Audio Engineering in a Nutshell

Audio Engineering in a Nutshell - Or: The next two hours of your life

- Basics: the nature of Sound and its representation
    - Sound and hearing
    - Analog signals
    - Digital signals
- Basic digital operations
- Controlling timing: delays
- Controlling spectrum: equalization
- Controlling dynamics: compressors/expanders
- Space and spatialization

Steve Harris <S.W.Harris@ecs.soton.ac.uk>

Jörn Nettingsmeier <nettings@folkwang-hochschule.de>

A word of warning

This talk is absolutely packed. We will flood you with information.

For two hours straight.

Except there's the lunch break. :-D

And there will even be Maths.

Three things will help you survive:

- We'll start at the very beginning.
- You will be able to ask questions after each section, and possibly get answers.
- We'll explain everything twice, first from the analog point of view (Jörn), then from the digital (Steve).

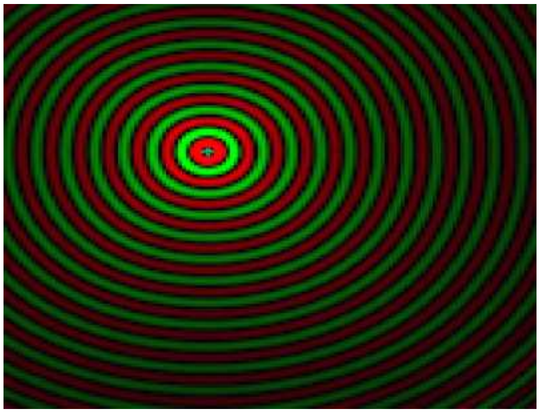
Note: Another general-purpose q & a session is planned at the end, so if your question does not relate to the current topic, please hold it until later.

Basics: The nature of sound and its representation

Sound and hearing

Sound waves

Sound is ***pressure waves*** travelling through a ***medium***.



Sound waves consist of alternating high and low pressure zones.

These zones oscillate in the direction of travel (a ***longitudinal wave***)

Sound waves travel through air at approx. 340 m/s (v_0, depending on temperature and humidity.

The oscillating speed of sound is called the ***frequency f***, measured in cycles per second or ***Hertz [Hz]***.

The combined length of one high- and one low-pressure zone is called the ***wavelength*** .

We find:

$$= v_0 / f$$

Hearing

We detect sound with our ear-drums. Like a microphone diaphragm, they move along with the pressure waves.

The softest sound humans can hear corresponds to an air pressure change of only 0.00002 Pa (Pascal). (The normal air pressure on earth is 102,300 Pa.)

The loudest sound we can endure has a pressure change of 200 Pa.

Our hearing range covers **10 orders of magnitude** of intensity!

The audible frequency range reaches from from 20 Hz to 20 kHz. The upper bound moves down with age.

We perceive frequency as ***pitch***, in a ***logarithmic*** fashion: pitch intervals are ***ratios*** of frequencies, not ***differences***.

The range from a given frequency to twice that frequency is called an ***octave***.

Signal levels, or: The dreaded Decibel...

To compute a ***sound pressure level L***, a given pressure p is first normalized to a ***reference***: our hearing threshold p_0 (0,00002 Pa).

Because of the huge range of pressures we can perceive, the value is then log'ed to make it more manageable:

$$L_p = 20 \log p / p_0$$

The unit of L_p is dB SPL, or "Decibel Sound Pressure Level".

For signals, we choose a ***reference voltage*** u_0 of either 1V or 0.775V:

$$L_u = 20 \log u / u_0$$

The unit of L_u is dBV (for 1V) or dBu (for 0.775V).

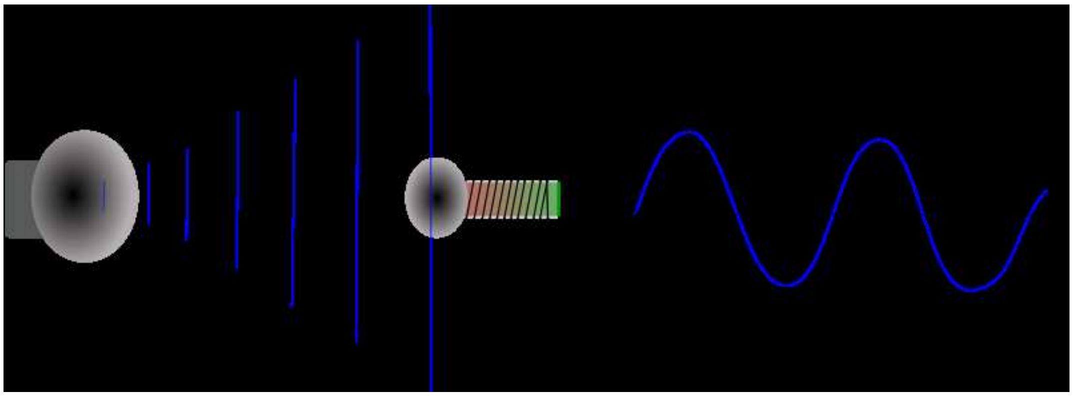
With dB's, a change of +6dB indicates roughly twice the level, -6dB is about half the level.

Note that the Decibel is not an absolute measure, but a **ratio**, and that it is meaningless without a ***reference***!

Analog Signals

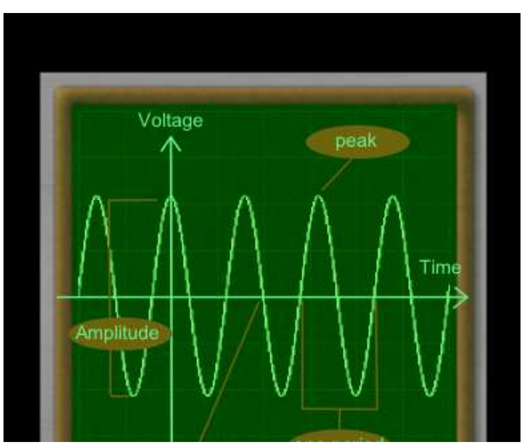
From wave to signal

A microphone diaphragm is moved back and forth by the pressure changes.
Attached to it is a coil within a magnetic field.
When it moves, an electric signal is ***induced*** in the coil.



The voltage changes with the speed and direction of the diaphragm movement.
Ideally, the signal is a perfect and complete representation of the sound.

Audio Signals



Audio signals can be visualized with an oscilloscope.

It displays ***voltage U*** vs. ***time t***.

This is called the ***time domain*** view.

The strength of the signal at any one time is called ***amplitude*** and roughly corresponds to ***loudness***.

Most audio signals are periodic - they follow a repeating pattern.

One ***period*** is the section between three adjacent ***zero crossings***.

The signal's ***frequency f*** (in ***Hertz***) is the number of such periods per second. It corresponds to ***pitch***.

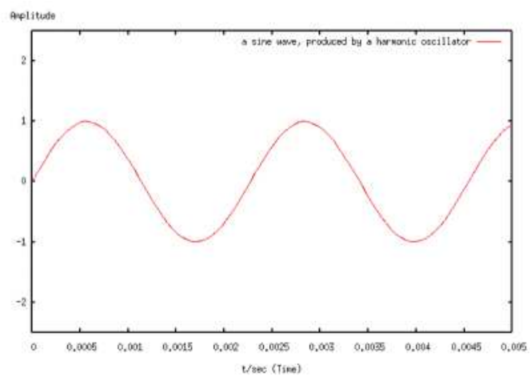
The points of maximum amplitude are called ***peaks***.

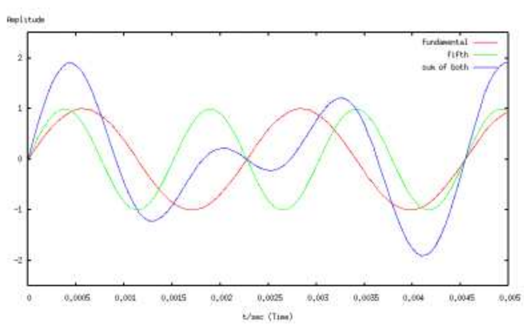
Sine waves (1)

The most fundamental sound phenomenon is the ***harmonic oscillation***.

It produces just one single pitch. [sine@440Hz]

In the time-domain view, it looks like a ***sine*** function.





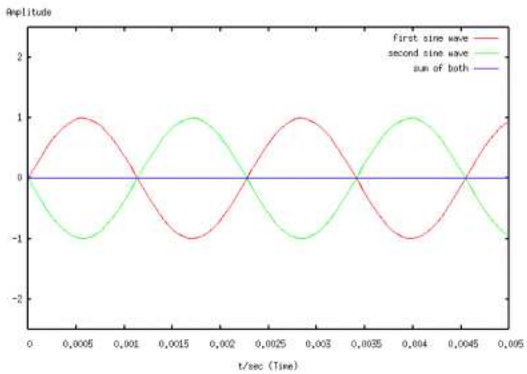
Sine waves can be mixed into complex sounds.

The amplitudes of the components at any one time add up to the result.

[sine@659.255Hz], [both]

Sine waves (2)

Identical waves can amplify each other:



Or they can cancel each other out, depending on their relative position:

This relative position is called the ***phase*** .

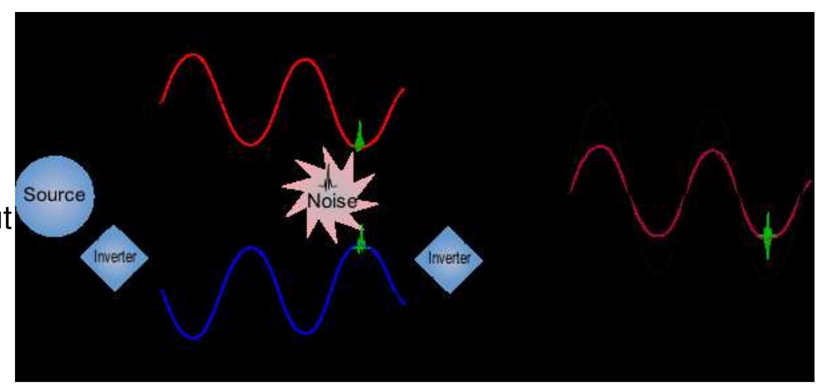
Detour: balanced connections

Professional ***balanced*** audio cables have three wires.

Pin 1 (screen) is the shielding foil or braid and is grounded.

Pin 2 (hot) carries the audio signal.

Pin 3 (cold) carries the inverted audio signal (180° out of phase).



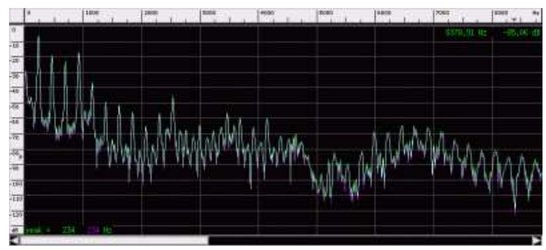
When electro-magnetic fields interfere with the cable, both hot and cold wires pick them up the same way.

