CONTINUOUS SURROUND PANNING FOR 5-SPEAKER REPRODUCTION

PETER G. CRAVEN

consultant to Meridian Audio Ltd., Huntingdon, UK.
pgcraven@algol.co.uk

We have constructed 5-speaker panning laws in which each speaker feed is a sum of circular harmonics. The coefficients of the harmonics have been optimized numerically to give best performance on a particular loudspeaker layout according to various psychoacoustic criteria. A panning law using fourth order circular harmonics has been auditioned on the ITU layout and judged superior to conventional pairwise panning in several respects.

INTRODUCTION

Panning laws have been a subject of much discussion in the literature. In two speaker stereo, amplitude panning with a "constant power" law is the norm, and indeed there is little that can be done to improve it other than possibly to use a "shuffler" to compensate for the fact that high frequency sounds tend to pull towards the speakers more than low frequency sounds.

With three frontal speakers, Gerzon [1] and others (e.g. [2]) have explored ways of panning a sound continuously across centre-stage, so that it does not "fall into" the centre speaker. Gerzon's work attempts to satisfy two or more localisation theories simultaneously, so that the various cues reinforce each other, resulting in more stable images and less listening fatigue.

In surround sound, the possibilities are so numerous that many have reverted to simple pairwise amplitude panning between adjacent speakers. West [3], in a master's thesis, analysed the principal panning laws known at the time, and conducted listening comparisons. West's "*Five-channel 'optimal' implementation*" constructs a surround panning law by feeding frontal sources to the front three speakers, using a Gerzon law designed for three frontal speakers, and by using a Gerzon four-speaker law to cover the rest of the stage using the L_f, L_s, R_s and R_f speakers. However this approach is essentially discontinuous between the front stage and the sides, and will result in sources at the angles of the L_f and R_f speakers "falling into" those speakers.

In this work we have derived panning laws that cover the full 360° horizontal stage in a continuous manner. Both the ITU speaker layout and non-ITU layouts have been considered, though in this paper we have concentrated on a particular law designed for the ITU layout. The aims of the work have been to provide:

- The best image quality possible for each individual panned direction
- Consistency as a source is moved round the listener
- Support for speaker re-mapping and hierarchical [4] transmission possibilities.

1 A CONTINUOUS PANNING LAW

We start by presenting the principal panning law that is discussed in this paper. This law is designed for the ITU layout having speakers at angles of 0°, ±30° and ±110°. The five feeds are plotted in Figure 1, and the equations are given in the appendix.

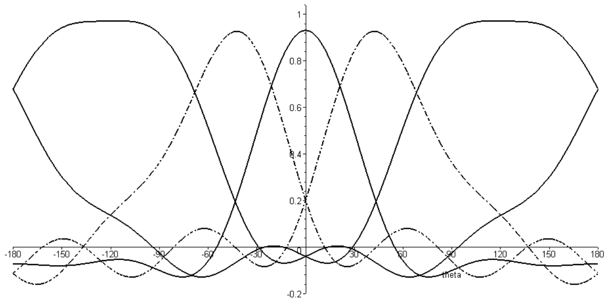

Figure 1: Feeds for each the five speakers L_s, L_f, C, R_f, R_s as a function of panning angle θ.

2 PSYCHOACOUSTIC CRITERIA

2.1 Fundamental principles

The ear's horizontal directional acuity is, under favorable conditions, of order one degree of arc. Since the installation of 360 loudspeakers seems to be domestically unacceptable, we are forced to consider phantom images if the consumer is to experience sound from all round.

We design phantom images by proposing a model for the response of human hearing, and attempting to adjust the feeds to a number of loudspeakers to produce a response "as close as possible" to the response from a real source in the intended direction.

For evaluation of a system, we may well use a psychoacoustic model that takes into account Head Related Transfer Functions (HRTFs) and simulates other neural auditory processing (c.f. [8]). At present these models are too complex for use as design tools, and we seek something simpler.

Although exceedingly complex, the ear's processing is constrained by the laws of Information Theory. Gerzon sought to cut through complexity and indeed ignorance by focusing instead on what information is available for the ear to process. In 1977 he circulated a document that was published fifteen years later as a "Metatheory" [9] and that provides a rational way at looking at these matters. However, this document is not for the mathematically faint-hearted.

