Audio Engineering Society

Convention Paper

Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Stereo and Surround Panning in Practice

David Griesinger[1]

[1] *Lexicon, Bedford, MA 01736, USA*
Correspondence should be addressed to dg@lexicon.com

ABSTRACT

The apparent position of speech and music sound sources was investigated using both a two channel loudspeaker array and a three channel loudspeaker array. The results showed that a sine-cosine pan law was reasonably accurate for the three channel array, but consistently produced sharp images who's positions were consistently wider than expected with a two channel array. The discrepancy was investigated using a headphone model. We found the apparent position depends strongly on the spectrum of the source, with speech frequencies tending to dominate the overall impression.

INTRODUCTION

Nearly all mixing desks are equipped with a control that sets the apparent horizontal position of a single sound source by manipulating the level of that sound source in a number of loudspeakers. The most common of these controls varies the output level of a single channel in two output channels, using a "sine-cosine" pan law, where one of the two outputs is found by multiplying the input by the sine of a control value, and the other by multiplying by the cosine of a control value. The sine-cosine law has the advantage of maintaining constant energy as the apparent position is varied, and has a long history in use. As an example, if:

Left_output = cos(p)*input
Right_output = sin(p)*input

Where p is assumed to vary from 0 degrees (full left) to 90 degrees (full right), with center at p = 45 degrees. Note that then

(Left_output)^2 + (right_output)^2 = (input)^2,

Which indicates constant loudness for many types of signals.

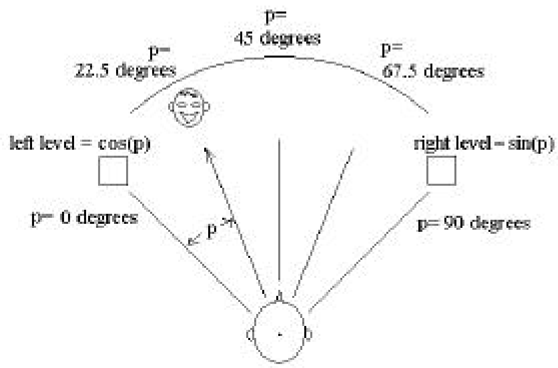

Fig. 1. If the sine-cosine pan law is accurate, the sound image will be heard exactly half way between center and left when p = 22.5 degrees.

If we assume the two channel loudspeakers are separated by +-45 degrees then the apparent position of a sound source should correspond to the setting of the control p. p=0 for left, p=90 for right, and p=45 for center. The full left and full right positions are accurately predicted, since in these cases the sound is being reproduced by a single loudspeaker. The center position is also accurately predicted whenever the two front loudspeakers have the same sound pressure. Although we would expect this result by symmetry, it is not obvious that a central image would be stable as it is when the listener makes small rotations of the head. The stability of the front image is beyond the scope of this preprint.

Many engineers have noticed that the sine-cosine pan law (sometimes referred to as the "tangent" law) is not particularly accurate when a sound is panned somewhere between center and left or between center and right.

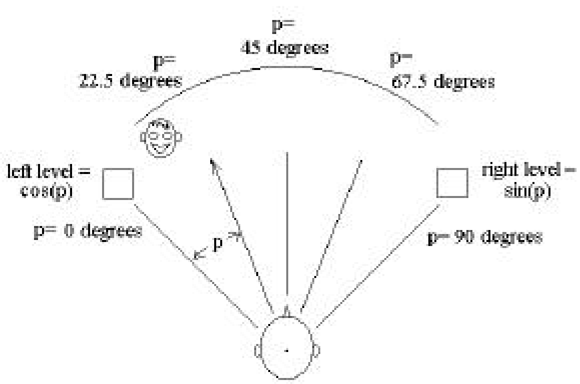

Fig. 2. In practice, the sound image when p = 22.5 is closer to the left loudspeaker. The positions of sound images are consistently wider than the sine-cosine pan law would predict.

The sine-cosine law predicts that when p = 22.5 degrees the apparent position of a sound source will be half-way between center and left. This is usually not the case. In our experience when the control is set to 22.5 the apparent position of the sound source is closer to the left loudspeaker, at a position the sine-cosine law would call 15 degrees (with zero as full left.) See figure 2.

