

IN DEPTH INFORMATION - CONTENTS

In Depth Information

<i>Room Simulation for Multi-channel Film and Post prod.</i>	<i>2</i>
<i>ADA 24/96 Sample Rate Conversion filters</i>	<i>7</i>
<i>Clock, synchronization and digital interface design of System 6000</i>	<i>9</i>

ROOM SIMULATION - FOR MULTICHANNEL FILM AND MUSIC

Introduction

For more than 10 years TC Electronic has put considerable research resources into perfecting digital room simulation. Our achievements have been made possible by combining scientific room models with intensive listening tests and perceptual adjustments.

This is also the case with the VSS5.1 room simulator which is the crown jewel of the algorithms included with System 6000.

For this specific research project, the work started in 1995 on powerful Silicon Graphics computers, so we could ready ourselves for today with plenty of real-time DSP power available. Some of our work was made public in an AES paper written in April 1999 and published September at AES in New York City.

The unique combination of objective tests and scientific modeling described has won acceptance in our industry and is now the angle generally used for finding the right balance between emotional impact, controllability and range of credible results.

VSS5.1 covers from natural over super-natural to wild, wild spaces.

Room Simulation for Multichannel Film and Music

Presented at the AES 107th Convention, New York, NY, 1999 September 24-27. Republished with permission from the AES, www.aes.org

KNUD BANK CHRISTENSEN AND THOMAS LUND

TC Electronic A/S
Sindalsvej 34,
DK-8240 Risskov,
DENMARK

To fully exploit new Reverb and Spatial algorithms for 5.1 and 7.1 environments in music and film, effect mixing procedures may have to be changed slightly. The paper will describe production techniques for positioning of sources and listener in virtual rooms, and an algorithm structure to achieve this.

0. INTRODUCTION

Typical production techniques involve only one or two inputs to the room simulation system, thereby limiting the precision of source positioning to be only a matter of send level differences and power panning. This "one source - one listener" model is not very satisfying when producing for mono or stereo, but even worse when the reproduction system is multichannel.

Multichannel recording and reproduction is an opportunity for the production engineer to discriminate deliberately between scenes or instruments heard from a distance, and sources directly engaging the listener.

For film work, engaging audio has a very pronounced effect for stimulating the viewer emotionally, and may therefore significantly add to the illusion presented by the picture.

In the search of more authenticity in artificial room generation, long term studies of natural early reflection patterns have led us to propose new production and algorithm techniques. Using ray tracing in conjunction with careful adjustments by ear, we have achieved simulation models with higher naturalness and flexibility, which is the basis of true source positioning.

The paper will discuss two aspects of precise room simulation for multi source, multichannel environments to cover distant and engaged listening:

- Present different production techniques
- Describe an algorithm structure to achieve the objectives

1. SINGLE SOURCE REVERB

By having only one or two inputs in a room simulator, the rendering is based upon multiple sources sharing the same early reflection pattern, and therefore it is not really convincing.

In the real world, all actors or instruments are not piled up on top of each other.

ROOM SIMULATION - FOR MULTICHANNEL FILM AND MUSIC

1.1 Music Production

In many studios, one good reverb is used to render the basic environment of a particular mix. One aux send, set at different levels on the different channels, is used to obtain depth and some complexity in the sound image.

To obtain a sound image of a higher complexity and depth, several auxes and reverbs have normally been used.

Tuning of the levels, pans and reverb parameters in such a setup may be very time-consuming.

For effect purposes, anything goes, but if the goal is a representation of a natural room or a consistent rendering of a virtual room, it may be hard to achieve using conventional reverbs.

1.2 Film and Post Production

For applications where picture is added to the sound, several psychological studies have proven audio to be better at generating entertainment pleasure and emotions than visual inputs. When it comes to counting neurological synapses to the brain, vision has long been known to be our dominant input source. However, a study by Karl Küpfmüller [4] has suggested, that stimulation of even our conscious mind is almost equally well achieved from visual compared to auditive inputs.

Sense	No of Synapses	Conscious Input, bps
Eye	10.000.000	40
Ear	100.000	30
Skin	1.000.000	5
Smell	100.000	1
Taste	1.000	1

Stimulation of conscious mind [4]

Realism in audio is just as important when it is accompanied by picture.

In multichannel work for film, several reverbs configured as mono in - mono out are often used on discrete sources. By doing so, the direct sound and the diffused field are easy to position in the surround environment. The technique is therefore especially effective for point source distance simulation.

As an alternative, several stereo reverbs are used on the same sources to achieve a number of de-correlated outputs routed to different reproduction channels.

With both approaches, adjustments can be very time-consuming, and a truly engaging listening experience is difficult to achieve.

2. MULTIPLE SOURCE ROOM SIMULATION

To obtain the most natural sounding and precise room simulation, an artificial reverb system should be based upon positioning of multiple sources in a virtual room. Each source should have individual early reflection properties with regards to timing, direction, filtering and level.

We have found this to be true for both stereo and multichannel presentations.

If the target format is 5.1, at least two directional configurations should exist in the room simulator, namely for home (110 degree surround speakers) and theatre (side array surround speakers) reproduction.

The room simulator should also be flexible enough to easily adopt to new multichannel formats, e.g. the Dolby EX scheme.

By changing the production technique slightly, multiple sends from e.g. the Auxes, Group busses or Direct outs of the mixing console can be used to define several discreet positions as inputs to the room simulation system.

From a production point of view, multiple source room simulation can be configured two ways, as described below. Any large scale console build for stereo production can adapt to both routing schemes.

2.1 The Additive Approach

The conventional approach to reverb is additive. Dry signals are fed to the reverb system, and wet-only signals are returned and added at the mixer.

With a multiple input room simulator, this configuration works much better than with an single source reverb, because at least each source can be approximated to fit the nearest position rendered. However, normal power panning still needs to be applied in the mixer.

An even more precise rendering can be achieved using the integrated approach described below.

2.2 The Integrated Positioning Approach

The sources in a mix needing the most precise positioning and room simulation, should be treated this way:

The source is completely positioned and rendered into a precise position by passing the dry signal through the simulation system, from which a composite output from a number of source positions are available.

XY positioning to any target format, stereo or multichannel, will be rendered as a best fit. The positioning parameters (replacing conventional power panning) can be controlled from a screen, a joystick or discrete X and Y controls.

With all positioning done in the room simulator, consoles made for stereo production may thereby overcome some of their limitations.

3. ALGORITHM STRUCTURE

This part of our paper describes a generic algorithm currently in use for Multichannel Room Simulator development. It is not a description of any particular present or future product, but rather a presentation of the framework and way-of-thinking that has produced our latest Room Simulation products and is expected to produce more in the future.