At the input stage, the cold signal is inverted again and added to the hot.

Since they are back in phase, they amplify each other.

But the two halves of the noise are out-of-phase now - when added, they cancel out.

Sine waves (3): analysis



Any periodic sound is composed of sine waves:

The ***fundamental frequency*** determines the ***pitch***. The ***overtones*** or ***harmonics*** affect the ***timbre*** ("sound").

The sine components of a complex sound can be found with a ***spectrum analyser***.

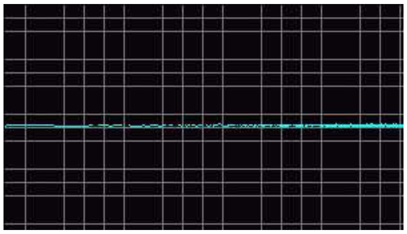
A ***spectrogram*** plots intensity vs. frequency.

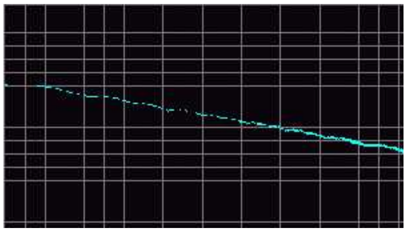
This is called the ***frequency domain view***, as opposed to the ***time domain view*** of the oscilloscope.

Sine waves (3): analysis (cont'd)

Noise has a ***non-periodic*** waveform.

It is not composed of a few distinct sine waves, but of an infinite number of arbitrarily close ones. The noise spectrum is continous, without any discernible overtones.





[***White noise***] contains all frequencies with equal loudness, i.e. the octave from 10 kHz to 20kHz contains as much energy as all the other octaves below.

Since humans perceive pitch logarithmically, white noise sounds very bright with exaggerated treble.

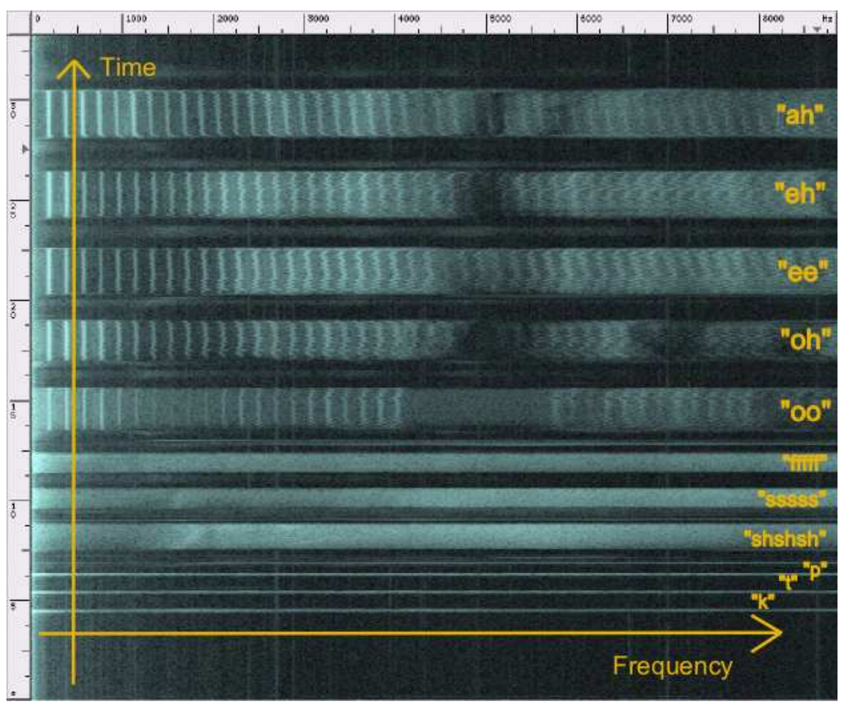
To accomodate for this, low-pass-filtered white noise is used for acoustic measurements. It is called [***pink noise***] and contains a constant amount of energy per octave.

Sine waves (3): analysis (cont'd)

A ***sonogram*** displays intensity (color) vs. time (y-axis) vs. frequency (x-axis). When running, it scrolls up as time passes.

Here are some speech sounds: sung vowels at the same pitch, sibilants and plosives.

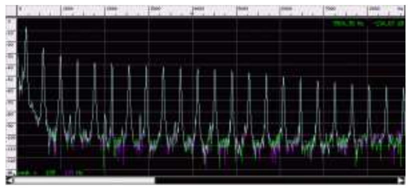
Note the identical fundamental and distinct harmonics of the vowels, and the noisy spectrum of the consonants.

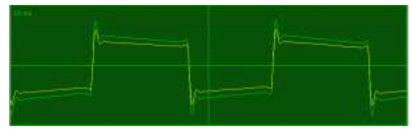


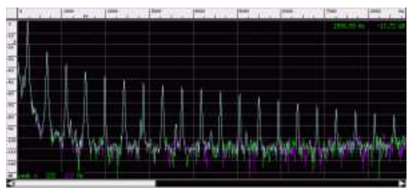
Sine waves (4): synthesis

New sounds can be generated by adding sine waves of different ***frequency***, ***amplitude*** and ***phase***. This is called ***additive synthesis***.

[***square wave***], only odd harmonics, slow rolloff

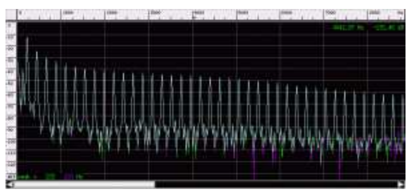






[***triangular wave***], only odd harmonics, fast rolloff

[***sawtooth wave***], all harmonics (bowed instruments)



[Fourier approximation of a square wave], [Fourier approximation of a triangular wave]

Problems with analog signals: noise and clipping

There are many sources of noise in analog signal processing:

- All electronic components produce ***thermal noise*** through the jiggling of their atoms (***Brownian motion***).
- Cables pick up noise via electro-magnetic interference (induction or capacitive coupling).
- Analog media introduce additional noise (***tape hiss*** or ***vinyl crackle***).

Analog signal stages will ***clip*** the signal if the amplitude is too high, producing ***distortion***.

A clipping stage can not follow the input signal, and flattens out the peaks of the wave.

The result now resembles a square wave.

Clipping generates new harmonics, which produce the typical distorted sound.

Both noise and clipping **cannot be undone**. Noise reduction and peak reconstruction are lossy processes.

Problems with analog signals: dynamic range

Audio systems are bounded between the ***noise floor*** and the clipping level.

The usable room between those boundaries is the ***dynamic range*** or ***signal-to-noise ratio (S/N)*** of a system.

The human ear has a dynamic range of about 120dB from hearing to pain threshold.

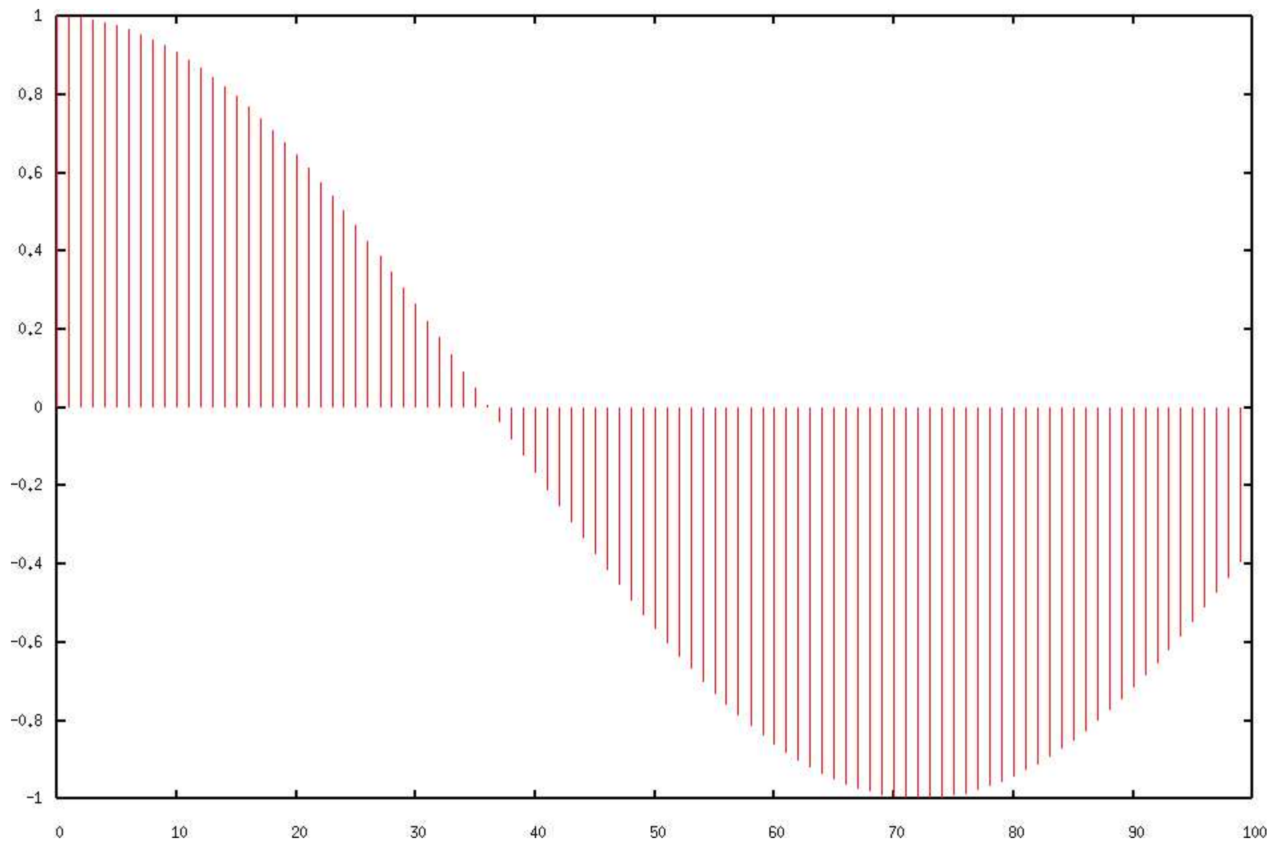
Studio-grade analog tape recorders achieve 70dB S/N, consumer machines are a lot worse.

Therefore, care must be taken to always fully exploit the available range. Analog signals should always be kept as hot as possible without clipping.

Questions?

