AVB Application Requiring Better than 5uS Accuracy of the PTP Clock

Carl Christensen
Principal Engineer
Harman Pro
cchristensen@harman.com (801) 352-2471

Recently there has been some discussion regarding how much accuracy is needed in the real time clocks of AVB nodes. Specifically, the possibilities of needing no better than 5us of accuracy have been discussed. One area where the accuracy of the clock between nodes could have an impact is in speaker arrays.

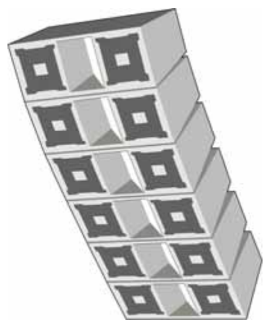

Figure 0

One example of a speaker array is the line array that has seen a large increase in popularity over the last decade. Figure 0 is an illustration of a line array. Figure 1 shows a block diagram of one element of an AVB-based line array. In each element an AVB receiver block receives the audio stream from the AVB network and outputs data to a digital-to-analog converter. The AVB receiver block also uses PTP to maintain a real-time clock (RTC) represented in the block diagram by the analog wall clock symbol. A word clock generation block uses this RTC to develop the clocks used by the DAC. The output of the DAC is amplified and feeds a speaker. In such a system, all DSP including intentional timing adjustments between elements is done outside of the individual elements. Figure 2 shows a conceptual view of this type of system. The accuracy or amount of difference allowed between any two of the RTC or wall clocks is the primary concern of this paper.

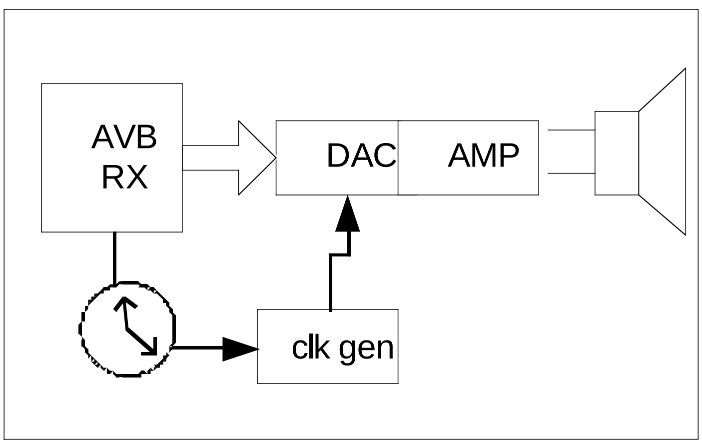

Figure 1

Line arrays are designed to produce sound in a large area using carefully designed speaker pattern overlap often referred to as beam engineering. Modern beam engineering uses a combination of physical placement and digital signal processing. Uncertainty in the phase relationship of the word clocks driving the digital to analog converters is a known source of system degradation. Accuracy of the AVB RTC will have a direct impact on the phase accuracy of the word clocks.

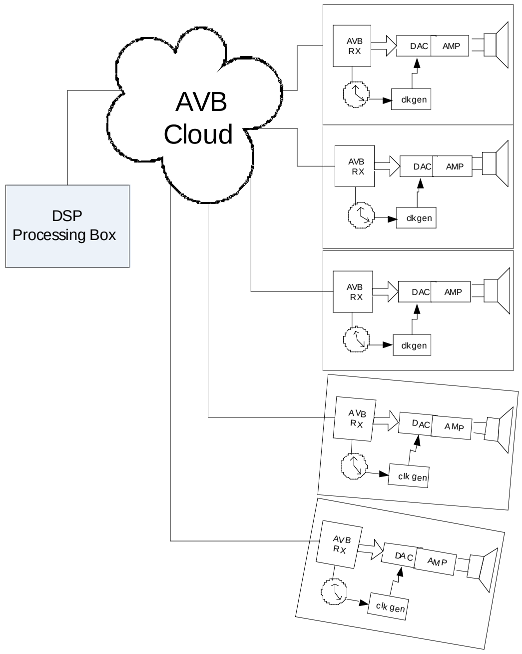

Figure 2

To illustrate how phase differences in the clocks driving the DACs can cause degradation, let's look at a simplified example of two elements and what the listener would hear at a location where the beam patterns of the two elements overlap (see Figure 3). For our example, the signal or "tone" can be represented by the equation:

$g(t) = Sin(t) + 2sin(3t) + 0.3sin(10t)$

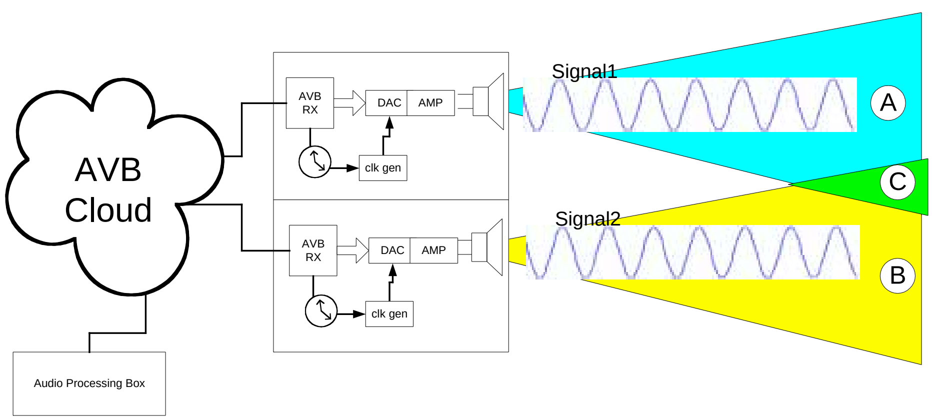

Figure 3

A plot of this signal varying over time is shown in Figure 4.

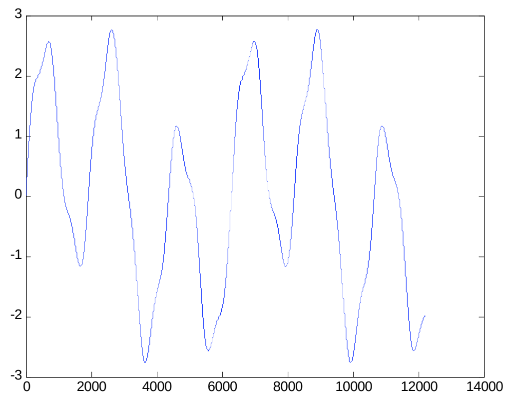

Figure 4

For our example we will assume that the top element's word clock is "correct" and we will vary the phase of the word clock in the second element. Thus a listener at point A would hear the signal g(t). A listener at point B would hear g(t + ***pe***) where ***pe*** is the time difference between the word clock in the lower element relative to the word clock in the upper element.

B(t) = g(t + ***pe***)

Our simplified example will assume that a listener at point C will hear a linear mix of the two signals.

C(t) = ½ A(t) + ½ B(t) = ½ g(t) + ½ g(t + ***pe***)

A plot of the waveforms of the sound pressure at the three points is shown in Figure 5.

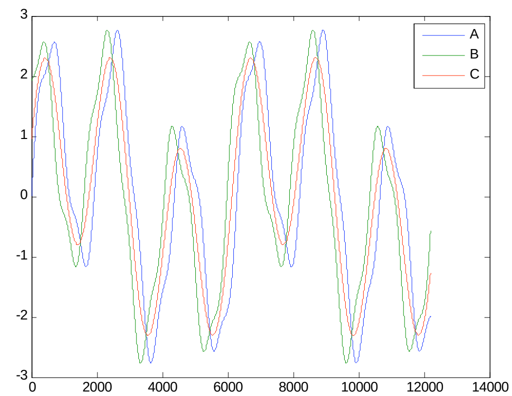

Figure 5

In this example it is evident that the shape of waveform C is different from that of the waveform at A or B. The phase difference in the word clocks has resulted in what we call destructive interference. A reasonable way to quantify the difference in shape of A and C would be to adjust the phase of C by ½ of pe (see Figure 6) and subtract C from A.

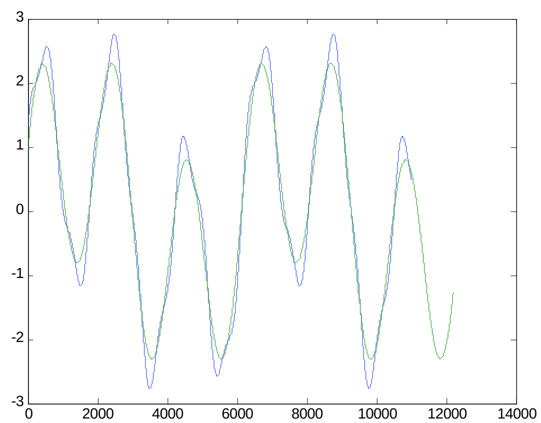

Figure 6

Error = A(t) – C(t + ***pe***/2)

A plot of the magnitude of the error over time is given in Figure 7.