ROOM SIMULATION - FOR MULTICHANNEL FILM AND MUSIC

3.1 Design conditions

The overall system requirements can be stated as follows:

- The system must be able to produce a natural-sounding simulation of a number of sources in acoustic environments ranging from "phone-booth" to "canyon"
- The system should not be limited to simulating natural acoustics: Often quite unnatural reverb effects are desired, e.g. for pop music or science fiction film effects.
- The system should be able to render the simulation via a number of different reproduction setups, e.g. 5.1, 7.1, stereo etc.
- The system should be modular so that new rooms, new source positions in existing rooms, new source types or new target reproduction setups can be added with minimal change to existing elements.
- The system should be easily tuneable: In our experience, no semi-automatic physical modeling scheme, however elaborate, is likely to produce subjective results as good as those obtained by skilled people tuning a user-friendly, interactive development prototype by ear.

Fortunately there are a few factors that make the job easier for us:

- There are no strict requirements for simulation accuracy: Certainly not physical accuracy (the sound field around the listener's head), and not even perceptual accuracy (the listener's mental image of the simulated event and environment). The listener has no way of A/B switching between the simulation and the real thing, so only credibility and predictability counts: The simulation must not in any way sound artificial, unless intended to, and the perceived room geometries and source positions should be relatively, but not absolutely, accurate.
- Moore's Law is with us. The continual exponential growth in memory and calculation capacity available within a given budget frame has two effects: It constantly expands the practical limits for algorithm complexity, and it makes it increasingly feasible to trade in a bit of code overhead for improved modularity, tuneability, etc.
- There are physical modeling systems readily available, which may provide a starting point for the simulation.

3.2 Block diagram

The overall block diagram of the Room Simulator is shown in fig. 1. As often seen, the system is divided into two main paths: An early reflections synthesis system consisting of a so-called Early Pattern Generator (EPG) for each source and a common Direction Rendering Unit (DRU) that renders the early reflections through the chosen reproduction setup. And a Reverb system producing the late, diffuse part of the sound field. Note that - contrary to what is normally the case - there is no direct signal path. The dry source signals are merely 0th order reflections produced by their respective EPGs. In the following, a more detailed description of the individual blocks is given.

3.3 Early Pattern Generators

Each EPG takes one dry source input and produces a large set of early reflections, including the direct signal, sorted and processed in the following "dimensions"

- Level
- Delay
- Diffusion
- Color
- Direction

The Level and Delay dimensions are easily implemented with high precision, the other 3 dimensions are each quantized into a number of predefined steps, for instance 12 different directions. Normally, the direct signal will not be subjected to Diffusion or Color. The quantization and step definition of the Direction dimension must be the same for all sources, because it is implemented in the common Direction Rendering Unit. Physical modeling programs such as Odeon [1] may provide an initial setting of the EPG.

3.4 Direction Rendering Unit

The purpose of this unit is to render a number of inputs to an equal number of different, predefined subjective directions-of-arrival at the listening position via the chosen reproduction setup, typically a 5-channel speaker system. Thus, the DRU may be a simple, general panning matrix, a VBAP [2] system or an HRTF- or Ambisonics-based [3] system.

3.5 Reverb Feed Matrix

The reverb feed matrix determines each source's contribution to each Reverberator input channel. Besides gain and delay controls, some filtering may also be beneficial here.

3.6 Reverberator

To ensure maximum de-correlation between output channels, each has its own independent reverb "tail" generator. Controllable parameters include:

Reverberation time as a function of frequency $Tr(f)$

- Diffusion
- Modulation
- Smoothness

We take particular pride in the fact that our "tail" can achieve such smoothness in both time and frequency, and that modulation may be omitted entirely. This eliminates the risk of pitch distortion and even the slightest Doppler effect, which tends to destroy focus of the individual sources in a multichannel room simulator.

Again, an initial setting of $Tr(f)$ may be obtained from Odeon.

ROOM SIMULATION - FOR MULTICHANNEL FILM AND MUSIC

3.7 Speaker Control

This block is by default just a direct connection from input to output. But it may also be used to check the stereo- and mono compatibility of the final simulation result by applying a down-mixing to these formats. Also it provides delay- and gain compensation for non-uniform loudspeaker setups, which may also - as a rough approximation - be used the other way around to emulate non-uniform or misplaced setups and thus check the simulation's robustness to such imperfections.

4. CONCLUSION

The system described above is evidently a very open system under continual development. At the time of writing these words, our test system is running in real time on a multiprocessor SGI server with an 18-window graphical user interface providing interactive access to approximately 2000 low- and higher-level parameters. However, this is not the time or place to go into more details. When this paper is presented at the 107th AES Convention in about 4 months, we will have more real life experience with the system.

If integrated positioning is used with multi-source room simulation, our experiments have already shown how much there is to gain in terms of realism and working speed. But even with the less radical additive approach, virtual rooms may be rendered more convincingly with multi-source simulators.

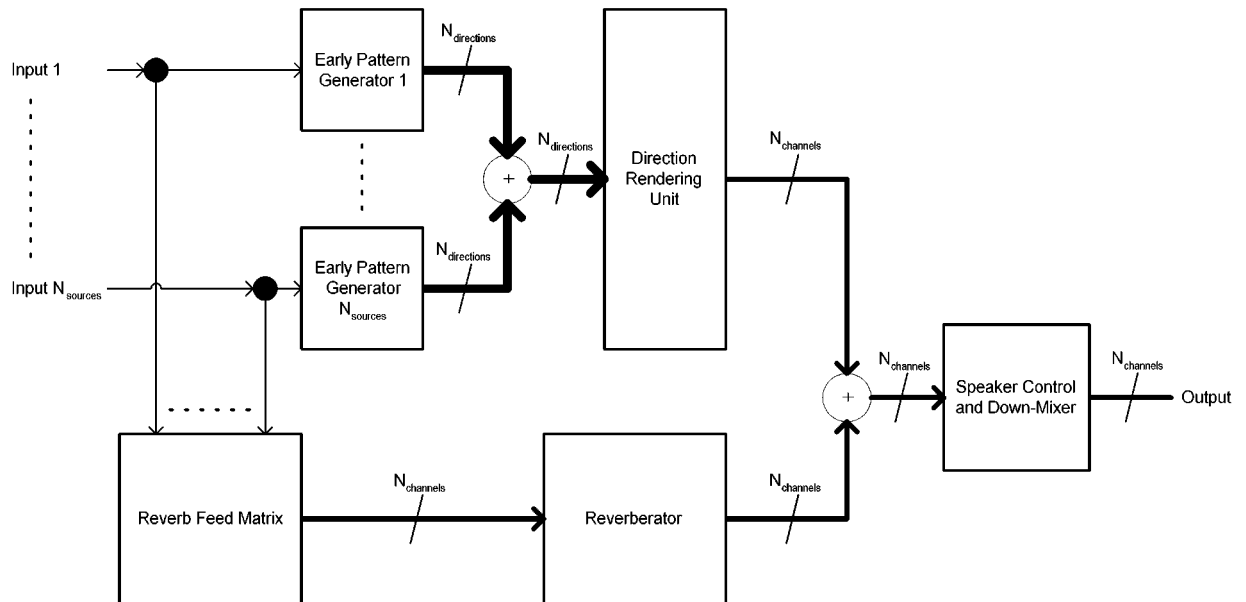
For applications where picture is added to the sound, the most stimulating source will be one, where audio and video are treated with equal attention to quality and detail. The new possibilities available from multi-source room processors may be exploited to generate a real quality improvement at the end listener, especially when his reproduction system is multichannel. More convincing sound generates more convincing picture.