Digital Signals

Continous v’s. discrete signals

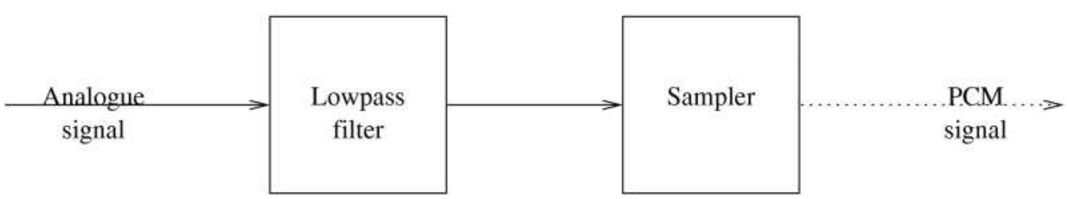


Sampling (AD Conversion)

Turning the analogue signal representation into digits

Most commonly used encoding in software is ***Pulse Codule Modulation*** (PCM)

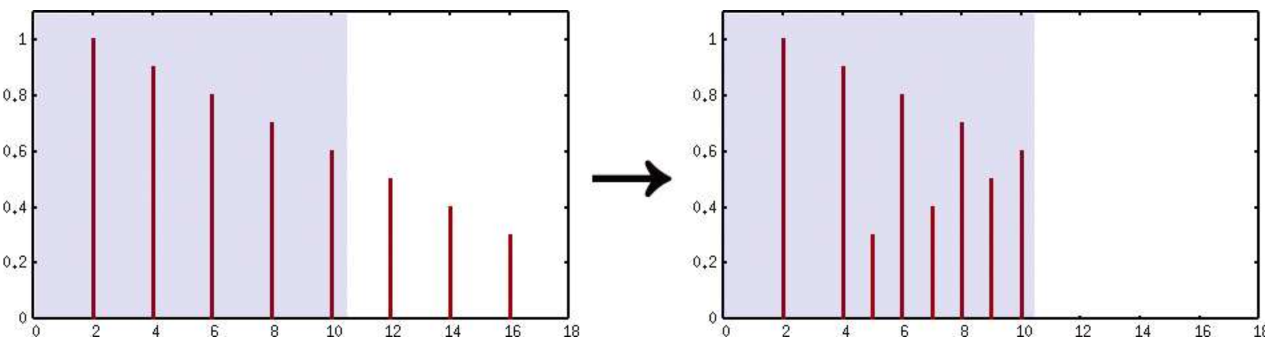
- Represented as impulses *n* times per second - a ***pulse train***
- *n* is the sample rate
- Amplitude of the impulse is measured by passing the signal through a lowpass filter



Nyquist criterion and aliasing

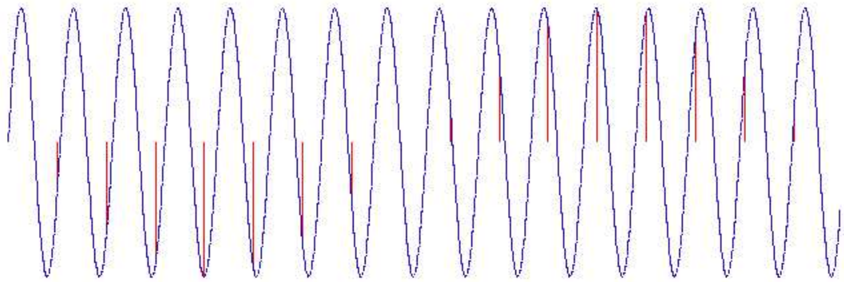
Nyquist criterion states that in order to be correctly sampled, all components of the signal must at less than half the sample rate

Any compenents at a frequency at or over the nyquist frequency will appear as "aliased" components when converted to PCM

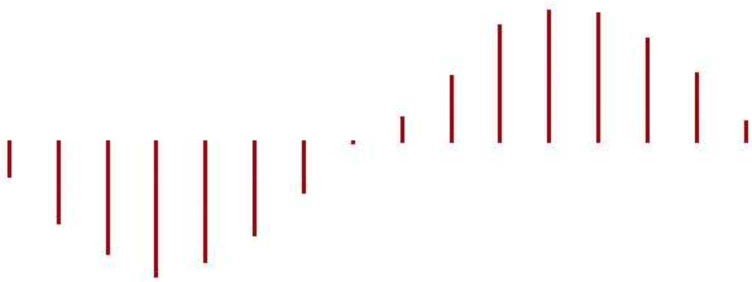


Trying to PCM endode a signal at 21kHz that contains components above 10.5kHz

Why does aliasing do that?



Sinewave at about 0.9 x the sampling frequency (1.8 x nyquist frequnecy)

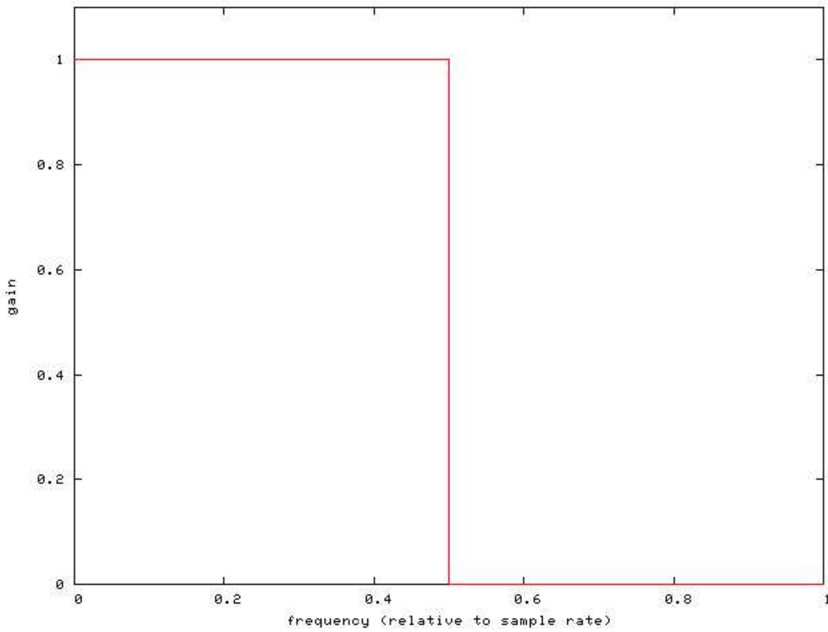


Samples shown without wave

Aliasing results in a sinewave appearing at a lower frequency

Anti-aliasing filters

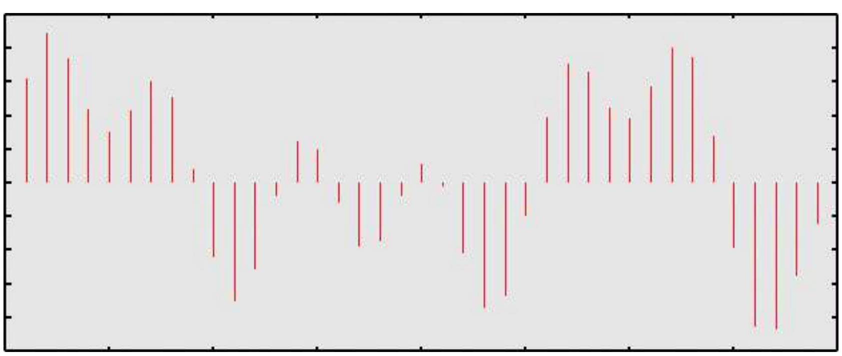
The best thing to prevent aliasing would be a bricakwall lowpass filter:

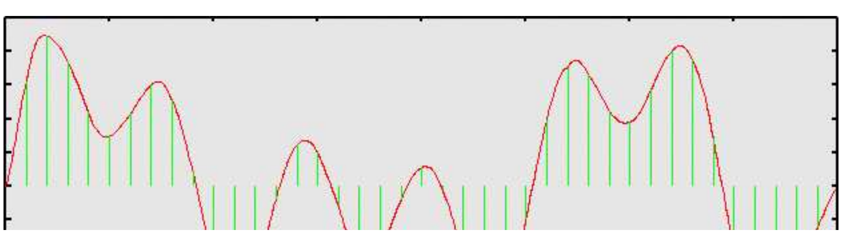


In practice you can’t get a response that good, but its the ideal

Reconstrunction (DA Conversion)

In theory the PCM data can be turned back into an anlogue signal by passing the impulse train through a lowpass filter





The ideal filter would be the same brickwall filter we’d like for anti aliasing.

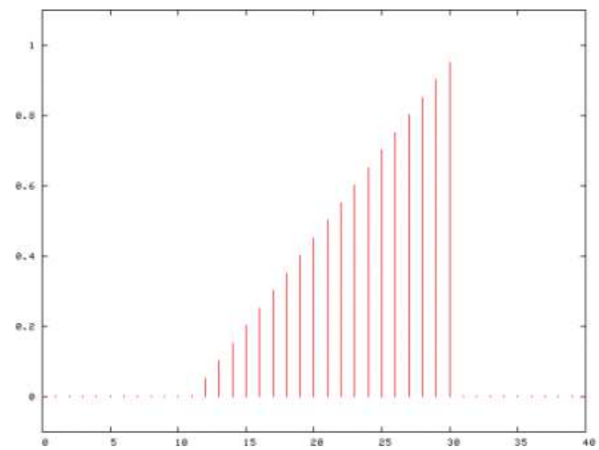
The reconstruction here was done using a filter thats very close to that

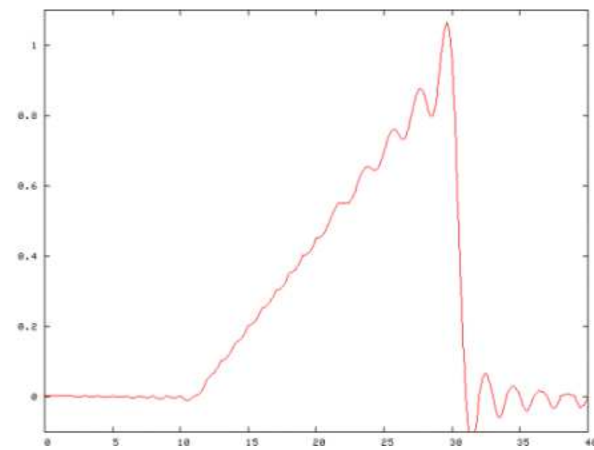
Problems: aliasing

Even though the analogue signal will have been antialiased on the way in it is still possible for aliasing to be created by processing

This can happen if some processing causes the frequency compenents to be shifted below 0 or obove the Nyquist frequency

Or if the PCM data is treated as if it were continuous in synthesis:

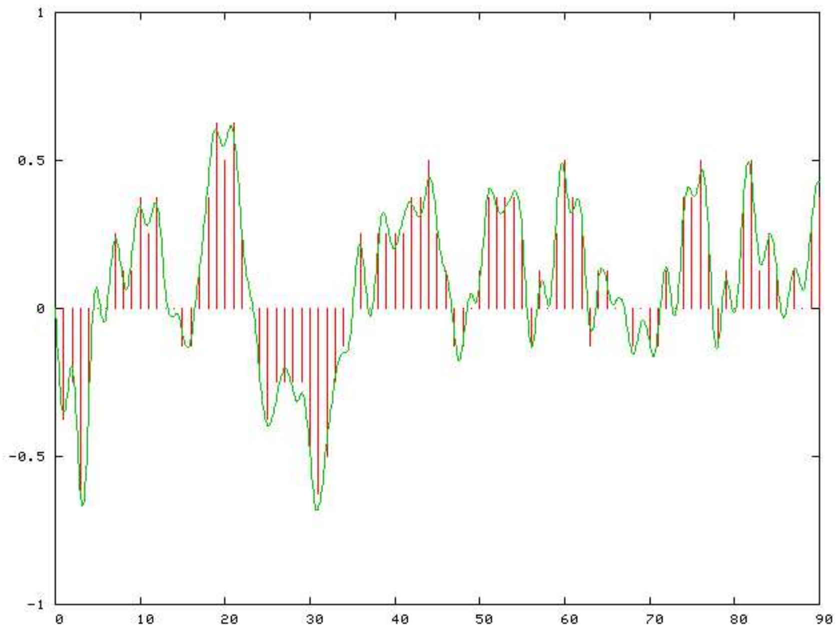




If the signal is built up from sinewaves, all below the Nyquist freuency then the signal will look a bit odd in PCM form, but will reconstruct to something that sounds like a saw wave