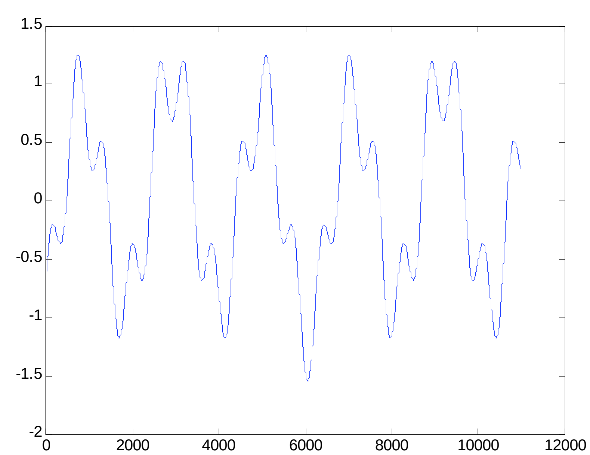

Figure 7

Notice that the error is quite periodic and thus both measurable and audible.

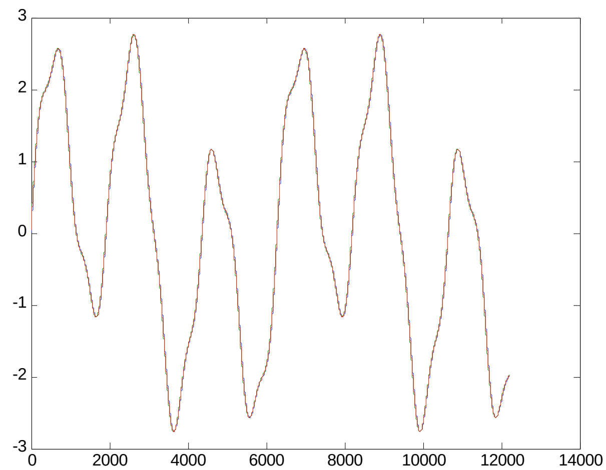

Figure 8

Now let's reduce the magnitude of the phase error between the two word clocks. A plot of the three signals is given in Figure 8. A zoomed image is shown in Figure 9.

A plot of the error signal shows a similar shape (see Figure 10), but a reduced magnitude. Figure 11 gives a visual representation of the difference in magnitude.

It should be noted that this is a very simple example that ignores many realities of acoustic engineering. We've assumed a perfect wave combining, ignored the structure or the room and associated reflections, and a host of other issues. The example is to illustrate the method by which phase differences in word clocks cause destructive interference and impact audio fidelity. It does not represent a real world example.

In the real world, the complexities of audio engineering make the effects of any phase differences in the word clocks highly application-specific. As such, a line array designer will normally specify phase accuracy for word clocks from previous experience / rules of thumb and design the array and then test and characterize the system at the extremes for a number of different signal types.