REFERENCES

- [1] <http://www.dat.dtu.dk/~odeon/>
- [2] Ville Pulkki: "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", JAES Vol. 45, No. 6, pp. 456, 1997.
- [3] Jérôme Daniel, Jean-Bernard Rault & Jean-Dominique Polack: "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions", AES Preprint no. 4795, 1998.
- [4] Küpfmüller, Karl: "Nachrichtenverarbeitung im Menschen", University of Darmstadt, 1975.

ROOM SIMULATION - FOR MULTICHANNEL FILM AND MUSIC

Fig 1
Overall block diagram of Room Simulation Algorithm



SAMPLE RATE CONVERSION FILTERS - ADA 24/96

Sample Rate Conversion Filters

Introduction

The Delta-Sigma Converters of the ADA 24/96 operate at 6.144MHz (48/96kHz sample rate) or 5.6448MHz (44.1/88.2kHz) at the ends facing the analog world. But the processing, transmission and storage of high-quality audio material are often done at 48 or 44.1kHz. Conversion to/from this low sample rate involves a series of digital filters removing the signal components above one half the sampling frequency (the Nyquist frequency) at all points in the chain. The properties of the lowest filter stage converting to/from 44.1 or 48kHz has profound influence on sound quality, and getting down to 44.1kHz digital and back to analog without severe loss of "air", transparency and spatial definition is usually considered impossible. Furthermore, experience suggests that the optimum design of this critical filter is program material dependent. And of course there is always the matter of taste.

Therefore, when we designed the ADA 24/96, we did not just do a careful circuit design around a set of state-of-the-art 24-bit converter chips. We threw in the powerful Motorola 56303 DSP chip as well, enabling us to offer you a choice of different, carefully optimized filters for the super-critical conversion stage from 96 to 48kHz.

The 5 different filter types described below have been optimized individually for each of the two target sampling rates (44.1 and 48kHz), enabling us to take full advantage of the (comparatively) more relaxed design conditions at 48kHz. The optimization involved A/B comparison to a direct analog transmission of live musical instruments.

Linear phase - non linear phase

When aiming for perfect sound reproduction, the best one can hope for is a "straight wire": Nothing added, nothing removed and nothing changed. The output waveform is an exact replica of the input waveform. In technical terms this means

- Infinite signal-to-noise ratio and no distortion
- Linear phase response "from DC to light".
- Dead flat magnitude response "from DC to light"

How does digital audio transmission and processing score compared to this ideal?

- With state-of-the-art 24-bit converters and no-less-than 24-bit processing, the signal-to-noise ratio today is sufficiently close to "infinite" and with proper dithering, distortion is no longer considered a problem, even at low levels.
- Linear Phase is easily obtained: Perfect digital transmission is essentially just copying or storing and retrieving numbers, and if filtering is required, linear-phase FIR filters are the simplest kind.

- The Sampling Theorem puts a dramatic restraint on bandwidth: Any remaining signal components above the Nyquist frequency (only 22.05 kHz at 44.1 kHz sampling rate) turn into aliasing. Therefore the magnitude response has to drop sharply just above 20 kHz.

So far most designers of digital audio equipment have settled for 2 out of 3: Fought noise as best they could and kept the phase response linear. However, when the conditions change so dramatically from "Ideal" to "Having to give up on one goal out of three", there's no law saying that the other two goals don't move! And - according to our experience - the linear-phase goal does indeed move! Given the necessity of a sharp hi-cut filter, a perfectly linear phase response is not optimal!

With a bit of mathematical and psycho-acoustical reasoning, this is not surprising:

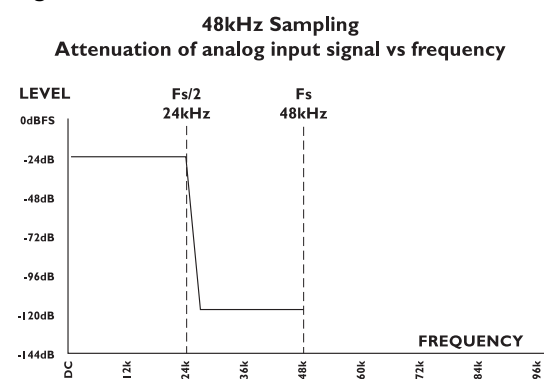
The sharp hi-cut filter is bound to add a lot of ringing to the system's impulse response. The phase response affects the distribution of the ringing in time: Linear phase implies a time-symmetric impulse response with equal amounts of ringing before and after the actual impulse. Human hearing is not time-symmetric, as anyone who has ever played a tape backwards will know. On the other hand, we are not insensitive to phase distortion either, so a compromise has to be found.

The Filters

Through weeks of repeated listening tests and phase, as well as magnitude response adjustments, we came up with the Natural and Vintage filters that use slightly non-linear phase responses.

The filters include some that are aliasing-free in the sense that they achieve full attenuation at the Nyquist frequency where aliasing starts to occur, as opposed to the cheaper half-band filters often used, that are only 6 dB down at the Nyquist please see fig. 1 and 2.

Fig. 1



SAMPLE RATE CONVERSION FILTERS - ADA 24/96

Fig. 2

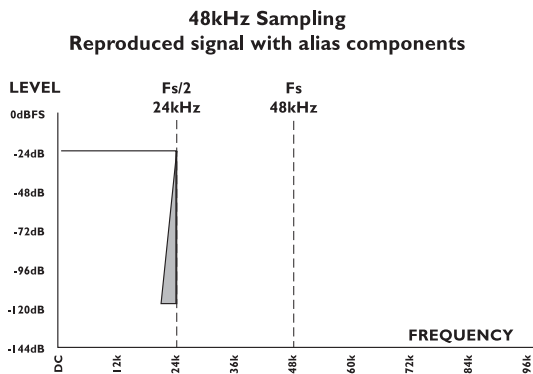


Fig. 3 illustrates how attenuation is performed just above FS/2 when operating at 48kHz sampling rate. As shown in fig. 4 alias components will be reflected around FS/2 and will disturb the high but still audible frequencies.