A recent pair of papers [1], [2] investigated the performance of different pan laws, specifically the sine cosine law (labeled the tangent law) and another law (the "sine" law). In these papers the two laws are described as:

If g1 = the gain of left channel, and g2 = gain of the right channel, then

Tangent law:
Apparent position = arctan(tan(45)*(g1-g2)./(g1+g2));

Sine law:
Apparent position = arcsin(sin(45)*(g1-g2)./(g1+g2));

The sine law predicts that the apparent position of a sound source (adjusted by a conventional sine cosine pan pot to the half way point between center and left) will actually be located at about 30 degrees (again with zero at full left.) Thus although the sine law is also accurate at the end-points (full left and full right) as well as at the center, the discrepancy between the predicted and observed positions of a full bandwidth speech signal is larger than for the sine cosine law.

Pan law accuracy matters!

With the advent of surround mixing, the accuracy of the pan law matters. When the only concern was two channel stereo the discrepancy in the sine cosine pan law did not make much difference. Engineers simply adjusted the pan control by ear, and did not much care that its final position was not in accordance with the apparent position of the source.

However, say you are trying to prepare a mix that will be released in both two channel and five channel formats. You want to pan a particular voice half-way between center and left, and you have a mixer that allows you to create this pan both as a two channel pan (using a sine cosine law) and a three channel pan, where the sound energy is divided between the center loudspeaker and the left loudspeaker. Suddenly you need two pan laws, one for a three front loudspeaker system, and one for a two front loudspeaker system. If either of the two pan laws is defective, the apparent position of the voice in the two mixes will be different.

In another application, say you are trying to make a machine that will automatically translate (decode) a two channel recording into a five or seven channel recording, and you want the sound images in the front (using three loudspeakers) to sound in exactly the same place as they sound in the two channel mix (using two loudspeakers). To make this translation you will have to use an inverse pan law to determine the true apparent direction of an incoming sound source, and then you will have to use matrix technology to derive a three channel signal (using a three channel pan law) to create the new image. Once again if either pan law is in error, the images will not be in the same places.

We are in the business of designing such decoders. At their most fundamental level, a good 2-5 or 2-7 decoder consists of a two channel sound direction detector, and apparatus to adjust the gain of the output channels to create the same (or a similar) direction using a larger number of loudspeakers.

Both problems are tricky to solve, and errors in the direction detector can easily be blamed on errors in the panning apparatus, and vice versa. Because we must always derive the center channel from a mix of the left and right input channels, the resulting three front channel sound images are often narrower (closer to the center) than we would like. However at least in theory if we can make a perfect direction detector, and we have accurate pan laws, the image positions will be exactly the same.

We recently put a lot of effort into designing a better sound direction detector, which began to pass all our tests with flying colors. Unfortunately, the resulting three front channel images remained narrower than the originals. Clearly something was wrong – and it had to be the pan laws.

We started an investigation into two channel and three channel pan laws. Surprisingly perhaps, the sine cosine law appeared to work pretty well for a three channel pan. That is, if you make the center channel speaker and the left channel speaker equally loud, then the apparent sound image is pretty much half way between center and left. But the two channel image was much more complicated. This paper presents some of the work that went into understanding the problems of two channel panning.

A few results

It turns out that the psychoacoustics of direction perception is based on some relatively simple physics and neurology, but introducing amplitude panning creates some interesting complexity.

Consider a sound panned half way between center and left. Sound from the left loudspeaker travels directly to the left ear, but must diffract around the head to reach the right ear. The head diffraction adds a certain amount of delay, and attenuates the sound amplitude. This delayed and attenuated signal from the left will interfere with the direct sound from the right loudspeaker.

At some frequencies the interference will be constructive, and a high amplitude at the right ear will result. At other frequencies the interference will be destructive, and the amplitude at the right ear might disappear altogether. Thus for a narrow band signal the apparent position of the sound image is highly frequency dependent.

With a broad band source human hearing appears to simply average over the various possible sound directions to determine the best-guess position for the source. However, this averaging is frequency dependent – or at least it looks like it is frequency dependent to engineers with a sound level meter.

It is well known that the sensitivity of the external ear and the middle ear is not equal at all frequencies. Low frequencies are severely attenuated by the middle ear, and the pinnae system concentrates the energy in a broad peak centered at about 3kHz. Figure 3 shows the energy transfer through the external ear and the middle ear, from a paper by Moore et al, reference 3.

