ELECTRONOTES

WEBNOTE 08/30/2009

ENWN4

LINEARITY – USE OF THE TERM IN MUSIC SYNTHESIS

There is a fair amount of confusion about the term "linearity" as it is used in engineering and music synthesis. The term seems to imply a "line" – presumably a straight line. It is often said that a system with an input/output relationship that is a straight line is linear. Is it? Is that all there is to it?

SUPERPOSITION – and STRAIGHT LINES!

Well let's look at another definition. In a general sense, we look for a linear system to support "superposition". That is, if a system S produces an output y_1 in response to an input x_1 and an output y_2 for an input x_2, then if the inputs are superimposed (x_1+x_2), the outputs should be superimposed as well, (y_1+y_2). This provides a well-defined test.

Consider a system y = S(x) = 2x. It just doubles the value of the input, a voltage amplifier with a gain of 2 if you like. For example, if x_1=2 then y_1=4. If x_2=3 than y_2=6. It is true that if we first sum 2 and 3, giving 5, than the output is 10, the sum of 6 and 4. Superposition applies. The system is linear. The relationship is a straight line (Fig. 1).

Another system might be y=S(x) = 2x + 1. This is also a straight line (see again Fig. 1). Input 2 and you get an output of 5. Input 3 and you output 7. Input 2+3=5 and you get an output of 11. But the sum of the outputs of the individual cases is 5+7=12. So superposition does not apply, and the system (while a straight line) is

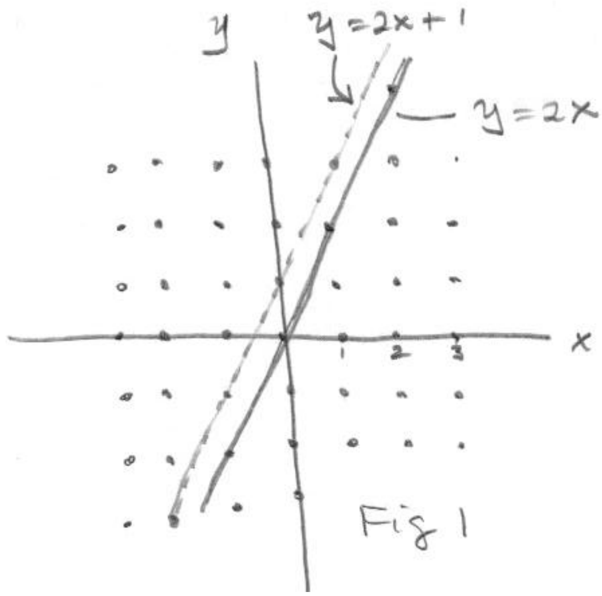


Fig 1

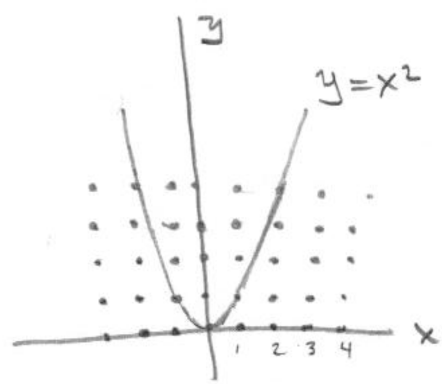


Fig 2

not linear. [You can easily see that we might require the straight line to go through the point x=0, y=0 to get linearity.]

Likewise, $y = S(x) = x^2$ is not linear. If x=2 than y=4. If x=3 then y=9. But if x=2+3=5, y=25, which is not the sum of 9+4. This relationship is not a straight line (Fig. 2). Neither is y= S(x) = sin(x) linear, for another example.

So, we discover something. Linearity is mathematically rare – very rare. On the other hand, in electrical engineering, linearity is common – very common. More on this a bit later.

LINEAR EQUATIONS ARE DIFFERENT

Now, you may be screaming that the equation y = 2x + 1 is a linear equation. Indeed, the problem of having two "simultaneous linear equations" is common in engineering. Indeed, we could have a second linear equation y = -3x -4. We can easily solve these to see that the two lines intersect at x=-1, y=-1. So here we have a different meaning of "linearity". The equations are linear in that they are only first order in the variables x and y. We don't see any x^2 or y^π or cos(x).

LINEAR SYSTEMS – RESPONSE TO SINE WAVES

Another way to tell if a system is linear is to input a sine wave (in general, a complex exponential). If what we get out is a complex exponential of the same frequency, and nothing else, the system is linear. Indeed, both the amplitude and the phase of the sine wave generally change from input to output – this is what we mean by the "frequency response" of the linear system (like a filter).