Fig. 3

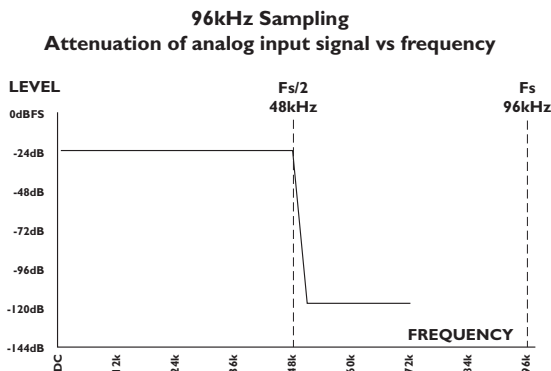
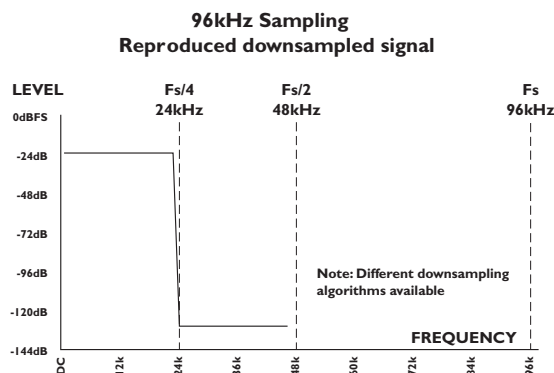


Fig. 4



At 96kHz FS/2 is 48kHz. Since the alias components are produced just below FS/2 they are in a non-audible area when 96kHz processing is possible. This is one of the reasons why 96kHz is preferable.

"Linear" filter

These filters are linear-phase and non-aliasing (the stop-band starts below the Nyquist frequency). The pass-band response is extremely smooth and non-equiripple, extending beyond 20 kHz. With the "Linear" filters you'll have a hard time discriminating between the sound of the conversion chain and direct analog, even at 44.1 kHz!

"Natural" filter

Based on the "Linear" filter class, but with a carefully adjusted non-linear phase response, these filters obtain an almost "better-than-live" reproduction of space while retaining crystal-clear imaging and absolute tonal neutrality. The "Natural" filters too are non-aliasing.

"Vintage" filter

Based on the "Natural" filters, here we've added a bit of warmth and roundness to the treble by introducing a smoother "tube like" roll-off. This filter would be an exceptionally good choice when mastering material that seems too hard in the high-end frequencies. These filters too are non-aliasing and non-linear phase.

"Bright" filter

These filters are something entirely different: Ultra-short impulse response, linear phase and quite a bit of deliberate aliasing produces a "digital" and slightly aggressive sound adding plenty of top-end life to e.g. Rock and Techno recordings, or giving you the feeling of air you need when you are mastering a somewhat dark sounding source material.

"Standard" filter (STD)

This filter emulates the response of typical mid-end converters: Equiripple half-band filters that are precisely 6dB down at the Nyquist frequency.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

By Christian G. Frandsen

Introduction

This document will discuss the clock, synchronization and interface design of TC System 6000 and deal with several of the factors that must be considered when using a digital studio. We will go through different aspects in this area e.g.

- What is jitter, what causes it and how is it removed on System 6000.
- Measurements comparing a conventional clock design to that of System 6000.
- Synchronization in digital audio studios in general.

The article is addressed to the users of System 6000 and other high-end digital studio equipment. It is meant to be a guide to an optimized digital studio setup. It is our experience that many of the problems in a digital setup can be solved by knowledge alone, so hopefully the document will help to answer some of the questions and to clear some of the typical misunderstandings relating to timing and clock generation.

Quality of AD and DA conversion

When designing System 6000 and its ADA24/94 analog conversion card we looked carefully at a lot of parameters in order to reach the highest possible overall performance...

- Frequency response
- Distortion
- Noise
- Crosstalk
- Common mode signals
- Alias filtering
- Jitter
- Analog domain pre/post scaling of converters
- Analog outputs optimized for balanced as well as unbalanced operation

After AD conversion, the analog signal is represented only as a level and timing component. Lack of precision in either area is detrimental to the process, so talking about a converter being "20" or even "24" bit provides very little information if the timing source is not equally well quantified.

What is jitter?

Jitter is the variation in time of a clock signal from the ideal.

The amount and rate of the variation are the important parameters. Research made through the last years suggests that variations faster than 500 Hz are the most audible [1]. If you are not familiar with the jitter terminology it is important to notice that the sampling frequency (F_s) could be 48.000 kHz (or maybe 256 times F_s equals 12.288 MHz) and that it is the variation of this clock that has to be faster than 500Hz.

Jitter performance is measured in seconds and a typical value could be in the area of 100 ps peak to 50 ns peak (jitter is also often measured in peak to peak or RMS). The typical jitter frequency spectrum tends to be low frequency weighted.

Clock wander (clock frequency change over long time) is also a kind of jitter but is so slow that it probably hasn't got a direct influence on the sound. Clock wander is typically due to temperature change and aging of the clock crystals in a device.

Sampling jitter.

Jitter only affects the sound quality when occurring in relation with converters (Digital to Analog Converter DAC, ADC or an Asynchronous Sample Rate Converter ASRC).

If the clock controlling an analog to digital conversion is subject to significant jitter, the signal gets converted at unknown points in time. Then, when playing back the digital signal even by using a rock steady clock, the signal is no longer identical to the first analog signal. You could say that the time variation on the first clock has been modulated onto the signal and therefore the signal is now distorted.

Sampling jitter is a potential problem every time the signal changes domain like in a DAC, ADC or an ASRC and as a rule of thumb the internal clock of a digital audio device is better jitter wise than when the device is clocked from a digital input.

This is why quality conscious mastering studios put their ADC in internal clock mode when feeding in analog material, but switch to DA clocking when all material is loaded. The converter performing the most critical task (capturing or monitoring) is assigned the master clock role.

System 6000 is one of the first pro audio devices seriously tackling the dilemma of which clock to use as master - AD or DA. First of all, AD and DA share the same very local, high quality clock. Secondly, even if an external clock is used, its jitter is so attenuated, that it has no effect on the conversion.

Interface jitter.

Interface jitter does not directly influence the sound. In extreme cases, however, interface jitter can be so severe that the transmission breaks down. At this point, of course, audio quality will indeed be affected.