Quantisation

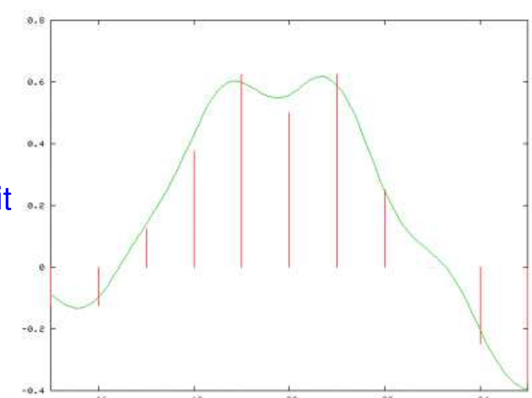
Quatisation is the approxiation of the level of the analogue signal to the most appropraite digital representation, eg. a 4 bit (16 level) sample of an arbitrary signal is quantised as:

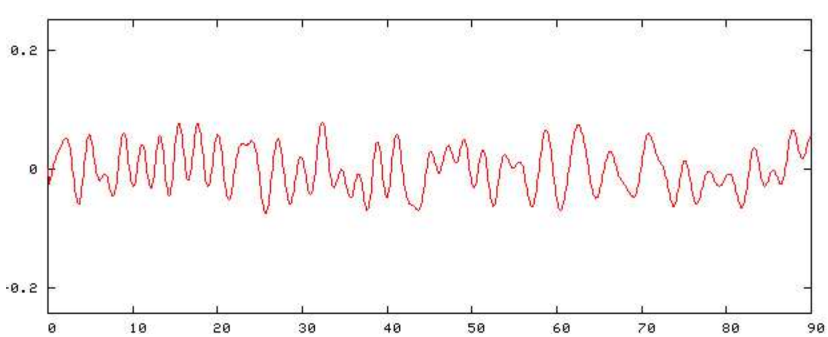


Problem: quantisation noise

The quantised signal does not match the original curve exactly, some points are over, and some are under

The differenece between the source signal and the quantised one will be percieved as noise on top of the wanted signal: [source sine wave], [4 bit sine wave], [quantisation noise]

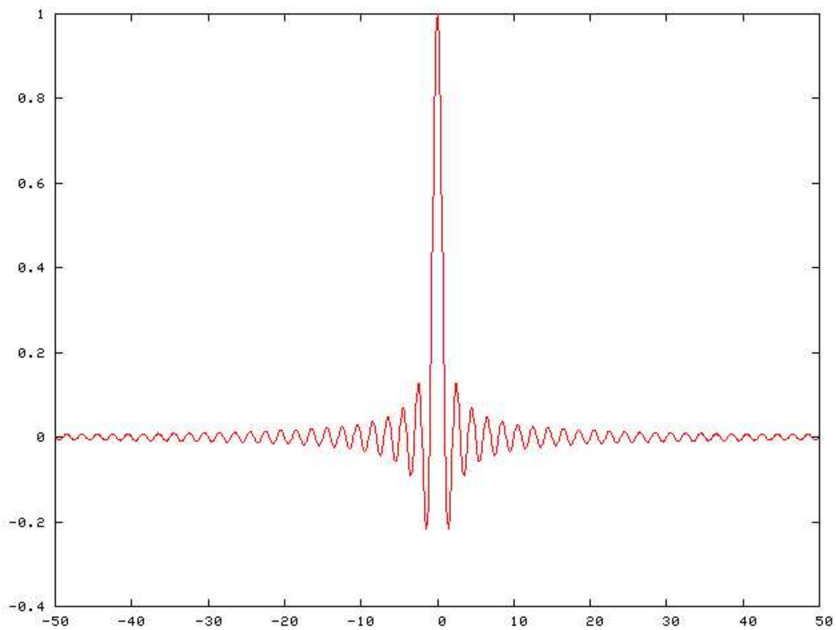




Interpolation

To reconstruct a discrete signal we can lowpass filter it with a brickwall lowpass filter, so it makes sense we can do the same to interpolate

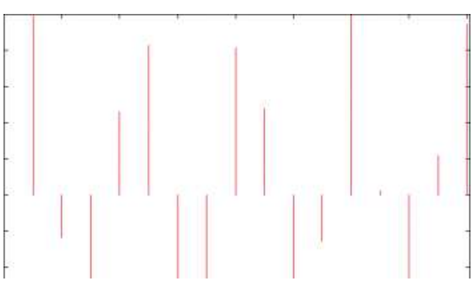
We can get a brickwall filter by replacing every impulse in the descrete signal with the ***"sinc"*** function:

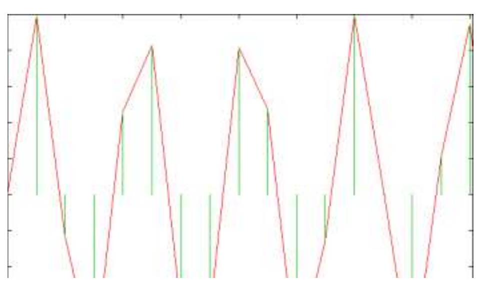


$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

The only problem with this is that the sinc function is infitinly wide, so we have to approiximate it

Interpolation (2) When do you need sinc

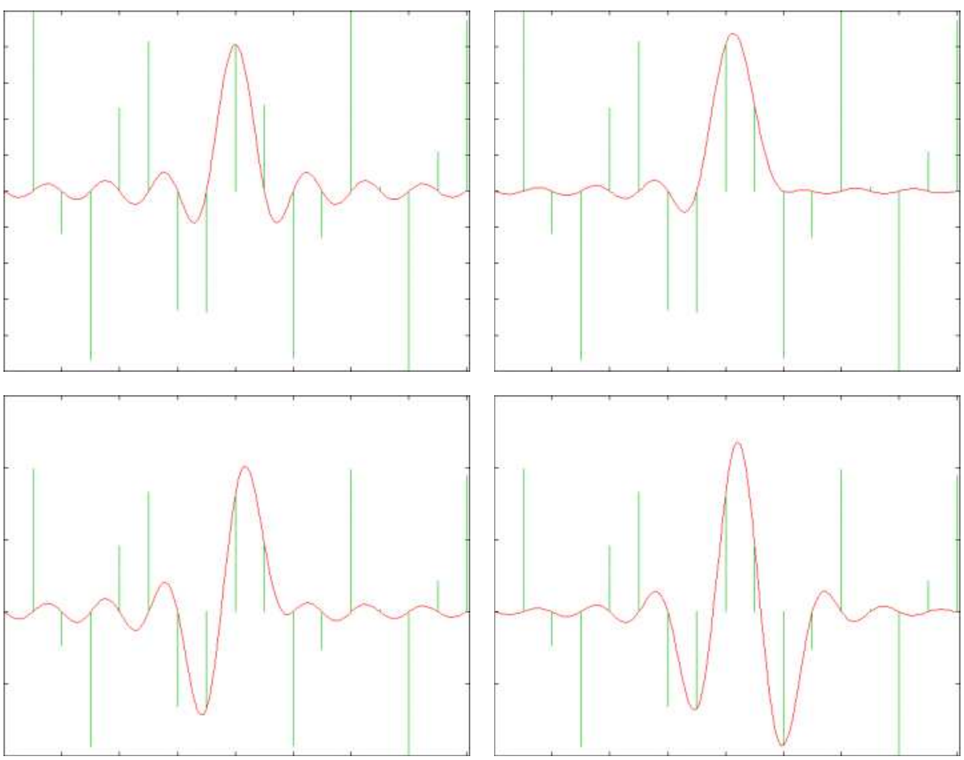




The original PCM data doesnt look like much, and linearly interpolating it just gives rubbish

But when sinc interpolated it can clearly be seen as a sinewave

Interpolation (3) Example stages of sinc interpolation

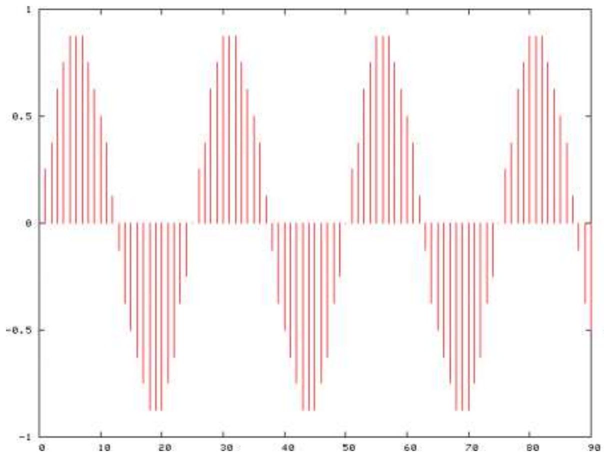


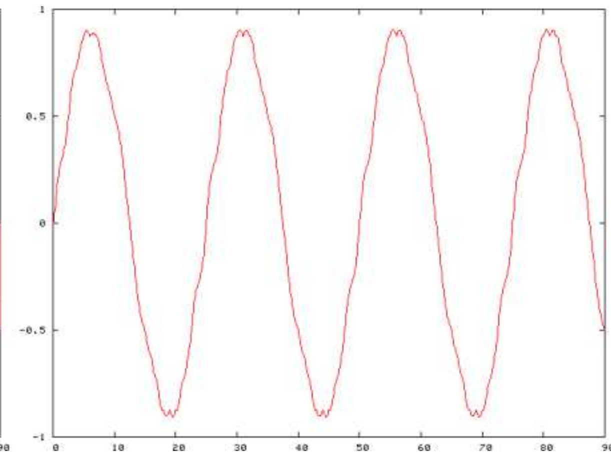
Sinc waveforms are progressivly added inplace of the impulses