Gerzon's document describes a hierarchy (actually a double hierarchy) of models. If one were to adopt a sufficiently high order model one would have an essentially complete description of the sound field in the vicinity of the listener's head. However such a high order model would again be too complicated for practical design work. The argument then is whether the low-order models from Gerzon's metatheory are successful in abstracting the quantities that are practically relevant.

2.2 Pressure, velocity and energy

We shall use "Pressure", "Velocity vector" and "Energy vector" as the main parameters to be monitored in matching a panning law. These are low-order parameters from the Gerzon metatheory referred to above, but they also underpin many of the ideas in common circulation, such as the "law of sines" [6], that have nothing to do with Gerzon's work.

Gerzon provides a readable description of these parameters in [1]: below we shall attempt to explain them with minimal recourse to mathematics.

When a source emits a sound, the physical parameters that characterize the wave are pressure and velocity. Pressure is a scalar quantity, but velocity is a vector quantity.

For a single source of sound, the pressure and the magnitude of velocity are always in the same ratio, given by the characteristic impedance of air.

A single loudspeaker can reproduce the same pressure, at the listener's ear, as a source, regardless of direction. However the velocity will be correct only if the loudspeaker is in the same direction as the intended source.

When we use two loudspeakers, we can simulate the velocity produced by any source between the loudspeakers by suitably adjusting the amplitudes of the feeds to the two loudspeakers. Makita [10] proposed that, at low frequencies, the perceived direction of a sound is the direction of velocity that it produces.[1] It is this velocity theory that leads to the law of sines [6] for perceived angles.

When we drive two loudspeakers equally, the pressure at a listener (assumed to be at the exact same distance from each) increases by +6dB, but the velocity increases by less than 6dB if the speakers are in different directions. Indeed, if the speakers are at 180º from each other, the velocities will cancel.

Thus in general the reproduced velocity is deficient relative to the velocity that would have been obtained from a real source delivering the same pressure.

Gerzon defines the *velocity vector* that results from driving several loudspeakers, as the vector that has:

- A direction θ_v that is the direction of the air motion
- A length r_v that is the *ratio* between the magnitude of the air velocity, and the velocity that would have been expected from a single source producing the same pressure.

For a single real source, the velocity vector length r_v will be unity, and we try to reproduce this situation when rendering to loudspeakers. In practice r_v tends to be less than unity, though it is possible to produce $r_v>1$ by driving loudspeakers out-of-phase.

In the velocity theory, we assume that each loudspeaker contributes a velocity, and that these velocities add as vectors. However the situation becomes extremely complicated, as shown in Poletti's plots [5], when there are path length differences that are significant compared to a wavelength.

For higher frequencies therefore we tend to concentrate on energy theories, which relate to the *incoherent* addition of sounds from the loudspeakers in an array.

The energy radiated by a loudspeaker is always positive, being proportional to the square of the drive. Energy is a scalar quantity so at the listener the total sound intensity is obtained by simple summation of the energy from each speaker.

However, the ear does have some sensitivity to the direction from which energy arrives. Consider low-mid frequencies, at which head shadowing is starting to have an effect, so that each HRTF is starting to deviate from the omnidirectional form that it has at low frequencies. Suppose we can model the each HRTF approximately as a cardioid or a subcardioid pointing due left or due right. If we model the ear's "interaural amplitude difference" localization processing as comprising a square-law detector on the output of each ear, followed

[1] While an individual ear is, at low frequencies, essentially a pressure transducer, it is assumed that the listener is able to sense sideways velocity by subtracting the pressure sensed at his two ears. The difference will represent a phase shift between the two ears, and this mechanism is known as "interaural time difference" localization.

by a subtraction, that will give a signal that, for each source, is proportional to the sine of its angle relative to the median plane. By head rotation we can obtain a similar measure in the orthogonal direction, and putting the two measures together we can synthesise a vector quantity that is related to energy. The *energy vector* is obtained by normalizing this vector quantity with respect to the (scalar) sound intensity at the listener.

The energy vector has:

- A direction θ_e that is a broadband measure of the direction from which most of the energy is arriving
- A length r_e that indicates the extent to which the energy is concentrated in one direction

For a single real source, the energy vector length is unity. For an isotropic diffuse field, the energy vector length is zero. For several separated sources or loudspeakers, the length is always less than unity—there is no possibility of increasing its length by using out-of-phase sources as there is with the velocity vector.

Please see section 4 for further discussion of the psychoacoustic significance of r_v and r_e.