Interface jitter is the variation in time of the electrical signal (carrying the digitized audio) being transferred between two devices. The main issue in order for the interface to work is that the receiving device is able to follow the timing variations well enough to receive the correct data.

Often, a device receiving a digital signal is slave to the incoming signal. This means that the device extracts the clock from the incoming digital signal in order to be

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

synchronized to the transmitting device. In conventional circuit designs the extracted clock is typically used directly for the converters. This means that the jitter on the digital interface is fed nearly unaltered to the converters and therefore manifests as sampling jitter.

Causes of Jitter

There are several ways that jitter find it's way into a digital studio setup.

Noise induced on cables

A digital receiver typically detects a rising or a falling edge on a digital signal at approx. halfway level. Due to finite rise/fall times on the signal, noise then can disturb the detection so the receiver detects the edges imprecisely. Therefore, both noise and other interference imposed on the signal line and the slope of the signal edge has influence on the precision of the receiver. Some digital formats are unbalanced (coaxial-S/PDIF) and others are balanced (AES/EBU). The balanced signals are more immune to induced noise due to the noise being treated as a common mode signal, which is suppressed to some extent in the receiving device.

Data jitter (or program jitter)

Data jitter is caused by high frequency loss in cables and the nature of some digital formats (e.g. AES/EBU and S/PDIF). Because the electrical data patterns are irregular and changes all the time, a specific edge in the signal can arrive at different times depending on the data pattern prior to the edge. If there weren't any high frequency loss in the cable this wouldn't be the case.

By using cables with incorrect impedance there will be a non-ideal transmission line that potentially contributes to the sloped edges and high frequency loss, and therefore indirectly generates jitter. In this respect, unbalanced formats (like S/PDIF) is often superior to its balanced counterparts.

Optical formats

Some digital formats are optical (Toslink-S/PDIF and ADAT) and they have a reputation of being bad formats jitter wise. One of the reasons for this is that the most common circuits used for converting between electrical and optical signal are better at making a rising than a falling edge. This causes asymmetries in the transferred digital signal, which also contribute to data jitter.

Internal design

Every oscillator or PLL (phase locked loop) will be uncertain about the time to some extent. (A PLL is typically used to multiply frequencies or to filter a clock signal in order to reduce jitter - jitter rejection).

This kind of basic uncertainty is called intrinsic jitter and for cheap designs it can be quite severe (there are examples of up to 300ns peak where the limit for the AES format is 4 ns peak @ 48 kHz Fs, BW: 700 to 100 kHz [3]).

Devices that feature jitter rejection will typically be well designed regarding intrinsic jitter as well.

Jitter accumulation

Jitter accumulation can happen in a chain of devices due to intrinsic jitter PLUS jitter gain (see The clock design on System 6000) in devices PLUS cable introduced jitter. Every device and cable will add a bit of jitter and in the end the jitter amount can get disturbing. There are ways to overcome this potential problem (see Synchronization).

How to detect jitter in the system

How to detect sampling jitter

The higher rise/fall time of the program signal the more sensitive it is to sample clock jitter and therefore one of the best ways to analyze a converter performance jitter wise is to apply a full-scale high frequency sine to the converter. The sample clock jitter will then be modulated onto the audio signal and it is now possible to measure the jitter frequency spectrum by performing an FFT on the converted audio signal.

In Figure 1 the DAC has been converting a 12 kHz sine. The two curves illustrate the difference with and without 5 kHz 3.5ns RMS jitter being applied on the digital interface. In this example the device has no rejection of the jitter appearing on the digital interface so it is nearly directly transferred to the converter where it is modulated into the audio signal. A conventional design like this is discussed in more details later.

The two jitter spikes are at the frequencies 12 kHz +/- 5kHz and the level approx. -80 dB corresponds to the 5kHz 3.5ns RMS jitter being applied. Sampling jitter (for jitter frequencies below $F_s/2$) will appear symmetrically around the sine being converted. Jitter frequencies above $F_s/2$ will be modulated into the audio signal in a more complex way. Another thing to notice on Figure 1 is that on the curve with 5kHz jitter there is also a tendency of some low frequency < 2 kHz (note 12 kHz +/- 2 kHz) noise jitter. This might be due to noise in the circuit generating the 5 kHz jitter.

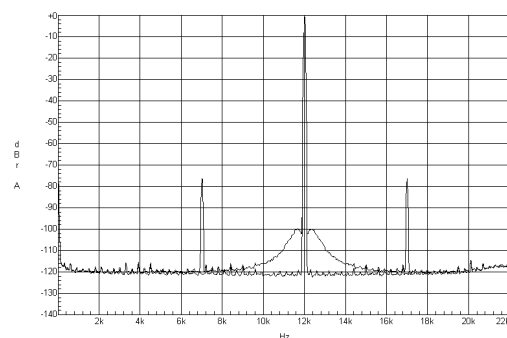


Figure 1 FFT on a DAC. Measurement made on Audio Precision System 2 Cascade using a 2k point FFT with 256 times average, and equiripple window. Two curves: Upper with a 12kHz spike and two 12kHz +/- 5kHz spikes. Lower only with a 12kHz spike.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

How to detect interface jitter.

The typical way to investigate interface jitter is by measuring the clock variations directly on the digital signal. There are devices made specifically for interface testing. The way they usually work is by applying a PLL circuit like the ones used for jitter rejection (see The clock design on System 6000) and then measure the amount that has been stopped by the PLL.

This circuit will act like a low pas filter towards the jitter variations and therefore it is the high frequencies that are stopped by the PLL. This way you will measure the jitter noise with a band limited filter that typically will have settings like 50 to 100 kHz, 700 to 100 kHz and 1200 to 100kHz.

By applying different filters and therefore getting different results you will have an idea of what jitter amount your system is operating at and what jitter frequencies that might be the potential problem.

Examples of interface jitter amounts

A test setup was made with 4 different devices connected using AES/EBU. Short impedance matched cables were used so there was only an insignificant amount of jitter coming from this potential source.



Figure 2. Test setup for interface jitter measurement.

Devices 1 to 3 are conventional designs with no jitter rejection below 10 kHz.

Device 4 is System 6000 (with jitter rejection).

Band width	After device 1	After device 2	After device 3	After device 4
50 to 100 kHz	1.0 ns peak	2.2 ns peak	2.5 ns peak	1.1 ns peak
700 to 100 kHz	0.9 ns peak	1.6 ns peak	1.9 ns peak	0.9 ns peak
1200 to 100 kHz	0.9 ns peak	1.5 ns peak	1.7 ns peak	0.9 ns peak

Table 1 Interface jitter measurements.

In Table 1 the results from the interface jitter measurements are shown. After device 1 the results reflect only intrinsic jitter of this machine. Notice the slight increase at 50Hz, which is common. After device 2 the jitter amount has increased which in this case was due to intrinsic jitter in device 2. The same thing was the case after device 3. After System 6000 (device 4) jitter level has dropped due to the jitter rejection in the system.