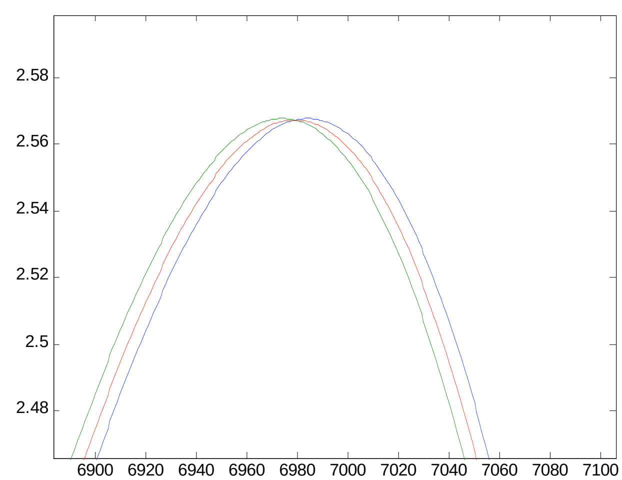

Figure 9

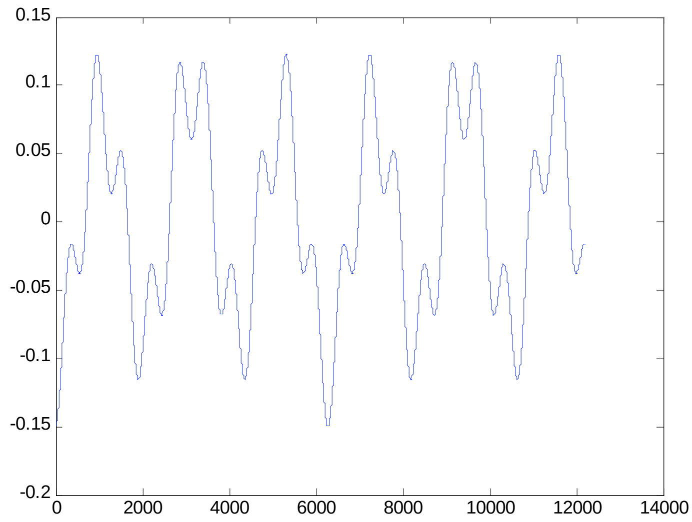

Figure 10

Current practice is to specify a maximum phase error of 15 to 30 degrees of the highest frequency of interest. That has proven to minimize destructive interference in line arrays and other systems that require channel-to-channel timing requirements. It would be desirable if future technologies could enable reducing this to 10 degrees or less. The highest frequency of interest is normally specified as 16 or 20KHz.

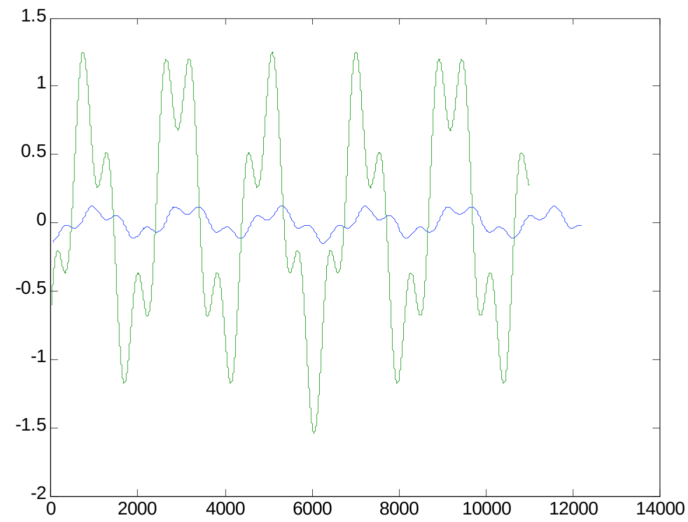

Figure 11

Table 1 shows the allowable time shift for various combinations.

Table 1

Freq of interest	Degrees of phase shift	shift in %	allowable time shift
20Khz	30	0.083333	4.17 uSec
20Khz	20	0.055556	2.78 uSec
20Khz	15	0.041667	2.08 uSec
20Khz	10	0.027778	1.39 uSec
16Khz	30	0.083333	5.21 uSec
16Khz	20	0.055556	3.47 uSec
16Khz	15	0.041667	2.60 uSec
16Khz	10	0.027778	1.74 uSec

This is the accuracy needed between the edges of the derived word clocks. Obviously the method used to create these word clocks from the AVB RTC will also have an affect on this number and intuitively requires that the accuracy of the PTP clocks be better than the numbers in the table to allow some uncertainty in the word clock generation.

The same basic ideas are also relevant in other applications where mixing is involved. This would include applications where the mixing is electrical rather than acoustical. In such applications the phase error of the clocks driving the analog-to-digital converters is what must be minimized. Some current systems that mix digital audio streams today use AES-3 as the transport mechanism and AES-11 as a timing specification. AES-11 requires that outputs be aligned within 5% of the reference clock. Interestingly 5% of a 48KHz sample is 1.04 uS.

The same principles apply to multi-element speakers. While today multi-element speakers are normally driven by a single source and rely on analog crossover circuits to create the signals for each element, a uniquely-generated stream delivered over AVB has many potential advantages and would follow previous patterns of technology moving from the professional to consumer arenas.

Some of the early work in AVB assumed that an accuracy of 1uS could be achieved. Such accuracy would allow for the desired move to 10% phase error and leave some room for uncertainty in word clock generation. An accuracy of 5 uS is not adequate for higher quality systems having 20KHz bandwidth and is barely usable in the 16KHz bandwidth. An accuracy in the 2 – 3 uS would be more useful, but clearly not as desirable as the 1uS accuracy of the early AVB goal.