3 PAIRWISE PANNING

The best known panning law is pairwise panning. In pairwise panning, the soundstage is considered as divided into segments, each segment being bounded by the two neighboring speakers. Within a segment, the sound is placed by adjusting the amplitudes fed to these two speakers, the others being silent.

The speaker feeds for a Pairwise Constant Power Panning (PCPP) law with speakers in the ITU layout, are illustrated in Figure 2. The horizontal axis is the desired panning angle θ in the range –180º to +180º. It will be seen that when θ is coincident with one of the speaker positions of –110º, –30º, 0º, +30º and +30º, just that one speaker is "illuminated", and with unit amplitude. At other positions two adjacent speakers are illuminated, with amplitudes adjusted for the same total power.

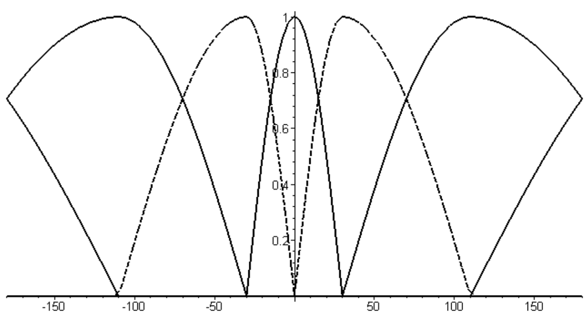

Figure 2: Speaker feeds for Pairwise Constant Power Panning (PCPP).

The imaging mechanism for an individual source is similar in PCPP to conventional stereo. We have been living with that domestically for nearly half a century, and we know that 2-speaker stereo imaging can be somewhat fragile. The problems get worse with large speaker angles and when the included segment is elsewhere than directly in front of the listener.

In terms of the psychoacoustic parameters introduced above, we can identify three non-idealities:

- R_v is less than unity
- R_e is less than unity
- θ_e is different from θ_v except at points of symmetry.

4 SIGNIFICANCE OF R_V AND R_E

If the two speakers are at angles ±φ from their line of symmetry, and are driven equally to produce an image on the line of symmetry, then r_v=r_e=cos(φ). For the angles relevant to the ITU speaker layout, the values are shown in Table 1.

Segment	φ	$r_v = r_e = \cos(\varphi)$
Front	15º	.9659
Side	40º	.7660
Rear	70º	.3418

Table 1: Half angle φ between the bounding speakers, and cos(φ), for each segment of the ITU 5-speaker layout.

For normal two-channel stereo with speakers 60º apart, we have φ=30º, cos(φ) = .866. Thus for a central image, r_e =r_v =.866. In terms of deviation from unity, the r_v and r_e from pairwise panning in the ITU layout are nearly four times as good at the front, nearly twice as bad at the sides, and about five times as bad at the rear.

The effect of having r_e significantly lower than unity is that the image is not as stable with respect to listener movement.[2]

The effect of having r_v lower than unity is, *inter alia*, that the image position changes with head rotation. Gerzon and Barton [7] consider that for frontal reproduction, a relatively wide tolerance such as $0.8 \le r_v \le 1.2$ is acceptable, but that for side imaging the relationship $r_v \approx 1$ is important.

In this paper we are not considering frequency-dependent panning laws. If we were to do so, it would be more important to optimize r_v at low frequencies and r_e at higher frequencies [7].

[2] In contexts other than pairwise panning, a low value of r_e can indicate significant "crosstalk" to a speaker far removed from the desired source direction. In this case there may be the possibility of front-rear image flips. It is wise to monitor crosstalk figures, especially front-to-rear crosstalk, as well as r_e.

Pairwise panning delivers the largest possible r_e. For a given speaker layout and image direction, r_e can only be made worse if one departs from pairwise panning.

R_v however can be considerably improved, generally by introducing *small* out-of-phase signals to speakers remote from the intended source direction. Because energy is proportional to the square of the drive, a small drive to these remote speakers results in a very small energy emitted from them. Therefore, a significant improvement in r_v can be obtained for very little reduction in r_e.

