Stereo Processing

Cross-Correlation - A Measure of Signal Similarity

When we are dealing with stereo signals, we have effectively two separate signals, one for each channel. We will denote the left channel signal as $x_L(t)$ and the right channel signal as $x_R(t)$. Those two signals - although well separated - are usually not independent from each other. Instead, they exhibit a certain similarity. In the case where they are indeed exactly equal, we have nothing more than a mono-signal, just played over two channels. To quantify the apparent width of a stereo-signal, we need a means to measure the similarity between the two signals x_L and x_R. The cross-correlation between the two signals is such a measure and it has the following properties:

- If both signals are equal, the cross-correlation between them is 1.
- If one signal is the negative of the other, the cross-correlation is -1.
- If the two signals are not similar at all, the cross-correlation is 0.

The cross-correlation is computed by multiplying both signals, averaging this product over time and finally normalizing this value to the range between $-1...1$. Let's denote the cross-correlation with the variable C, then the formula for calculating the cross-correlation becomes:

$$C = \frac{E\{x_L(t) \cdot x_R(t)\}}{\sqrt{E\{x_L^2(t)\} \cdot E\{x_R^2(t)\}}} \tag{1}$$

where $E\{\ldots\}$ means to take the expectation value (the average, that is) of the expression inside the braces. Thus, the numerator is the time averaged product of both signals. The denominator is a normalization factor which ensures that C is always in the range $-1...1$, regardless of the overall volume of the two signals. In practice, the expectation values are replaced with short time averagers (simple first order filters can be used for this) and the short-time averaged value can be shown in a metering display.

Correlation of phase-shifted Sinosoids

Assume, both signals x_L and x_R are sinosoids of the same frequency ω and with the same amplitude A but different phases φ_1, φ_2, such that:

$$x_L(t) = A \sin(\omega t + \varphi_1) \qquad \text{and} \qquad x_R(t) = A \sin(\omega t + \varphi_2) \tag{2}$$

Then, a relationship between the phase-difference $\varphi = \varphi_2 - \varphi_1$ and the cross-correlation C can be established:

$$C = \cos(\varphi) \qquad \Leftrightarrow \qquad \varphi = \arccos(C) \tag{3}$$

where a cross-correlation of $C = 1$ maps to $0°$ phase-shift, $C = 0$ maps to $90°$ and $C = -1$ maps to $180°$. A word about 'phase-inversion' is in order: for sinusoids, it is true that inverting the signal (switching the polarity) is the same as phase-shifting the sinusoid by $180°$. This, however is not generally true for other periodic signals (a $180°$ phase-shifted saw is not the same as an inverted saw). Note also, that talking about a signals phase makes sense only for periodic signals. Thus, terms like phase-inversion or phase-correlation are to be taken with a grain of salt - frankly, i regard them actually as misnomers.

Stereo-Width and Mono-Compatibility

When we need to convert a stereo signal into a mono signal (in order to play it on a mono-system, for example), we usually simply add the left and right channel signal (and possibly multiply the result with a scale factor) to obtain the mono signal. Let's stick to the case that x_L and x_R are sinusoids of the same frequency and amplitude. What this summed mono-signal $x_M = x_R + x_R$ will look like, depends crucially on the phase relationship between the two sinusoids. To illustrate that, figure 1 depicts 3 sinusoids with the same frequencies and amplitudes but with different phase-shifts. $s_1(t)$ has a phase shift of $\varphi = 0°$, $s_2(t)$ has a phase-shift of $\varphi = 90°$ and $s_3(t)$ has a phase shift of $\varphi = 180°$. Now consider what happens, when