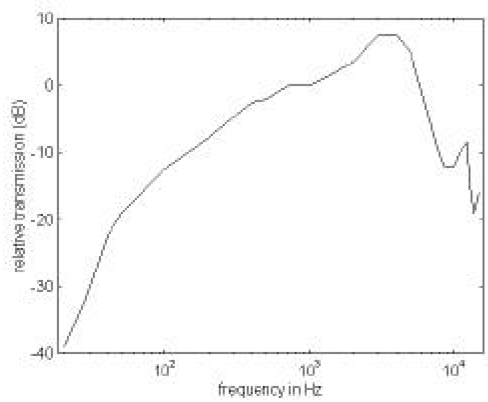

Fig. 3. Relative transmission from an external sound field through the external ear and middle ear. From [3] figure 2.

Nerve firings are thus more numerous in the band normally used for speech intelligibility – the band between 700Hz and 4kHz. When the hearing mechanism averages positions over frequency, it does so by weighting nerve firings, and thus the speech band is dominant.

The apparent position of a sound source is also strongly influenced by past history. A source which appears to be located at a particular position when it contains one mix of frequencies will remain in that position when the frequency mix is changed. For example, if we play a segment of speech with full bandwidth in a panning experiment we can determine an apparent position of the sound image. If we immediately follow with a presentation of the same speech segment that has been highly filtered, the apparent position of the sound image will be unchanged.

If we present the exact same signal from the opposite side of the listener – by simply reversing the right and left outputs of the amplifier – the apparent angle from the center is likely to be very different. By switching the side of the presentation we break the tendency to assign the signal to the previous position. Thus we found in all our experiments it is better to switch the presentation of a sound from left to right and ask the subject to estimate the width between the two images, rather than estimating the position of a single sound image.

A broadband source is perceived as being in a single horizontal position. The fact that different frequency components present in the source have different positions when they are separately analyzed does not contribute to a broadening of the source width. The hearing mechanism performs a weighted average over the incoming signal power, and assigns an apparent position to the source as a whole.

This preprint is not intended as the final word on this subject. For one reason, the number of subjects is small, consisting of the author and a few friends. Experiments with loudspeakers were conducted in the author's studio, in the demo room at Lexicon, and in an automobile. Thus the data should be considered very preliminary. However, even with this uncertainty the data has proved extremely useful in decoder design. With a corrected pan law – simply widened from a sine-cosine law for images panned between the center and left or right, the differences between the image positions in two channel sound presentation and 2-5 or 2-7 presentation has been almost entirely eliminated.

The intent of this paper is to present some of the theory of binaural hearing as it applies to sound panning, and show some of the pitfalls we encountered as we attempted to study the problem. The experiments were more complicated and more interesting than we expected when we started, and we intend to continue the work as time allows.

EXPERIMENTAL MODELS

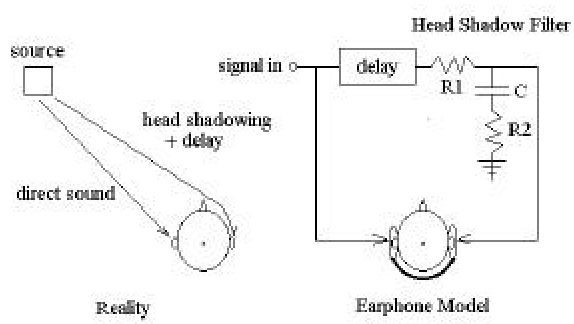

Fig. 4. We can model head diffraction and delay with a simple circuit. Here it is shown as an analog filter. Typically R1 ~= R2, and the 3dB point of the filter is about 1500Hz. The total delay (including the group delay of the filter) is about 0.5ms for a loudspeaker position of 45 degrees.

Figure three shows as simple head model, which we have been using for many years. The entire model can be constructed in the digital domain. For these experiments we used the Matlab programming language. The source codes are available on the author's web page: www.world.std.com/~griesngr. The author would appreciate email communication from anyone continuing this work with the Matlab codes.

Use of this simple model allows us to investigate the time and amplitude differences between the two ears from a geometric representation of the speaker-head system.

Problems with the model

Although this model has many advantages for calculations, there are some problems in its use. The settings for the interaural time delay and the shadow filter are obviously important. It is relatively easy to use the model to generate binaural signals that can be heard through earphones. Although listening to such signals is extremely useful for understanding how binaural hearing works, there are likely to be significant differences from natural hearing.