There is still an increased level at 50Hz where the System 6000 jitter rejection filter has it's corner frequency and this is why the jitter at 50Hz isn't as reduced as the level above 50Hz.

The clock design on System 6000

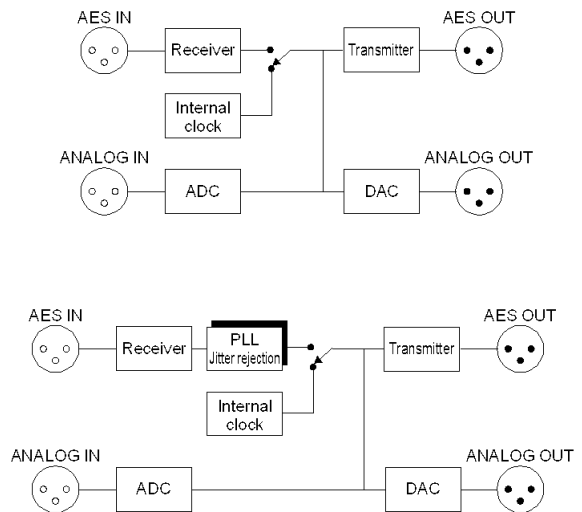


Figure 3 Clock circuit in a conventional design and in System 6000

Jitter rejection

The secret is the PLL that is used to remove jitter that might appear on the digital input and in a conventional design is transferred nearly directly to the converters. The PLL acts as a low pass filter and will reduce jitter noise of higher frequencies than the corner frequency of the filter. On System 6000 the corner frequency is as low as 50 Hz and at 1.4 kHz the noise will be reduced by at least 100 dB. Figure 4 shows -69dB at 500 Hz, which corresponds to filter suggestions, made from research in the last year's [1].

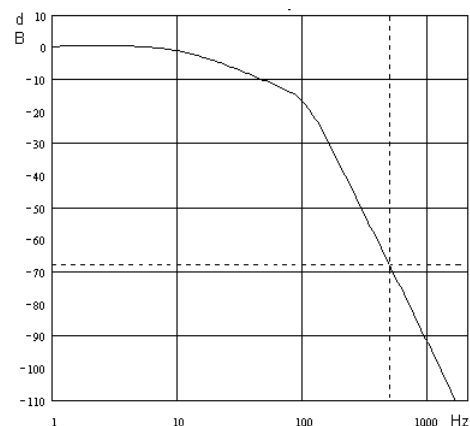


Figure 4 Jitter rejection filter (4th order filter)

It is very difficult to design a low pass filter with a steep slope without gain in the pass band. This type of gain can contribute to jitter accumulation in a chain of devices. The System 6000 gains as little as < 1dB at 2 Hz.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

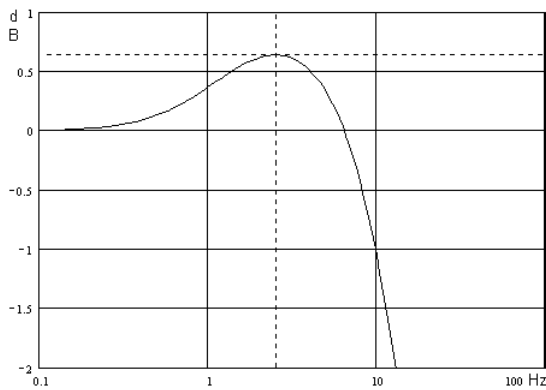


Figure 5 Zoomed picture of the jitter rejection filter.

Lock range

There are some devices on the market that feature jitter rejection to some extent but most of them uses a common technology (VCXO type of oscillator in the PLL). This technology is limited in the terms of lock range. This means that the jitter rejection only can be done at the known frequencies e.g. 96 kHz, 88.2 kHz, 48 kHz +/- a few hundred PPM. This is typically enough given the fact that most clock oscillators have a precision of less than +/- 50ppm but if you want to use other sample rates (e.g. vary speed or the broadcast related 44.056 kHz) you will have to do without the jitter rejection.

System 6000 uses a special technology that makes it able to lock to all sample rates from: 30 to 34 kHz, 42.5 to 45.6 kHz, 46.5 to 48.5 kHz, 85 to 91 kHz and 93 to 97 kHz. This means that every signal at sample rates in these ranges will be treated with the same jitter rejection filter. The performance is not only good in a narrow range around one or two sample rates.

Intrinsic jitter

The intrinsic sampling jitter on System 6000 has been optimized through a long period of trimming in order to make the perfect AD- and DA-conversion on the ADA24/96 card. This is also why the intrinsic interface jitter is as low as < 1 ns peak, BW: 700 to 100 kHz that makes the perfect starting point for the digital setup.

System 6000 technical specifications regarding the clock circuitry:

Jitter rejection filter (4 th order filter):	< -3dB @ 50 Hz < -69dB @ 500 Hz < -100dB @ 1.4 kHz
Jitter gain:	< 1 dB @ 2Hz
Jitter rejection at external sample rates:	30 to 34 kHz, 42.5 to 45.6 kHz, 46.5 to 48.5 kHz, 85 to 91 kHz, 93 to 97 kHz.
Internal sample rates:	96 kHz, 88.2 kHz, 48 kHz and 44.1kHz.
Internal clock precision:	+/- 30 PPM.
Intrinsic interface jitter:	< 1 ns peak, BW: 700 to 100 Hz.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

Measurements on System 6000

As mentioned before the higher frequency in the program signal and the higher level the more sensitive it is to jitter on the sampling clock. Therefore the measurements in this document uses a 20 kHz sine at -1 dBFS and all measurements are done at 48 kHz sample rate. The measurements are done on the DA converter on the ADA24/96 card and the test system is Audio Precision System 2 Cascade.

To show the effect of the jitter being modulated via the clock into the audio signal the measurements are FFT's of the analog signal on the ADA24/96 output.

Every full view picture is made with an FFT: 2k point, 256 times average, and equiripple window. Every frequency zoomed picture is made with an FFT: 16k point, 32 times average, and equiripple window.

System 6000 performance

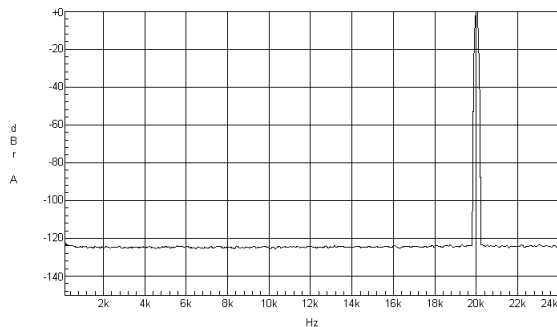


Figure 6 System 6000 DAC in master mode

The system is running in internal master mode thus there is no jitter being applied to the system. Notice the very flat noise floor with no spurious signals.