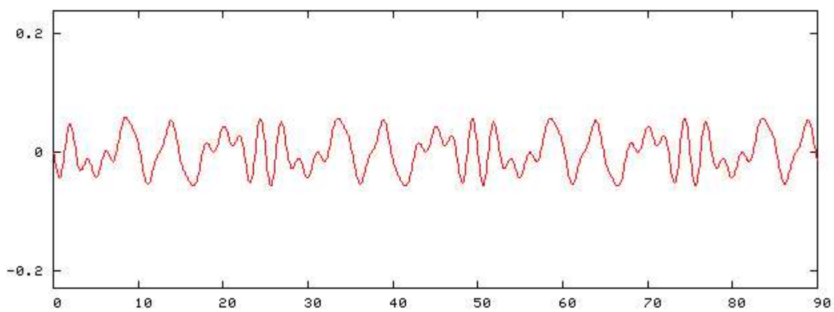
This process is called ***convolution***

Problem: intermodulation distortion

A 4 bit sample of a generated sine wave appears as:



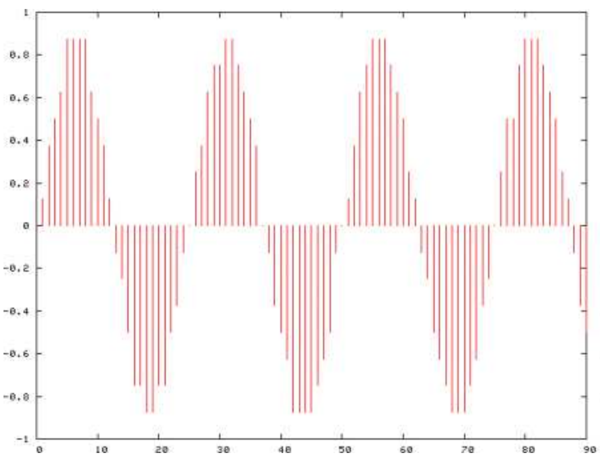


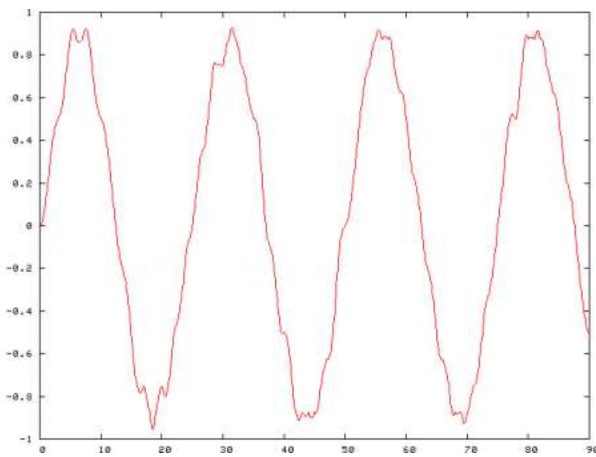


Note that the quantisation noise from the sinewave contains regularly repeating signals.

Dither

Dither is a way of masking the intermodulation distortion, a small ammount of noise is added to the signal before its truncated

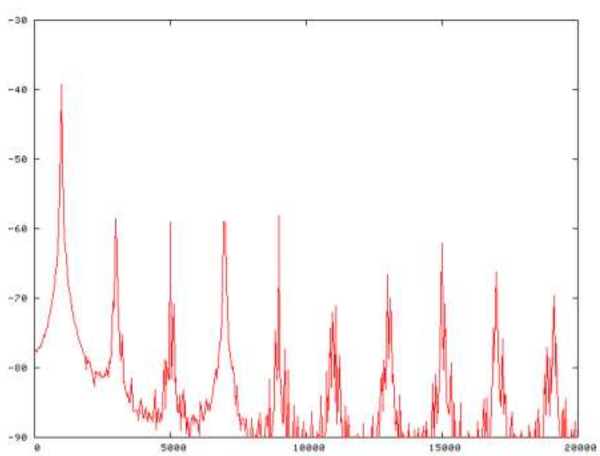


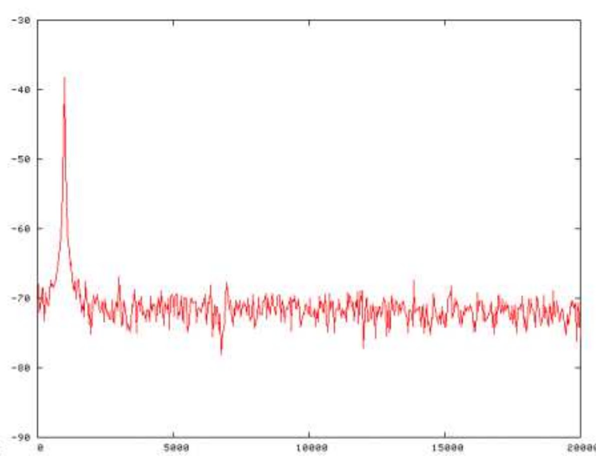


This makes the signal noisier (hurts the SNR), but makes the intermodulation distortion even whitenoise, unrelated to the source signal and so less anoying

Dither (2)

This is a 1k Hz sine wave at -40dB’s truncated to 8 bits, both undithered and dithered



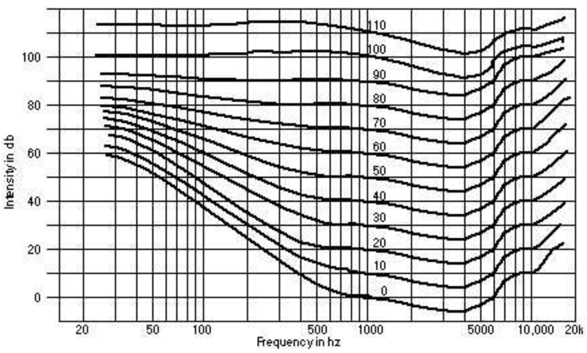


The graph on the left shows the intermodulation distortion caused by the truncated, undithered signal

The graph on the right shows the lack of distortion, but reduced signal to noise ratio

[undithered sine], [dithered sine]

Aside: equal loudness curves



These curves show signals of equal "loudness" (i.e. perceived volume)

They are known as the ***Fletcher-Munson*** curves

They indicate how loud a sinewave at a particular frequency will sound, compared to a sinewave at another frequency

This is relevant to dithering...

Triangular dithering

As the high frequencies are percieved less strongly by passing the noise through a gentle high pass filter and amplifiying it so it has the same ammount of energy we can still mask the distortion, but make the noise less "loud"

The peak noise is higher - but that peak is around 20kHz where human hearing is not very sensitive.

Around the crucial 1-4kHz mark the noise floor is lower (by about 5dB's) so we get more audible detail.

This is called ***triangular dithering*** because the noise floor slopes upwards.

[rectangular dithered sine], [triangular dithered sine]

Note this only works with audio, and only if you're not going to mess about with the frequency components afterwards, as its a psychoacoustic trick

Questions?

Basic signal operations

Mixing

Mixing audio is just the ∑ operation, i.e. +

So to mix signals x1[] and x2[] into y[] we simply do:

```
for (i=0; i<length; i++) {
        y[i] = x1[i] + x2[i];
}
```

Every time you add a channel the theoretical peak value doubles (if the peaks of the two signals are the same)

You can divide by the number of channels to guarantee no clipping, but then the volume will appear to go down as the number of chanels increases

Gain

Amplifying/attenuating audio is just the   operation, i.e. *

So to attenuate x[] by 20dB (-20dB gain) into y[] we simply do:

```
for (i=0; i<length; i++) {
        y[i] = x[i] * 0.1;
}
```

Gain is usally expressed in dB’s, relative to a gain of 1.0 (unity) so we can do:

```
double coefficient = pow(10.0, gain / 20.0);
```

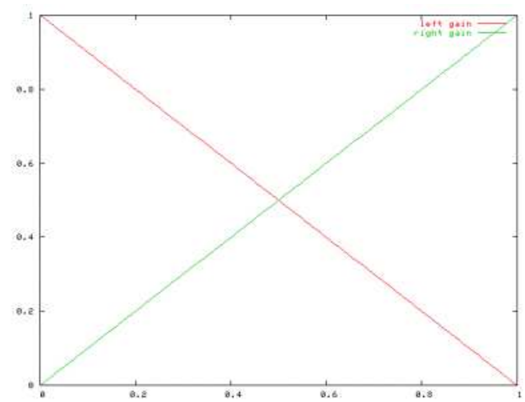
```
for (i=0; i<length; i++) {
        y[i] = x[i] * coefficient;
}
```

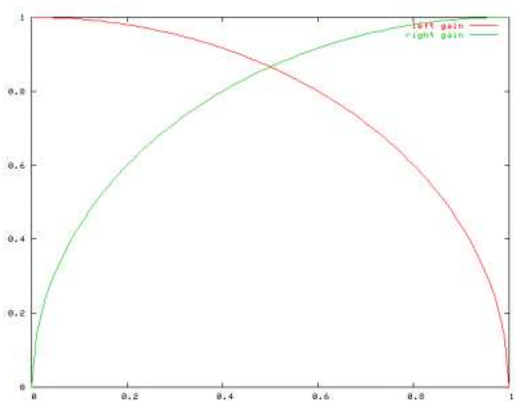
Pan

Pannning is apply gain to a/sum channels to make them appear to come from somewhere between a set of speakers

The simple case is for a single channel to be panned between two speakers

We want to maintain an equal power as the signal pans between the speakers, to avoid the dead spot if we do the naive thing of fading linearly between 0 at one end and 1.0 at the other





Constant power pan however try to retain a constant power aproximation (left^2 + right^2) to avoid the "dead spot"

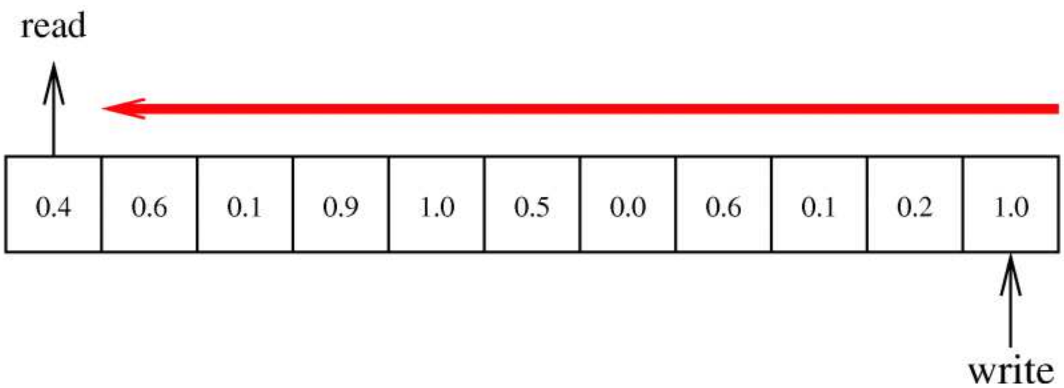
[linear pan], [constant power pan]

Delay

Another fundamental operation on digital signals is delay

Memory and delay are the same thing

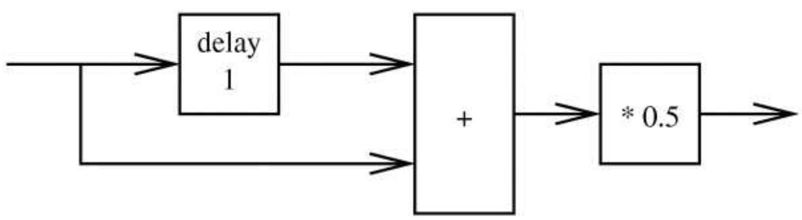
Simply storing samples and getting them back later



Phase shift

From the block diagram or euqation you can see that part of the signal that makes up the output signal comes from back in time

The delay in time varies with the frequnecy of the input signal



This is known as ***phase shift***. Its not possible to have a filter that does not cause phase shift, but it is possible to make the phase shift equal for all frequencies

Because the delay varies with frequency this means that filtering can cause different partials of the signal to become misaliged with each other in time. It’s rarely a problem, and sometimes it’s a desirable side-effect of filtering

Controlling timing (delays)

Delays (1): terms

There are two meanings to the word ***timing***:

- the ***time feel*** in the musical sense; whether the music is "sloppy" or "tight" and "ahead" or "laid-back";
- the way different sound sources ***coincide***.