One of the most consistent of these differences is a widening of the positions of the sound images. When one programs the model with interaural delays which are typically found in the literature – for example about 0.5ms for an incidence angle of 45 degrees, and 0.7ms for an incidence of 90 degrees – the resulting positions of the sound images are wider than they should be in earphones.

We are making progress in understanding the reasons for this discrepancy, and the results will be presented during the oral presentation of this paper. However for this preprint we only point out that these details of the model are not particularly important. The model is to be used as an aid to understanding the hearing process, not in modeling it precisely.

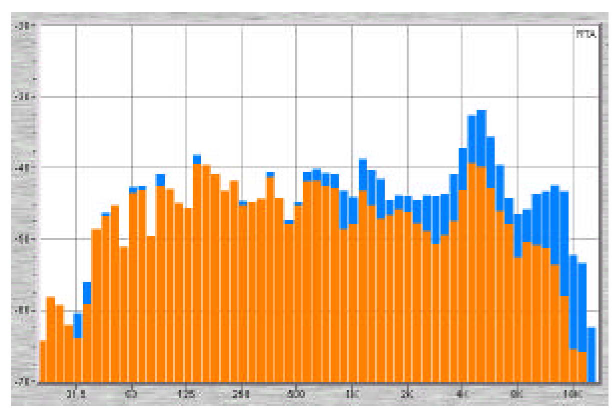

Fig. 5. Measured frequency response of a modified Neumann KU81 dummy head, right ear, 0 degrees and 45 degrees horizontal incidence. The shadowing appears to require a – 3dB point at about 1500Hz, and a shelf depth of 6 to 8dB.

Figure 5 shows the shadowing measured for a modified Neumann KU81 dummy head. We used these values for the shadowing filter. This head measures about 0.75ms of time delay at 90 degrees, and about 0.5 at 45 degrees. The interaural distance measured by the circumference around the front of the head (not including the nose) is about 12", and the interaural distance measured directly through the head is about 6". In making the model one must choose – is the interaural spacing closer to 6" or to 12"? In practice, an inteaural spacing between 15cm and 20cm seems to work adequately.

In addition, these measurements are about 20% larger than my own head, and as a consequence one would expect the positions of sound images using the dummy head measurements to sound wider than natural to me. They do.

The spectrum of the signals is also important, both for the apparent width of images in natural hearing, and in two channel panning. To get an accurate match between the width of binaural signals generated by a model, and the width observed in natural hearing, the spectrum of the two signals AT THE SURFACE OF THE EARDRUM must match precisely. Getting such a match with ordinary earphones – which sit outside the concha – is very difficult. We will leave further discussion of this issue for a later time.

According to binaural theory the ear uses the interaural time delay (ITD) to determine horizontal localization at frequencies below 1000Hz, and the interaural intensity difference (IID) at frequencies above 1000Hz. The model can be used to investigate the behavior of both the IDT and the IID as a function of frequency.

Figure 6 shows the results from using this model to investigate the effect of panning on the interaural time delay. We assumed an interaural spacing of 21cm – mostly because it made the clearest drawing! The results are plotted against a pair of axis which assume that the prediction made by the sine-cosine pan law is accurate. Thus on these axis the sine-cosine law plots as a straight line.

Note the high degree of frequency dependence of the curves. At 300Hz the ITD follows the sine law precisely. At about 600Hz, the ITD is well predicted by the sine-cosine law. Above 600Hz, the apparent image position moves rapidly toward the left loudspeaker. This happens because the sound diffracting around the head is interfering destructively with the direct sound. The total time delay around the head is approaching a half-wavelength.

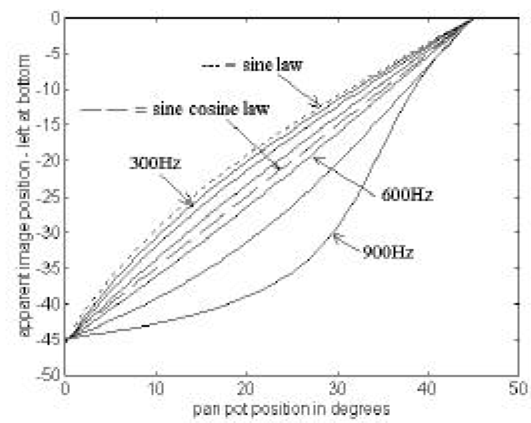

Fig. 5. ITD generated by signals panned between center and left by a sine-cosine pan law. A 21cm interaural spacing is assumed.