5 COMPUTATIONAL METHOD

Each speaker *sp* has a feed $feed_{sp}$ that depends on the desired panning angle θ. We have required this dependency to be expressible as a sum of circular harmonics:

$$feed_{sp} = \sum_{m=0}^{n} (\alpha_{sp,m} \cos(m\theta) + \beta_{sp,m} \sin(m\theta))$$

In this work, the maximum order n of the harmonics has been set variously between 1 and 11. When n=11, the angular resolution provided by the above Fourier series is so high that there is hardly any loss of generality, i.e. any reasonable continuous panning law can be expressed in the above form with negligible loss of accuracy.

The general approach is to combine various psychoacoustic criteria into a *penalty function*, and to use a numerical nonlinear optimisation method to adjust the α and β coefficients so as to minimise the penalty function. The psychoacoustic criteria we have used were:

- Reproduced energy should be substantially independent of panning angle.
- The velocity and energy vector directions θ_v and θ_e should be closely matched (c.f. [1])
- The angles θ_v and θ_e should be reasonably close to the panning angle θ.
- Velocity vector length r_v should be close to unity
- Energy vector length r_e should be as large as possible

We have tried various values of the maximum order n. With a small value such as n=4, there is some conflict between optimising the image quality in different directions. Frontal image quality has been given the highest weight, side images the next highest, and rear image quality the lowest.

Even with the small value n=4, there are 45 α and β coefficients to be adjusted in order to minimize the penalty function. This number can be reduced to 23 independent coefficients by taking advantage of left-right symmetry.

A conjugate-gradient method was used to solve the nonlinear optimization problem, with final convergence accelerated using second derivatives in a Newton iteration.

Sometimes nonlinear optimization can produce a result that depends on the starting-point, a result of the algorithm “getting stuck” in a local minimum that is not necessarily the global minimum. With general speaker layouts, no severe evidence of this type of behavior was observed, but that with the ITU layout, there is evidence of an alternative solution that has virtually no output from the C loudspeaker. With very small values of n, it seems that this essentially four-speaker solution gives a better result than driving all five loudspeakers.

6 OBJECTIVE EVALUATION

An objective comparison between the panning law given in the appendix, and PCPP, will be found in appendix B of the paper by Neher et. al. [11] published at this conference. The plots confirm that r_v has been significantly improved over PCPP for most panning angles, in particular at θ=±90º and at θ=180º. The directional error $|\theta_v - \theta_e|$ has also been greatly reduced over the front stage.

A trouble-spot occurs around θ=50º, where the directional error $|\theta_v - \theta_e|$ is substantial with both the new law and with PCPP. This problem can be reduced somewhat using higher values of n, but it seems difficult to eradicate it completely except at the expense of substantial worsening of the r_v performance. It seems to be a problem associated with the narrow angles of the front speakers in the ITU layout.

7 SUBJECTIVE EVALUATION

Members of the University of Surrey's Institute of Sound Recording conducted an informal listening session and kindly provided the following comments on the new panning law:

> *Panning a sound across the front gives a smooth transition. It's clearly better compared to constant power panning, both timbrally and spatially. There's hardly any change in quality as a sound passes through C and only a little bit more as a sound is panned beyond L and R. The sound image remains stable for about ±40 to 45 degrees (perceived angle).*
>
> *Sides images … certainly aren't worse than pairwise amplitude panning, but problematic in another way. As the image is panned from front to back or vice versa, it jumps easily between the speakers and it can broaden considerably (especially if you move your head).*
>
> *[At 180 degrees] there is less spread at LF … compared to pairwise panning, i.e. the image appears to be more focused.*

These comments suggest that the aim of obtaining continuity as sounds are panned between speakers has largely been achieved, at least for frontal sounds.

It is interesting that the region of perceived good performance extends beyond the stage spanned by the three front speakers. Whether the quoted 40º–45º limit for "stable" central images is correlated with the objectively poorer directional error in the region 45º–50º, is a matter for speculation.

The favorable remark on LF performance at 180° may well be explained psychoacoustically by the velocity vector length r_v = 0.693 at the rear, compared to r_v = 0.342 as given by pairwise panning. The improvement in r_v is achieved by feeding a small out-of-phase signal to the front speakers when sounds are panned to the rear.

8 CONCLUSIONS

We have shown that it is possible to derive a panning law that is functionally continuous as the panning angle rotates through 360º. The law presented here uses fourth order circular harmonics and is optimized for the ITU speaker layout.

The work reported here is mathematically based and the law is optimized on the assumption that certain psychoacoustic criteria are relevant. Initial subjective results are encouraging, nevertheless considerably more subjective testing of this and other laws would be desirable in order to refine the psychoacoustic model.