Both can be manipulated by delay stages. Manipulating the first is purely ***creative*** (i.e. a matter of taste), while the second is ***corrective*** and needs to be done precisely right.

Delays (2): Correlated signals

Signals are said to be ***correlated*** if there is a fixed-phase relationship between them. This usually means they came from the same source at some time.

Correlated signals can be ***misaligned***. This happens in a number of situations:

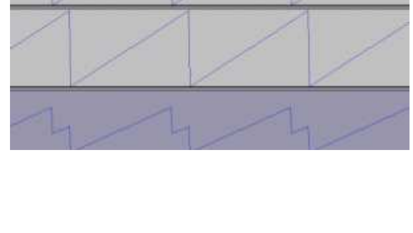
- many microphones are used for a large band, and each player is picked up both by their own and their neighbour's microphones (which are farther away);
- a single microphone picks up both direct and reflected sound (which takes a little longer);
- a signal is combined with a very short delay effect;
- a huge P.A. system uses speaker clusters at different distances to the listener.

When correlated but misaligned signals are mixed, there will be ***interference***. They can be realigned with corrective delay.

Uncorrelated signals can always be mixed together without artifacts: we are free to use creative delays as we please.

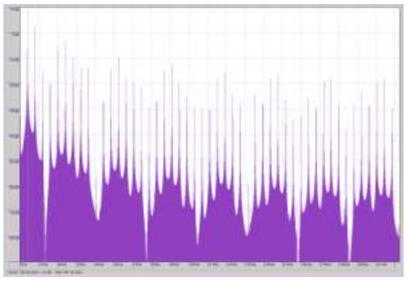
Delays (3): The comb-filter effect

Consider a sawtooth signal (maybe from a violin) being picked up by two microphones at different distances. The signals will look as follows (with the result at the bottom):



All wavelengths in the signal that divide the distance between the microphones will amplify, while those which leave a remainder of half the wavelength will cancel.

This is called the ***comb-filter effect***. Comb-filtered signals sound hollow and are totally immune to equalization.



It can be compensated by applying a corrective delay to the leading signal.

To find the correct amount, do a rough calculation (time = distance / speed of sound), start with that value and fine-tune by ear until the sound is full and natural.

Delays (4): The Haas effect

Timing is crucial to our perception of direction.

Two sound sources more than 30 ms apart will be heard as an echo. If their timing difference is less than that, they will "blur" into one.

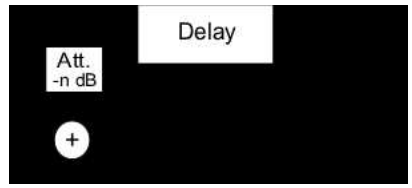
Still, we will perceive the sound to originate from the earlier one, even if the other is louder.

This is called the ***Haas effect***, and is exploited for huge P.A. systems at rock festivals:

A spectator standing behind the secondary P.A. rigs (the ***delay line***) will get most of the sound energy from them. Since they are delayed a little more than their actual distance to the main rig, the main rig hits earlier, and the sounds seems to originate from the stage.

You can do interesting panning tricks with short delays.

Delays (5): Creative use



Delays can be used as ***echo effects***:

Part of the signal is routed back through a delay, attenuated, and mixed into the original stream.

The attenuator is important to avoid runaway feedback.

Digital audio workstations or digital multitrack tapes provide ***track delay***. Individual tracks can be moved back in time.

Track delays can be used to change the "feel" of certain tracks, for example to make the bass more "laid back" compared to the drums.

They can also be used correctively, to align correlated signals.

Questions?

Controlling spectrum: equalization

Equalization (1): why and how

The ideal audio chain has a linear ***frequency response*** - it does not color the sound at all.

Of course, such a chain does not exist.

We need a way to tweak the frequency response, either to correct for deficiencies in the chain, or for musical reasons.

This can be accomplished by ***filters***.

A bank of filters is commonly referred to as an ***equalizer***, since originally it was used to equalize (flatten out) the frequency response of a system.

Filters are frequency-selective circuits.

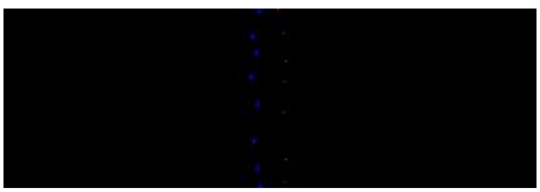
There are two simple components that show frequency-dependent behaviour towards a current: ***capacitors*** and ***inductors***.

We can use them to build analog filters.

Equalization (2): capacitors

A capacitor can be thought of as two opposite plates that can store electrical charge.

They are isolated from each other by an air gap or a special insulator (the dielectric), but are close enough to affect each other through an ***electric field***.



No electrons can pass through the air gap, but their electric field can push those on the other side away.

So, no DC goes through, but impulses do.

At very high frequencies, where the electrons hardly move long ways but only shuffle back and forth, the capacitor behaves like a piece of wire - its ***capacitive reactance X_C*** is low.

We have

$$X_C = 1 / (2\ \ f\ C)$$

where f is the frequency in Hertz and C is the ***capacity*** in Farad.

Capacitors introduce a ***phase shift***: the current ***leads*** the voltage by 90°.

Equalization (2): capacitors (cont'd)

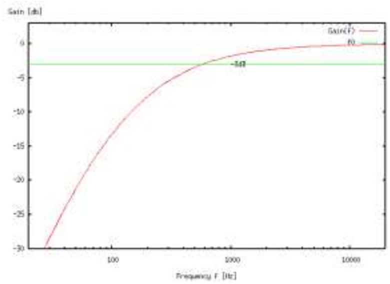
The capacitor blocks low frequencies, and lets high ones pass.

With an additional ***resistor*** acting as a ***voltage divider***, we can build a simple ***high-pass filter*** with a slope of 6dB per octave.

The ***cut-off frequency f_0*** (defined as the -3dB point) is

$$\mathbf{f_0 = 1 / (2 \quad R\ C)}$$

where R is the value of the resistor in ***Ohm*** and C is the capacity.



The frequency response of such an ***RC circuit*** is

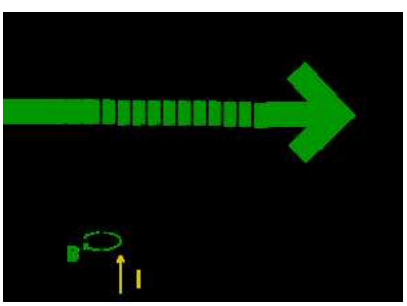
$$\mathbf{G(f) = 20 \log (U_{out} / U_{in})}$$

where G is the gain in dB, and U_{out} and U_{in} are proportional to the resistances across in- and output:

$$\mathbf{U_{out} \sim R,\ U_{in} \sim R + X_C,\ X_C = 1 / (2 \quad f\ C)}$$

Equalization (3): inductors

An inductor is a coil of wire, either hollow or wrapped around a ferrit core.



Coils produce a ***magnetic field*** when current flows through them, and a voltage is ***induced*** into them when they are in a *changing* magnetic field.

When DC flows through the coil, its magnetic field is *static* - hence no induction.

When the current changes (as with AC audio signals), so does the field - and a voltage is induced.

The coil induces a voltage into itself. It is opposite to the original and reduces the current that passes through the coil. The faster the field changes, the stronger this ***self-induction***.

Thus the higher the frequency, the higher the ***inductive reactance X_L***.

We find

$$\mathbf{X_L = 2 \quad f\ L}$$

where L is the coil's ***inductivity*** in Henry.

Inductors also introduce a ***phase shift***: the current ***lags*** the voltage by 90°.

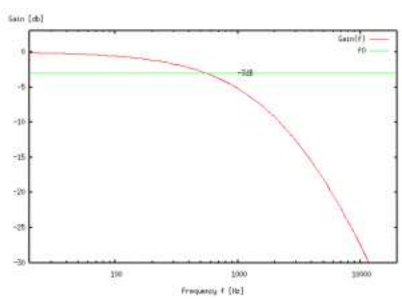
Equalization (3): inductors (cont'd)

The inductor blocks high frequencies, and lets low ones pass.

Again, with an additional ***resistor*** for ***voltage divider***, we can build a simple ***low-pass filter*** with a slope of 6dB per octave.

The cut-off frequency f_0 here is

$$f_0 = R / (2\ \ L)$$



The frequency response of an ***RL circuit*** is

$$G(f) = 20 \log (U_{out} / U_{in})$$

with

$$U_{out} \sim R,\ U_{in} \sim R + X_L,\ X_L = 2\ \ f\ L$$

Equalization (4): variants

Circuits that affect all frequencies above or below a certain point are called ***shelving filters***.

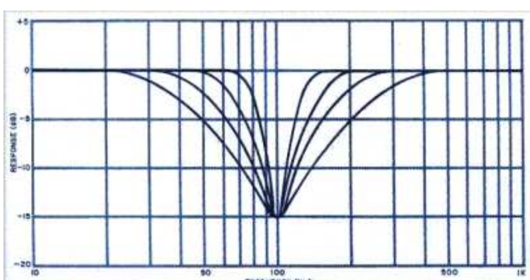
High- and low-pass filters can be combined into ***band-pass*** or ***peaking filters***.

Since they can only remove energy, they are called ***passive filters***.

By adding amplifiers, we can create ***active filters*** that allow boosting.

The filter properties can be improved by cascading filter circuits, and made tweakable by using potentiometers instead of resistors.

Equalization (5): the parametric EQ



Parametric equalizers have three controls:

- ***center frequency***,
- ***gain*** and
- ***bandwidth*** or ***q factor***.

The center frequency is the frequency of the peak, the gain is the amount of boost or cut at that peak, and the bandwidth is width of the bell in octaves between its -3dB points.

A filter with fixed bandwidth is called ***semi-parametric***.

A typical channel strip has high and low shelving filters and one or two (semi-)parametric bands.

