which refers to the non real-time nature of DAW's is no longer true, surely for high end systems likely to be used in professional facilities. Pro Tools, SADiE, Sonic Solutions, all run the vast majority of their available algorithms in real-time. Typically, very complicated multi-pass algorithms may not be able to be run in real-time (like

AudioSuite plugins in Pro Tools, but that's due to the nature of the algorithm not a limitation of the hardware). Of course, for low end host-based systems (no special audio hardware) such as Cubase, Digital Performer, etc., real-time performance problems do exist for complicated algorithms.

p. 259: Explanation of HF bias appears to be incorrect, or maybe incomplete. If the bias signal is not recorded or recorded at a very small level, then how does the actual biasing occur? Not being recorded means that the total signal value will fall back to be just the input component and hence you get back the zero-crossing problems. (I am not familiar at all with the physics of the actual recording process so I can't point out where the problem may be, but it appears that a small but important piece of information may be missing.)

p. 265: The footnote referring to backup speeds needs to be updated as current backup mechanisms to CD-R/DVD-R or data DATs are considerably faster than real-time. Only audio DAT backups would be limited to real-time.

p. 267: The notion of convolution presumes a linear (and time invariant) system. In that sense, modeling of "real" nonlinear audio devices requires more complicated operations than convolution. It also implies that invertibility is not always possible.

p. 332: ITU stands for International Telecommunication Union.

p. 370: In the third paragraph, first sentence, the term "formant" is not defined or discussed elsewhere, contrary to what's mentioned in the text.

p. 390: Footnote 267 appears to be missing (cited at the end of the first sentence in "Reverb Time".

p. A-3: In the description of Track 18, I do not understand the sentence: "Finally, we modulate the spectrum of the a band of noise" (sic).

p. A-4: Track 25 on the CD is identical to Track 24. Also, the explanation of Track 25 is not clear.

p. A-4: Referring to Track 26, the notion of auditioning the response curves of our ears can be confusing. The only way I can think of making sense of this is to make sure that one auditions all samples at 83-85 dB SPL so that he/she is in the most linear range of the Flether-Munson curves. Another way to look at this is the following. If we audition pink noise at 40 dB SPL, then what we hear *is* the frequency response of our hearing at 40 dB SPL! It is impossible to remove the hearing system from the process. If you apply the weighting curve, it's the same as applying the system function twice (one artificially, and one by our ears).