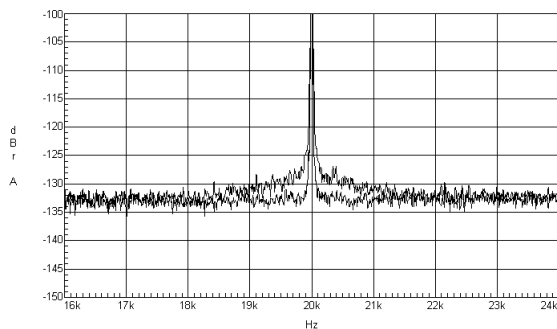


Figure 7 Zoomed version (both freq. and level) of Figure 6 plus System 6000 in slave mode.

On Figure 7 the system is running both internal master mode (lower) and external slave mode (upper) where the system is slave to the Cascade. There is no jitter being applied from the Cascade so what is shown here is approx. the difference in intrinsic jitter when the System 6000 is

slave to the digital input.

Notice that even with this much zoom the jitter on internal-mode is not visible.

The very little jitter that shows on slave-mode is low frequency weighted from 1kHz and down.

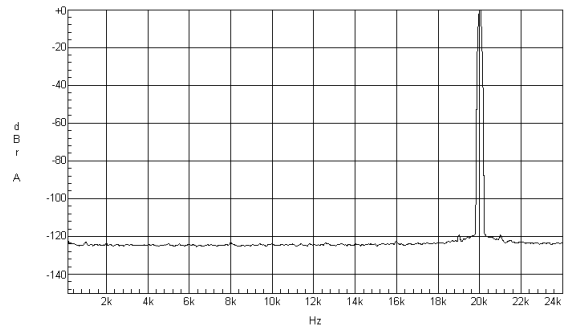


Figure 8 System 6000 DAC in slave mode, 1kHz jitter applied.

Figure 8 is a full view picture when System 6000 is slave and there is applied 1kHz, 1.3us peak sine jitter. This is a huge amount of jitter and actually beyond the tolerance level specified by AES3-1992 amendment 1-1997 [3] that all AES compatible devices should be able to lock to.

The performance of a conventional clock design compared to the performance of the System 6000

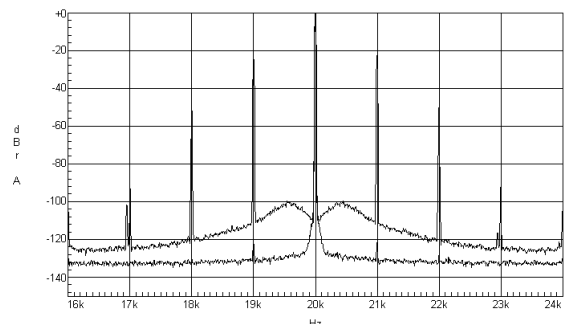


Figure 9 Zoomed (freq. only) version of Figure 8 (lower) plus the same level of jitter applied to a conventional design (upper).

On Figure 9 the 1 kHz, 1.3 us peak sine jitter is applied to System 6000 (lower) and a conventional design (upper) without jitter rejection. Notice that the jitter spikes on the conventional design reaches -22 dB with reference to the 20 kHz tone where the System 6000 has filtered out this jitter by approx. 100 dB down to -122 dB.

There are more spikes than 20 kHz +/- 1 kHz. There are also +/- 2, 3 and 4 kHz, which are harmonics on the sine jitter generator. The low frequency noise floor on the upper curve is probably noise in the jitter generator.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

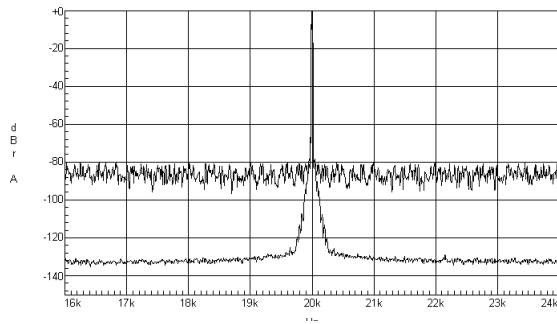


Figure 10 Zoomed (freq. only). System 6000 and conventional design in slave-mode. Wide band jitter applied.

On Figure 10 25 ns peak wide band jitter has been applied to both System 6000 (lower) and a conventional design (upper). The jitter level has formed a noise floor at approx. -80 dB with reference to the 20 kHz tone on the conventional design. The System 6000 curve reflects the jitter rejection filter curve up to approx. 300 Hz (20 kHz +/- 300 Hz on this picture). Beyond 300 Hz the System 6000 has reduced the jitter so much that it is hidden in the noise floor.

Synchronization

Synchronization: The digital signal, word clock or AES 11. There are several ways to obtain synchronization in a setup: Using a digital signal (carrying audio), a digital signal (not carrying audio) or a word clock.

- Digital signal (carrying audio).
This is the simplest way to obtain sync and it involves only the two (or more) audio devices that are connected.

Typically the transmitting device is the master and the receiving device is the slave.

As mentioned earlier there is jitter to take into account when selecting which device to be the master of timing. Typically the internal clock in a device is the cleanest and therefore when recording it is often the ADC that is the master (see Figure 11). To get the best DA conversion when mixing the DAC is often the master. But the DAC is a receiver of the audio signal and therefore it is necessary to send a synchronization signal from the DAC back to the rest of the system. This can be done with a digital signal that is or is not carrying audio and to which the system is slave.

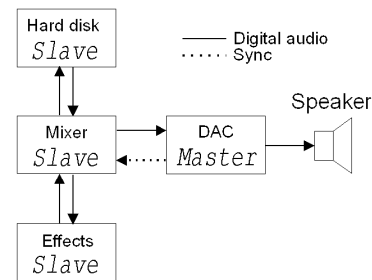
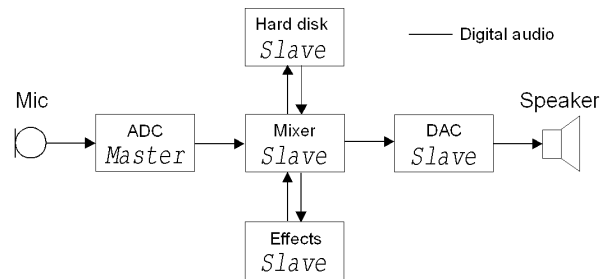


Figure 11 Digital studio setup with an ADC or DAC as master.