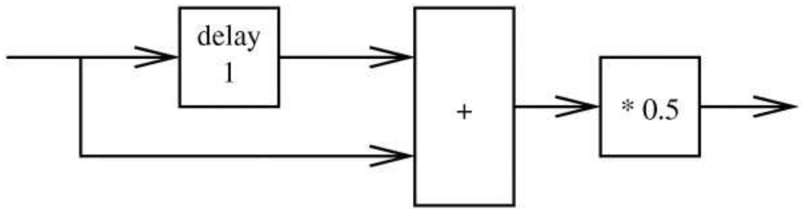
Filters

The most basic filter is simply a delay, mix and gain operation

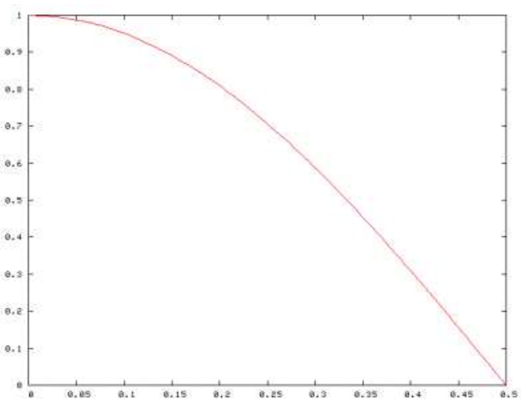
This is the simplest ***1st order lowpass filter***

In maths form its:

$y_n = (x_n + x_{n-1}) / 2$

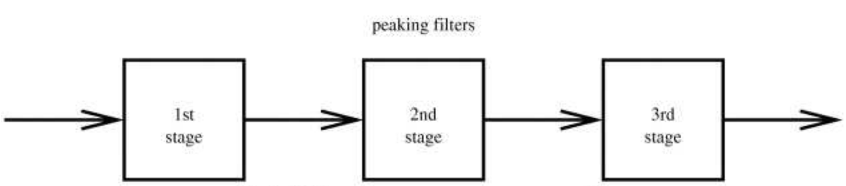


All normal filters can be represented as a combination of delays, mixes and gains

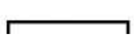


This frequency response of this filter shows a very shallow curve towards blocking the frequency componenets at half the sample rate.

Internal workings of EQs



Serial EQ arrangement e.g. parametric



In the serial arrangement the filters are peaking filters. This means that the phase shifts, rounding errors and so on build up with each stage.

In the parallel arrangement each band only goes through one filter, so the errors do not build up as much, but the bandpass filters must allign exactly, otherwise there will always be peaks and/or throughs in the signal where the bands do not meet exactly

Using an equalizer

0th Rule: **Listen!**

- Before you do anything else, walk up to the instrument or singer and listen quietly and patiently. This is the ideal you want to recreate.

1st Rule: **The best EQ is no EQ!**

- Don’t try to compensate bad mic placement, cheap mikes or shabby room acoustics with EQ.

2nd Rule: **Down is good, up is evil!**

- First try to listen for bands that need attenuation before boosting everything else.

3rd Rule: **Sweep is your friend.**

- If you have a semi-parametric filter, boost the gain and sweep through the available range to get a feel for how all the frequencies sound.

4th Rule: **Start with low Q.**

- If you have a fully parametric filter, set it to a broad bandwidth, zero in on the problem and then narrow the bell until the effect is lost, then back out a little.

Using an equalizer (cont’d)

5th Rule: **A/B frequently.**

- If you have an EQ bypass, use it often. The flat signal is your reference.

6th Rule: **You can’t boost what isn’t there.**

- If you have room cancellation effects on your recording, or there’s simply no sound in a particular range, there’s no point in cranking up the EQ.

7th Rule: **Watch out when boosting a bass shelving filter!**

- It will boost everything below, the deeper the louder, and you will almost certainly not realize the problem on your home monitors.

8th and Ultimate Rule: **Beware of smartasses telling you how to use an EQ!**

- Anything goes. After all, the mix has to work, and if it means to bend the signal till it screams, so what!

Questions?

Controlling dynamics (compressors/expanders)

Compression (1): terms

The difference between the softest and the loudest passages is called the ***dynamic range*** of a signal.

Macrodynamics is the loudness difference between sections (such as chorus, verse, bridge and solo). It affects the listener's perception of ***suspense*** and ***drama***.

Microdynamics is the peakiness of the signal (imagine drums vs. pipe organ). It affects the perceived ***loudness***, ***punch*** or ***power***.

Short, percussive peaks in the signal are called ***transients***.

In an ideal world, all media are good enough to faithfully reproduce the original dynamic range, all listening environments are quiet enough to reveal the soft passages, and all listeners are attentive and knowledgeable enough to enjoy it.

Ok, forget it. Let's fiddle.

Macrodynamics can be controlled by manual ***gain riding*** or mixing automation.

Manipulating microdynamics calls for ***automatic gain control***. Enter the ***compressor***.

Compression (2): how?

A compressor can be thought of as an extremely quick audio engineer. It listens to a signal and controls its level in accordance with predefined rules.

"Listening" here means measuring the signal loudness, either by looking at the ***peaks*** or by averaging the level over time (the ***root-mean-square*** or ***RMS*** algorithm), and producing a ***control voltage***.

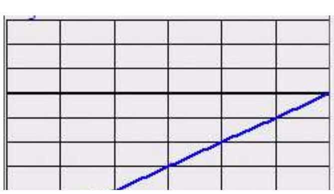
This is then fed into a ***VCA*** or voltage-controlled amplifier that manipulates the signal.

The behaviour of the VCA can be controlled by a number of parameters:

- ***Threshold***: the level at which the compressor starts to act;
- ***Ratio***: the amount of compression;
- ***Attack***: how fast the compressor reacts to transients;
- ***Release***: how quickly it stops compressing after a peak;
- ***Knee***: how harshly the compression sets in.

The current amount of compression can be monitored on the gain reduction meter.

Compression (2): how? (cont'd)

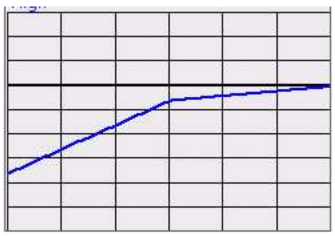


Here is the transfer function of a compressor with ratio 1:1 - it will not do anything.

Here, the ratio is 5:1 and the threshold is -30dB:

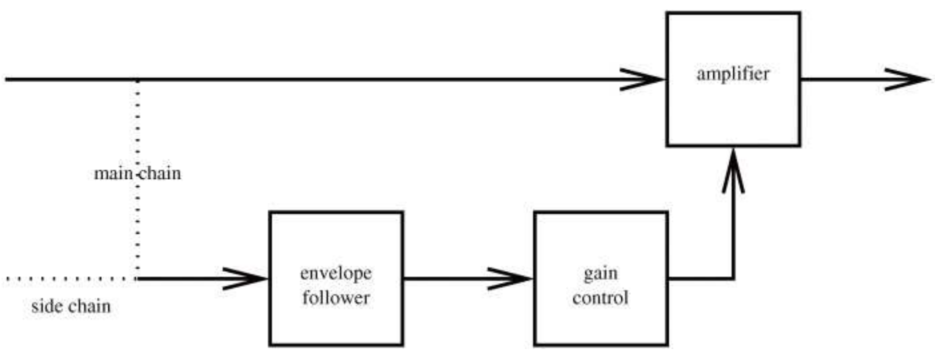
Above the threshold, the output will be compressed. 5:1 means 5dB more input result in only 1dB more output.

Note that a compressor at first only ***softens*** the signal.



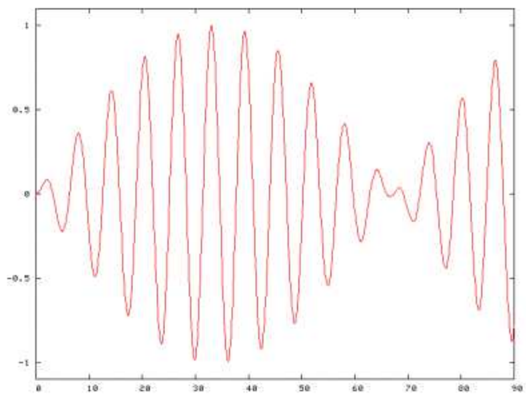
After the peaks are trimmed off, the level is raised to the old peak level with the ***makeup*** or ***output gain***. Now the signal is louder, even though the peaks are the same height.

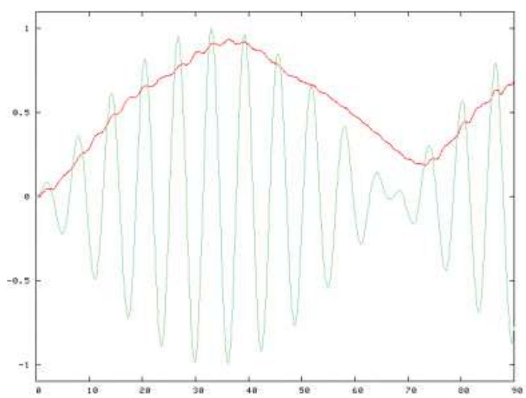
Internal workings of compressors



The envelope follower reads the signal level form either the side chain input or the main input, feeds it to the gain control module (which does the ratio conversion) and uses that to drive an amplifier on the main output

Compressor: envelope followers





The envelope follower tracks the ampltude of the incoming signal. The output is unipolar (positive values only) as we dont care about the sign of the signal, just its amplitude.

Notice that the output shape of the envelope follower is bumpy and doesn’t track the waveform very well. Thats because its a hard problem, and this is one just one I hacked up for this demonstation.

Tuning the envelope follower correctly is one of the hardest things to get right in digital compressors.

Compression (3): why?

Sometimes, gentle sum compression is needed because the dynamic range exceeds the S/N of the target medium or environment. This may be the case if you are mixing

- for analog tape, vinyl or radio;
- extremely demanding orchestra material for CD;
- for noisy environments (car radio etc.).

(Some of these situations only call for gentle macrodynamics reduction and can be handled with manual gain riding.)

Compression can be used on single tracks for purely esthetic or creative reasons, such as to give a singer or instrument more "presence".

Most often, extreme compression is used indiscriminately on the sum, to make the mix louder than anyone else. This is why commercial pop music is so boring.

Compression (4): Usage tips

When using a compressor for the first time, switch its attack and release controls to "automatic".

Set the threshold to "max" (= neutral) and decide whether you want strong (5:1 and up) or gentle compression (1.5:1 - 5:1).

Move the threshold down until you begin to hear the desired effect.

Calibrate the output gain so that the processed signal is as loud as the original, using the bypass switch.

Now turn the attack and release knobs to the middle and switch to manual control.

Use the attack time to control how much transients will be compressed. If the signal is too thin over all with very loud drum hits, use a short attack time to reduce them. If drums sound lifeless and squashed, set a longer attack.

If the compressor is "pumping" (i.e. loud transients are followed by a dent in loudness), reduce the release time.

When compressing stereo signals, make sure you link the control voltages for both channels (most units have a "stereo link" switch). Otherwise, the stereo image will jump around.