A typical linear system in music synthesis would be a VCA (amplitude change only), or indeed, a VCF (amplitude and phase changes depending on frequency).

Typical non-linear circuits in music synthesizers are "ring modulators" and clippers. Typically these produce what we call "distortions" but one man's distortion is usually another man's special effect!

Four more points about linearity in music synthesis:

(1) When we talk about linearity, we of course mean it is linear for practical purposes – there may be small non-linearities.

(2) The famous "exponential" relationship between pitch (frequency) vs. control voltage is highly non-linear. But it produces no distortions of the audio signal, which does not pass through the control-voltage converter.

(3) A combination of a linear system, and then a gradual non-linearity (such as rounding or clipping at higher amplitudes) is sometimes the hallmark of a device that "sounds good."

(4) Quantization following sampling is non-linear, although we often correctly consider sampled system, by themselves, to be linear.

LUCKY EE's

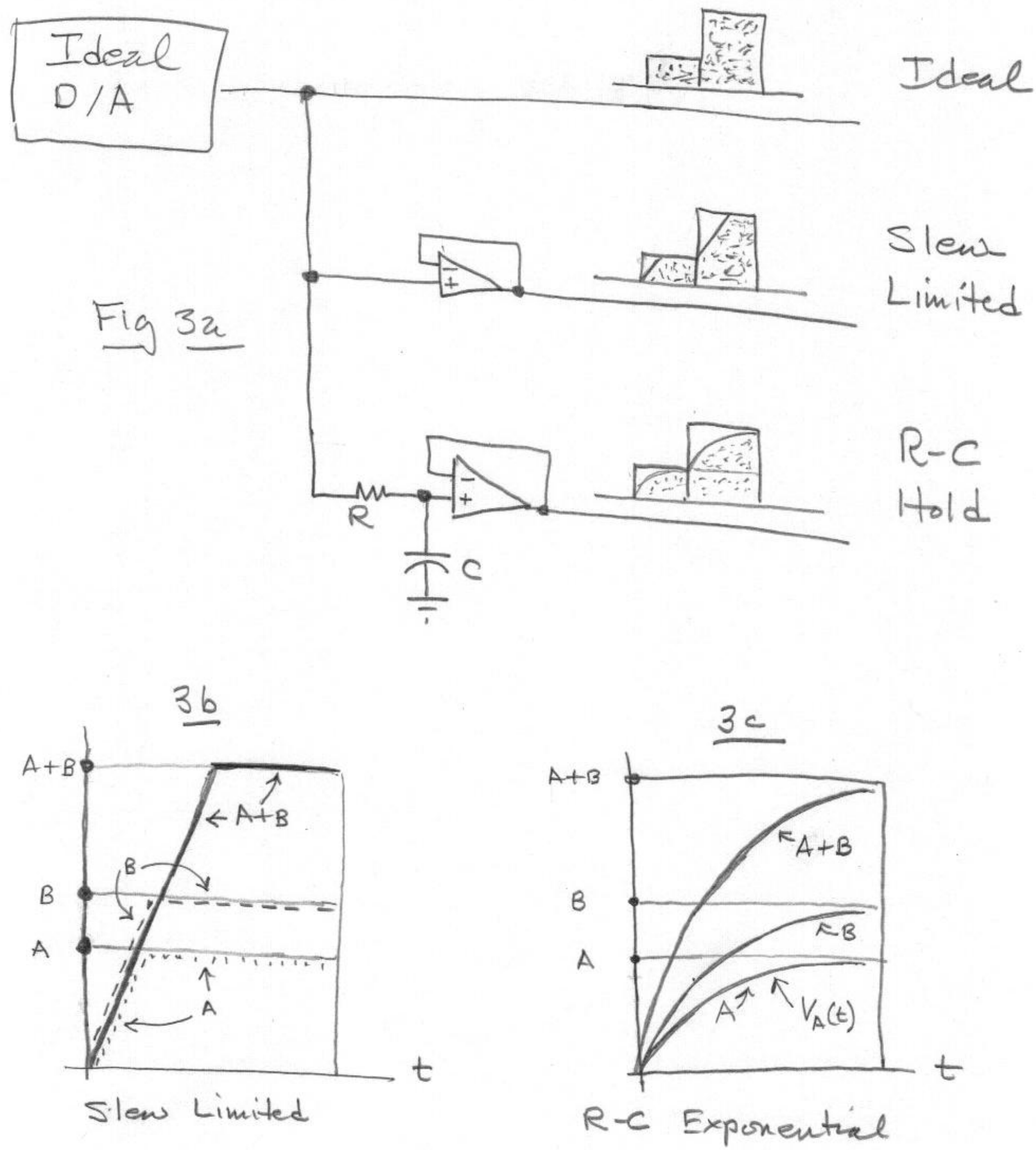
We have said that in electrical engineering, linearity is common. We do a lot with amplifiers and filters. Our devices, such as op-amps, support linear systems over wide ranges of amplitude. The same is not necessarily true for mechanical engineers, etc. They work with materials. For example, we all remember "Hook's Law", f=-kx, the force on a spring is proportional to its elongation. That's linear. But, very soon, as the amplitude increases, non-linear terms must be included. Electrical engineers needing to study non-linear systems are often encouraged to look for a course in the mechanical engineering school.