- Digital signal (not carrying audio) e.g. AES11 or word clock.

This way typically involves a device that makes synchronization signals for the audio processing devices in the setup.

This word clock generator is the master at all times and all the audio devices are slaves.

The word clock is a square wave signal (typically TTL level, 0 to 5V) with a frequency equal to the sample rate e.g. 48kHz. The impedance is 75 ohm and the connector is RCA phono or BNC.

The AES11 is an AES signal without audio and therefore the connector is XLR with a 110 ohm impedance.

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

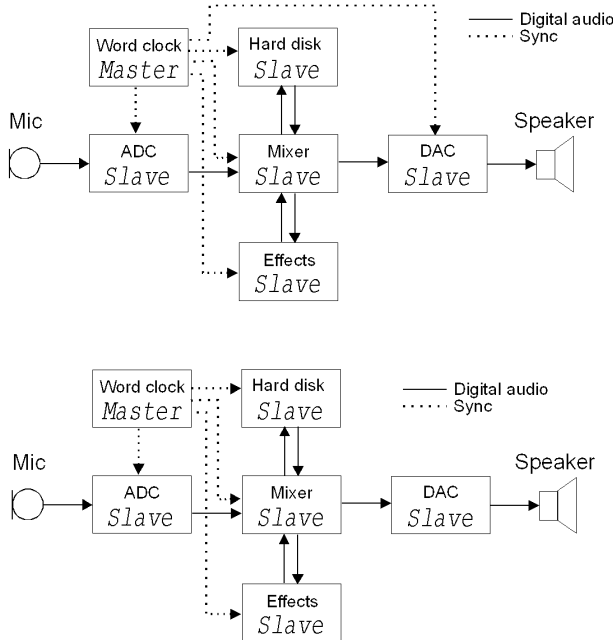


Figure 12 Same setup with a word clock generator. Example with DAC with and without word clock input.

Figure 12 shows the example setup with a word clock generator included. Some devices may not feature a word clock input and these devices will then slave to the digital signal carrying audio.

One of the advantages of using a word clock generator is that there is no constant switching between master and slave configurations. Just the normal patching signals around through the patch bay.

Another advantage is that potential jitter accumulation is reduced. The jitter sources from the clock master to where AD- or DA-conversion is done are reduced to the word clock generator intrinsic jitter, the word clock line to the converting device and the intrinsic jitter in this device. For setups with devices without word clock input the chain is a bit longer.

A disadvantage of using a setup with word clock might be that there will typically not be so much optimizing of the jitter e.g. the ADC being master when recording and the DAC being master when mixing. Even though the clock path is short (from the word clock generator to the converting device) having the converting device being the clock master could optimize the setup further jitter wise.

Nominal phase.

Digital mixers with a lot of digital inputs and other equipment e.g. surround devices have to be able to receive signals from several sources at the same time.

In Figure 13 the timing of different input and output signals is shown. AES specifies that a device must be able to

receive a signal with a phase of up to $\pm 25\%$ of the sample period away from the reference. This means that if the device is slave to input 1 the phase on the other input signals must fall within $\pm 25\%$ of the sample period away from the signal on input 1. If the phase of a signal (including two audio channels) is above the limit, the current sample in this signal can be interpreted as the previous or the next sample and this will add a delay to these specific two audio channels. A summing of the signals later in the setup (e.g. electrically or acoustically) will result in a potential audible phase error.

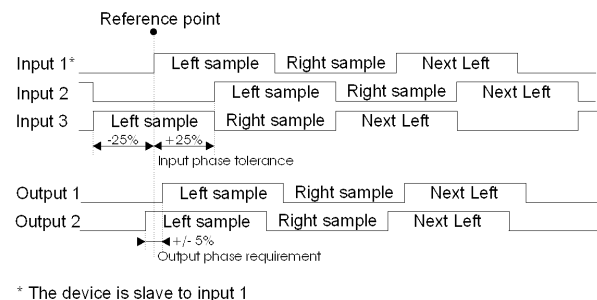


Figure 13 Nominal phase tolerances and requirements of AES signals.

It is difficult to make equipment transmitting a digital signal at exactly the same time as it receives a signal. Typically a device will have both a delay of a whole number of samples but also a sub sample delay. The delay of a whole number of samples is due to the routing of the signal in the device and the sub sample delay is due to the specific hardware in the device. AES specifies that the output must fall within $\pm 5\%$ of the sample period from the reference point (the incoming signal if the device is in slave mode). See Figure 13.

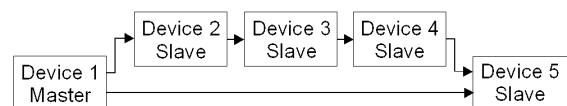


Figure 14 Example setup. Device 5 is slave to the signal from device 1.

Consider a setup like Figure 14 where device 2 to 4 adds a sub sample delay of 10% of the sample period. There will now be a difference between the two signals going into device 5 of 30% of a sample period. Now device 5 might interpret the two inputs wrong the way mentioned above. In a whole number of samples this might not be a problem because device 2 to 4 also has perhaps 10 samples delay pr. device equals a total delay of 30. So the resulting 1 sample extra delay in this chain due to several sub sample delays might not matter.

Another problem might be that perhaps the jitter level after device 4 is quite high. This jitter will perhaps make the signal from device 4 continuously cross the point in device 5 where the signal is interpreted as the current sample or

CLOCK AND SYNCHRONIZATION IN SYSTEM 6000

the next. This will make continuous slip samples (perhaps with audible clicks) on the input on device 5 receiving signal from device 4.

A way to work around this could be to have device 3 slaving directly to device 1 with an extra synchronization signal. This way the 10% sub sample delay in device 2 will be eliminated. A setup using word clock generator will typically not suffer from this kind of problem due to every device being slave to the same reference. Therefore there will not be any chain of devices that potentially could accumulate phase errors.

ADAT interfaces are typically more sensitive to sub sample delays or phase offsets than AES interfaces.

System 6000 technical specifications regarding in and output phase:

Digital Output Phase:	< 3 % of sample period
Input Variation Before Sample Slip:	+27 % / -73 % of sample period

Literature:

[1] Jitter: Specification and Assessment in Digital Audio Equipment by Julian Dunn.
Presented at AES 93rd Convention, October 1992
Available at www.nanophon.com

[2] Audio Precision, Technotes by Julian Dunn. Technote 23, 24 and 25.
Available at www.ap.com

[3] Standard: AES3-1992 amendment 1-1997.
Available at www.aes.org

Other literature on the subject:
"Everything you always wanted to know about jitter but where afraid to ask" by Bob Katz.
Available at www.digido.com

Measurement systems manufactured by:
Audio Precision www.ap.com
Prism Sound www.prismsound.com