At higher frequencies we expect the interaural level differences to become more important. As mentioned in the introduction, these level differences are also highly frequency dependent. Once again whenever the interaural delay is sufficient to create destructive interference at the entrance to the ear canal, the apparent width of the image increases.

To investigate this effect the model was used to generate a test signal consisting of a speech segment filtered into 1/3 octave bands. This signal was then presented to the subject panned to the middle between center and left by a sine-cosine pan pot. The panned signal was then alternated from the left side to the right side and back, and the subject was asked to estimate the angle between the two images.

As a reference signal the broadband panned signal was also presented. Thus the entire test signal consisted of a broadband panned speech signal on the left, broadband panned speech on the right, narrow band signal on the left, and finally narrow band signal on the right. This signal was continued repeatedly until the subject made an estimate of how much wider or narrower was the distance between the narrow band signal than the reference.

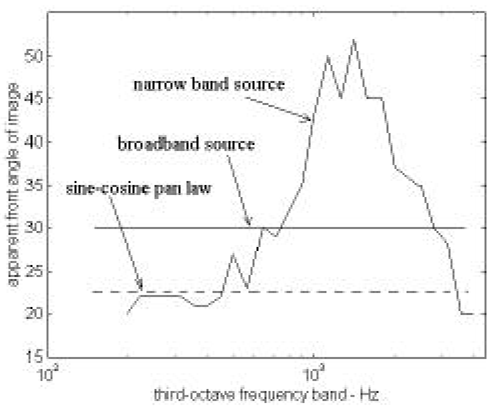

Fig. 7. The apparent position of a 1/3 octave speech band as a function of frequency (assuming a 15cm interaural spacing). Note the broadband apparent position is much wider than would be predicted by the sine-cosine law.

The particular speech segment was from the "Music for Archemedes" test CD produced by Soren Bech. It is a segment of female speech "In language, infinitely many words can be written with a small set of letters." The energy spectrum of the broadband speech spectrum is strongly weighted toward frequencies below 1000Hz. The energy weighting of the source combines with the transfer of energy shown in Fig. 1, and the hearing mechanism assigns he position of the broadband source by averaging the resulting narrow band positions, weighted by the amount of neural activity in each band.

The model was also used to generate signals identical to the ones used in the experiment above, but Matlab was asked to amplify the left and right signals independently so they had exactly the same RMS sound pressure.

The results from this experiment will be presented in more detail during the oral presentation. In brief, they showed that without the amplitude information the panned images were much narrower. When the same experiment was repeated using the same model, but this time with a hard sound source (not a panned source), it was found that the apparent position of the source moved about half-way from the low frequency position toward the center as the frequency was raised. In other words, the ear uses the ITD of the envelope of the signal as a direction cue at high frequencies, but combines it with amplitude differences to determine the final direction.

CONCLUSIONS

The apparent position of amplitude panned sound sources consisting of broadband speech or music is consistently wider than would be predicted by the sine-cosine pan law. The apparent position of the image appears to be determined from an average of narrow band position measurements, weighted by the power spectrum of the incoming signal and the transfer function of the external and middle ear. The narrow band direction detection is highly frequency dependent, due to interference between the direct sound from one speaker and the sound diffracting around the head from the other speaker. Due to the power weighting of the external and middle ears, the apparent position of energy in the speech range (700Hz to 4kHz) dominates the result. This frequency band is in the middle of strong interference between the direct sound and sound that diffracts around the head, and for this reason The positions of broad band speech and music images are consistently wider than would be expected from pan data measured at both lower and higher frequencies.

References

[1] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources, Part1: Stereophonic Panning," *J. Audio Eng. Soc.*, vol. 49, pp 739, 752 (2001).
[2] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources, Part2: Two and Three dimensional Panning," *J. Audio Eng. Soc*., vol. 49, pp 753, 767 (2001).
[3] B. C. J. Moore, B. R. Glasberg and T. Baer, "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.*, vol. 45, pp. 224-240 (1997).