In music synthesis, we have sometimes used non-linear chaotic systems to generate interesting complex signals. We had to make this happen. Mechanical engineers too often get chaos for free!

WHEN IS A STRAIGHT LINE NON-LINEAR AND AN EXPONENTIAL CURVE LINEAR?

True enough, instructors sometimes delight in throwing students a curve ball. The idea is not to simply confuse, but to force students to not just assume. One of my favorite examples is to consider a "hold" device following D/A conversion. Fig 3a shows a non-existent "ideal D/A" with two possible "real" op-amp output stages.

The first output stage is just an op-amp follower. Hit by the (ideal) instantaneous transition, the op-amp goes into slew limiting, heading in the required direction as fast as it can, in a straight line. When it gets there, it stops. The result is a triangular "chase" followed by a rectangular hold. Because the slope on the triangular side is a constant (like volts/second), the duration of the triangular chase is greater for larger step amplitudes, and the rectangular hold portion is correspondingly shortened. So if we had a step of size A and a step of size B, these would have different chase lengths (see Fig. 3b). A step of size A+B would have yet another, longer, chase length. The responses are not scaled versions of



each other. Thus the slew limiter, despite making straight line segments, fails the superposition test. The result <u>is not linear</u>.

Consider next the output stage consisting of an RC low-pass followed by an op-amp (slew limited real op-amp of course), sometimes called an "R-C Hold". Now when a step appears the output charges exponentially toward the level of the step, whether this level is A, B, or A+B (Fig. 3c). For example, the exponential rising toward A is:

$$V_A(t) = A(1 - e^{-t/RC}) \quad (1)$$

While the other two curves are:

$$V_B(t) = B(1 - e^{-t/RC}) \quad (2)$$

$$V_{A+B}(t) = (A+B)(1 - e^{-t/RC}) \quad (3)$$

which are clearly all scaled versions of the exponential charging, $(1 - e^{-t/RC})$. Here superposition applies and the system is linear. But the output "curves" are curves, exponential functions – NOT straight lines.

We thus arrive at the curious situation that the straight line output segments are NOT linear while the curved line output segments ARE linear. A good joke.

Two more points need to be made with regard to the choice of the RC product (the "cutoff" of the low-pass):

(1) Above for the R-C Hold we have assumed there is no slew-limiting, even though the op –amp in the R-C Hold is real and thus, like the one above it in Fig. 3a, has a finite slew. It is possible to choose the RC products such that no slew limiting can occur. Let's consider equation (1) for the case where A is the maximum step size we can expect – perhaps 10 volts. The slope of the charging curve is:

$$dV_A(t)/dt = (A/RC)e^{-t/RC}$$

It is easy to observe that the maximum slope is at t=0 (the slope is greatest at the start), and it is A/RC (volts/second) there.   We have only to choose RC so that A/RC is less that the data sheet slew rate of the op-amp. For example, for A=10 volts and max slew rate 0.5 volts/microsecond (a type 741), we get RC should be greater than 0.00002.

(2) Related to this, we have put a low-pass filter in the output. This means that we are going to get a roll-off with frequency and some phase shift. The R-C low-pass is of course a linear system, and we must expect this "frequency response" to be something we have to look at.   Where is the cutoff?   Well, it's at $f_c=1/2\pi RC$. For RC = 0.00002, this is 7958 Hz.   This might not be great for good audio. Is this a problem?

Not really. We can always use a faster op-amp (5 volts/microsecond types are cheap and common), or we can recognize that full size steps are highly unlikely, and reduce the RC product.