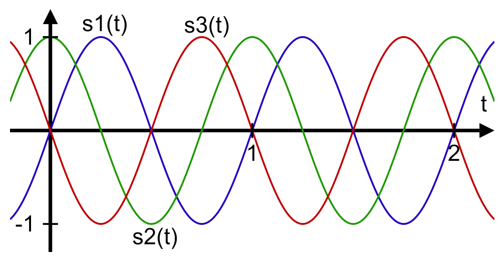

Figure 1: Sinusoids at various phases

summing various combinations of these sinusoids. If the two sinusoids to be summed have the same phase (the phase difference is 0° and $C = 1$), the sum will simply be a sinusoid with twice the amplitude of either of the two (just mentally add $s_1(t)$ to (a copy of) itself). If, on the other hand, the phase difference is 180° ($C = -1$), the two sinosoids will at each instant of time sum to zero (mentally add $s_1(t)$ to $s_3(t)$) - they cancel each other out in the mono sum. The intermediate case in between corresponds to a phase difference of 90° ($C = 0$). The sinusoids will sum up to another sinusoid with an amplitude of $\sqrt{2} = 1.414\ldots$ - this corresponds to adding $s_1(t)$ to $s_2(t)$. Often, when mixing together two channels (or more), one makes the assumption of uncorrelated signals ($C = 0$) in the individual channels. To compensate for the gain factor of $\sqrt{2}$, one often scales the summed signal with a factor $1/\sqrt{2}$. It should be apparent, that a stereo-signal with a cross-correlation of 1 is fully mono-compatible - the mono-sum will be exactly the same signal as both channels alone (up to a scale-factor). This stereo-signal is in fact a mono-signal on two channels and has absolutely no intrinsic stereo-width. The other extreme - a cross-correlation of -1 (one channel is the inverse of the other) - will at all times give zero for its mono-sum. It is maximally mono-incompatible therefore, its stereo-width is very wide but can sound rather odd. Sometimes we can achieve effects of localization out of the physical stereo-field with large negative cross-correlations. Uncorrelated channels reflect the assumption which is often made when summing to mono, in practice however, a correlation of 0 might be a bit too wide. Values in the ballpark between 0.2...0.5 often yield good sounding results. It should also be mentioned, that it is often desired to have a higher cross-correlation (near 1) at low frequencies than at mid and high frequencies ('the bass should be mono').

Mid-/Side Processing

We have established, that certain values for the cross-correlation between the channels are more desirable than others. So what can we do to achieve a specified cross-correlation? A simple means to adjust the cross-correlation (and with it, the stereo-width) provides the technique of mid-/side-processing. The mid-signal is simply the (scaled) mono-sum of both channels and the side-signal is taken to be the (scaled) difference between left and right channel:

$$x_M = \frac{x_L + x_R}{\sqrt{2}} \qquad \text{and} \qquad x_S = \frac{x_L - x_R}{\sqrt{2}} \tag{4}$$

The scaling factors of $1/\sqrt{2}$ are again due to the assumption of uncorrelatedness. A little algebra gives us the formulas to convert back from mid and side (M/S) to left and right (L/R):

$$x_L = \frac{x_M + x_S}{\sqrt{2}} \qquad \text{and} \qquad x_R = \frac{x_M - x_S}{\sqrt{2}} \tag{5}$$

The mid-signal represents those parts of the stereo-signal which are equal on both channels, the side-signal represents the differences between both channels. If we want to widen the stereo-width of the signal we just need to increase the relative amplitude of the side signal before converting back to L/R. Conversely, we can narrow the stereo field by decreasing the relative amplitude of the side-signal. 'Relative' is in both cases to be understood as 'with respect to the mid-signal'. In order to realize different cross-correlation values for different frequency bands, we could apply this kind of relative weighting in a frequency dependent manner by either doing the whole procedure described above separately for various frequency bands (multiband widening) or equalizing mid- and side-signals differently. All mid-/side based stereo widening techniques rely on an already existing (albeit possibly weak) stereo width. If you apply such techniques to a signal which is intrinsically mono in the first place, they simply won't work, because the side-signal is exactly zero at all times and amplifying a zero signal yields a signal which is still zero. To apply some stereo-width to mono-signals we need different techniques which can be summarized under the name...

Pseudo Stereo Effects aka Stereoizers

Pseudo stereo effects or 'Stereoizers' create a stereo signal from a source signal which is possibly (but not necessarily) a mono signal. Actually, many standard effects which are not specifically created to 'stereoize' such as reverb and chorus will do stereoizing as a side-effect, too. The fact, that various effects have stereo signals as their output even when their input is mono may make you suspect that dedicated stereoizers may also utilize different signal processing algorithms to achieve their goal. And this suspicion is right. A very simple means to stereoize a signal is to delay it a bit (in the range of a few to a few tens of milliseconds) and mix the delayed signal with positive phase to one channel and with negative phase to the other. Note that these added delayed signals will cancel each other out in th mono sum. Spectrally, this will result in comb-filtering both channels with complementary comb-filters. A refinement of this technique is to modulate the delay over time with an LFO. Another refinement would be to filter the delayed signal in order to limit the effect to a certain range of the spectrum - it is often desirable to leave low frequencies untouched ('dry bass'), this could be achieved by highpass-filtering the delayed signal. The delay could also be replaced by a frequency dependent delay (an allpass filter). Another approach would be to apply different (complementary) equalization to the left and right channel or use complementary filter banks. And, of course, one could also combine all these approaches and probably others as well - so, stereoization seems to offer a lot of room for creativity and experimentation.