Compression (5): Other usage examples

De-Essing

Some compressors have a ***side-chain*** input. Whatever you send there is used for the loudness measurement instead of the normal signal. You can use this feature to build a ***De-Esser***:

Patch a vocal track with excessive sibilance into the compressor, and also through an equalizer. Use the EQ to boost the sibilance even more, and feed the resulting signal into the side-chain.

You will get a compressor that is extremely sensitive to sibilance and will reduce the level whenever it gets ugly.

Limiting

A compressor with very high ratio (:1) and short attack time is called ***limiter*** and can be used to protect digital media from overloading or P.A. systems from excessive level. It can also be abused to increase loudness at the cost of naturalness and "punch".

Expansion

An ***expander*** works complementary to a compressor: it makes soft signals even quieter.

The most common use case for this is the ***noise gate***: whenever the program material is so quiet that the noise in the chain becomes audible, it can be reduced or completely muted by the gate.

It is important to get the attack and release times right, or the gate will be "stuttering" around its threshold.

Expanders can also be used to create the famous "In the Air Tonight" ***gated reverb*** drum sound. Mix drums with way too much reverb, and cut off the reverberation as soon as the drum sound has decayed.

Unfortunately, it is almost impossible to liven up material with an expander that has been compressed to death.

Questions?

Space and spatialisation

Spatial perception (1)

There are two aspects to spatial perception:

- to hear from which ***direction*** and ***distance*** a sound comes, and
- to get a picture of the ***acoustics*** of the room

From an audio engineering POV, the first boils down to stereo (and surround) techniques, and the second to reverberation.

Both aspects can be captured during recording, or they can be faked afterwards.

Spatial perception (2)

Our brain gets three clues to work out the direction of a sound:

- loudness differences between the ears,
- timing differences between the ears, and
- sound coloration due to damping and refraction around the earlobes and head.

Distance can be derived from

- the relative loudness,
- the amount of high-frequency damping (the duller, the further), and
- the ratio of direct to reverberated sound.

All these mechanisms can be (and are) used in recording.

Recording or faking distance information

The placement of the microphone is the most crucial factor.

For a certain microphone pattern in a certain room, there is a specific distance to the source at which the direct sound is as loud as the reverberated sound: the ***reverb radius***.

By moving the microphone in and out of the reverb radius, we can control the amount of reverb and thus the perceived distance.

Air damping can be faked by filtering the treble.

In pan-pot stereo, you can fake distance by reducing volume and adding more reverb and a a slight attenuation of the mid frequencies (the ***presence range***).

Stereophonic recording (1)

Distance information can in theory conveyed with only one source (***monaural*** or ***monophonic*** sound). But then everything seems to come from the same place, so this is not very interesting.

To create or recreate an impression of direction and spatial sound, it is essential to have at least two sources (***binaural*** or ***stereophonic*** sound).

There are two basic stereo techniques: ***intensity stereo*** and ***phase stereo***.

Things to keep in mind are ***mono compatibility*** and control of ***stereo width***.

Stereo (2): Intensity stereo

Intensity stereo utilizes the brain's sensitivity to loudness differences, and can be done with two directional microphones in one spot, turned slightly outwards (***X/Y*** miking).

The perceived width of the stereo field is determined by the ***opening angle*** of the microphones.

Intensity stereo recordings can be mixed down to mono without problems.

When increasing width, watch out for a hole in the middle of the stereo field.

Stereo (3): Phase stereo

Phase stereo works through timing differences, and is usually accomplished by two omnidirectional microphones at placed a distance (***A/B*** miking).

The perceived width is controlled by the distance between the microphones.

Again, there will be a hole in the middle of the stereo field if the mikes are pushed too far apart.

Phase stereo recordings can suffer from comb filtering effects when mixed down to mono.

Stereo (4): other techniques

It is also possible to combine the two techniques by using slightly angled directional microphones at ear distance (NOS and ORTF patterns). They generally convey a more convincing stereo image, since they recreate two different spatial clues.

A nice variant of intensity stereo is ***mid/side (M/S) stereo***. It consists of an omni or directional center microphone and a figure eight one with the lobes aimed left and right. It is equivalent to X/Y and must be converted as follows:

X = M + S, Y = M - S

The nice thing is you can control the width after the recording by changing the S level.

The crudest stereo technique is ***pan-pot stereo***, achieved by panning close-miked or DI mono sources. It needs artificial reverb to sound convincing.

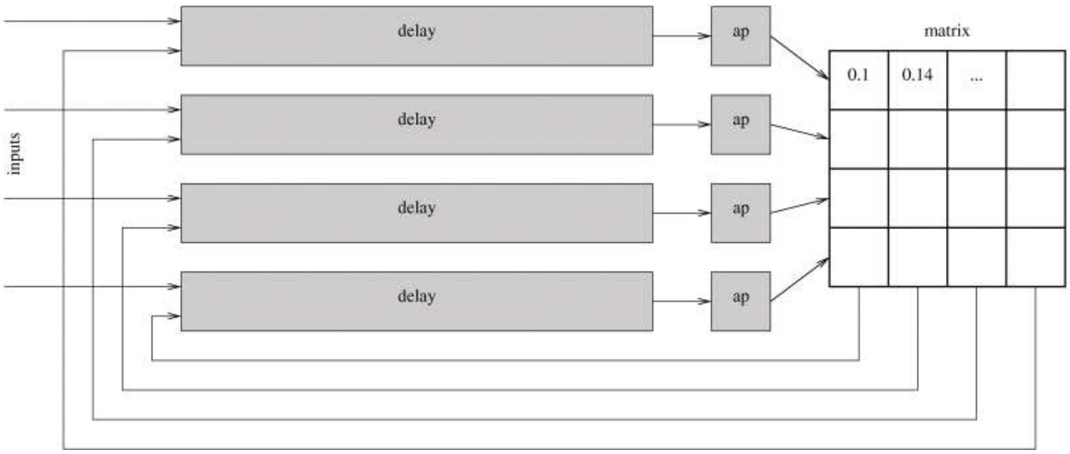
Reverb (1): some terms

Reverberating sound has a number of phases:

- the dry-sounding ***pre-delay*** phase before the sound has hit the walls;
- the distinct ***early reflections***; and
- the ***tail*** where the reflections are so numerous and so dense that they blur into a continous sound.

Reverb (2): Artificial reverb

Artificial reverb attempts to recreate the sense of space got from the reflections from the walls of a physical room



The delay lines (in reality there are more than 4) feed into allpass filters and then into a multiplication matrix that mixes portions of the input signals into each output.

The delay line lengths and allpass filters need to be very carefully tuned to prevent metalic or grainy sounds.

Reverb (3): Convolved reverb

Convolution can be used to recreate reverb from impulse responses

If the source signal is convolved with an impulse response captured from a room it will recreate all the linear (time and frequency response) characteristics of the room

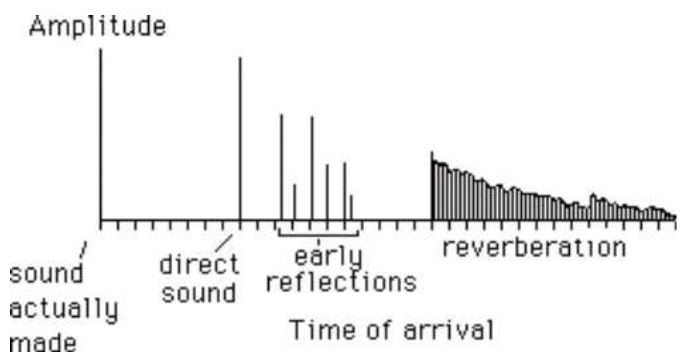
Impulse response captured from Volksbad

This impulse response is used inplace of each sample in the source signal, just as for sinc interpolation

Reverb (4): Capturing impulses

The simplest way to capture impulses is to play a single sample pulse in the environment you wish to capture and recording it with a microphone

Sometimes better results can be got by playing a "chirp" or swept sinewave and ***deconvolving*** it back to the equivalent impulse



[Volksbad impulse]

Impulses can also be captured by passing impulses through outboard equipment

One thing that cannot be captured properly is dynamics control (eg. compressors) and non-linear effects (eg. distortion)

Reverb (5): Reverb chambers

Another simple method to add reverb to a sound is to feed it to a speaker in a good-sounding room and pick it up with a microphone. When patched through an aux send, the amount of reverb can be controlled precisely.

With good rooms, speakers and microphones, the result can be very good.

On the downside, the reverb properties are fixed, as with convolution reverb.

Reverb chambers can be thought of as "acoustic convolvers".

Reverb (6): Using reverb units

Algorithmic reverb units have a number of tweakable parameters:

- ***Decay time*** is the time it takes for the reverb to fade away (= to fall -60dB below the original signal level). Rule of thumb is long decay for slow pieces, short decay for faster ones.
- ***Pre-Delay*** is the time span between the original signal and the first reflections. If set too long, it can create an unnatural delay effect. Used sparingly, it can help increase speech intelligibility without reducing the reverb density.
- ***High-frequency damping*** allows you to simulate hard or carpeted walls.

Some reverbs allow detailed configuration of the room dimensions, or to choose from simulations of mechanical reverb generators (plate, spring). Some also let you tune the levels of the early reflections and the reverb tail separately.

Some ideas for spatial-sounding mixes

Try to capture space during the recording stage. Choose good sounding rooms.

If you want natural sounding space, use the same reverb for all signals.

Use the reverb amount to control the perceived distance (the drier, the closer).

Use additional ambience mikes. If they color the close-miked sound too much, EQ the low and low-mid frequencies away and just keep the "air" quality.

Try omni mics, or directional mics at a greater distance.

To widen the stereo image, use phasing: it is sometimes possible to make a sound come from outside the speakers by adding some phase-reversed signal to the opposite channel.

All phasing tricks can potentially mess up the low end. Carefully check for cancellation effects.

References and recommended reading

Bernsee, Stephan M.: "The DSP Dimension" [http://dspdimension.com/]

Capel, Vivian: "Public Address Systems", Focal Press 1992

Görne, Thomas: "Mikrophone in Theorie und Praxis", Elektor-Verlag Aachen/Germany, 1994

Katz, Bob: "Mastering Audio, The Art and the Science", Focal Press 2002

Katz, Bob: "Digital Domain" [http://www.digido.com/]

Roads, Curtis: "The Computer Music Tutorial", MIT Press 1996

Smith, Steven W.: "The Scientist and Engineer's Guide to Digital Signal Processing", California Technical Publishing 2003 [http://www.dspguide.com]

Stark, Peter A.: "Electronics 101", 1998 - 2004 [http://www.users.cloud9.net/~stark/elbook.html]

Various authors: "Music-DSP Source Code Archive" [http://www.musicdsp.org]

Warstadt, Michael / Görne, Thomas: "Studiotechnik, Hintergrund und Praxiswissen", Elektor-Verlag Aachen/Germany 1994

Thank you.

Q & A