The work is ongoing and it is hoped to report in a subsequent paper on laws that use various orders of harmonics and that are optimized for speaker layouts that deviate from the ITU standard.

Readers who would like to perform subjective evaluations are invited to contact the author at the e-mail address given on the title page.

9 ACKNOWLEDGEMENTS

The author is grateful to Meridian Audio Ltd. for supporting this work and for permission to publish. Discussions with Chris Travis, Dave Malham and Tobias Neher are gratefully acknowledged.

REFERENCES

[1] Gerzon, M.A., "Panpot Laws for Multispeaker Stereo", AES 92nd convention, March 1992, Vienna: AES preprint no. 3309.

[2] Miles, M. T., "An Optimum Linear-Matrix Stereo Imaging System", AES 101st convention, Nov. 1996, Los Angeles: AES preprint no. 4364.

[3] est, J.R., "Five-Channel Panning Laws: An Analytical and Experimental Comparison" Master's Thesis, Music Engineering, University of Miami, 1998. Available at: *http://www.music.miami.edu/students/personal/jwest/master.html*

[4] Craven, P.G., Law, M.J., Stuart, J.R. and Wilson, R.J, "Hierarchical Lossless Transmission of Surround Sound using MLP", AES 24 th International Conference on Multichannel Audio, Banff, June 2003.

[5] Poletti, M., "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems", J.A.E.S.: Volume 44 Number 11 pp. 948-963; November 1996

[6] Bennett, J. C.; Barker, K.; Edeko, F. O., "A New Approach to the Assessment of Stereophonic Sound System Performance", J.A.E.S. Volume 33 Number 5 pp. 314·321; May 1985

[7] Gerzon, M.A. and Barton, G,J, "Ambisonic Decoders for HDTV", AES 92nd convention, March 1992, Vienna: AES preprint no. 3345.

[8] Wiggins, B., Paterson-Stephens, I., Schillebeeckx, P., "The Analysis of multi-channel sound reproduction algorithms using HRTF data", AES 19th International Conference on Surround Sound, Schloss Elmau, June 2001.

[9] Gerzon, M.A., "General Metatheory of Auditory Localisation" Preprint 3306; Convention 92; March 1992

[10] Makita, Y., "On the directional Localization of Sound in the Stereophonic Sound Field", E.B.U. Review, Part A no, 73, pp. 102–108 (1962)

[11] Neher, T, et. al., "Simulation of Ensemble Depth" Part 1 AES 24th International Conference on Multichannel Audio, Banff, June 2003.

APPENDIX

Here are the equations for the fourth order panner whose feeds are plotted in Figure 1. As one might expect, there is left/right symmetry of the cosine terms, and antisymmetry of the sine terms.

$$feed_{Rb} = 0.35579 - 0.35965\cos(\theta) - 0.42548\sin(\theta) - 0.06361\cos(2\,\theta) + 0.11778\sin(2\,\theta) + 0.00012\cos(3\,\theta) + 0.04692\sin(3\,\theta) + 0.02722\cos(4\,\theta) + 0.06146\sin(4\,\theta)$$

$$feed_{Rf} = 0.16656 + 0.24162\cos(\theta) - 0.27215\sin(\theta) - 0.05322\cos(2\,\theta) - 0.22189\sin(2\,\theta) - 0.08418\cos(3\,\theta) - 0.05939\sin(3\,\theta) - 0.06994\cos(4\,\theta) - 0.08435\sin(4\,\theta)$$

$$feed_{C} = 0.10492 + 0.33223\cos(\theta) + 0.26500\cos(2\,\theta) + 0.16902\cos(3\,\theta) + 0.05978\cos(4\,\theta)$$

$$feed_{Lf} = 0.16656 + 0.24162\cos(\theta) + 0.27215\sin(\theta) - 0.05322\cos(2\,\theta) + 0.22189\sin(2\,\theta) - 0.08418\cos(3\,\theta) + 0.05939\sin(3\,\theta) - 0.06994\cos(4\,\theta) + 0.08435\sin(4\,\theta)$$

$$feed_{Lb} = 0.35579 - 0.35965\cos(\theta) + 0.42548\sin(\theta) - 0.06361\cos(2\,\theta) - 0.11778\sin(2\,\theta) + 0.00012\cos(3\,\theta) - 0.04692\sin(3\,\theta) + 0.02722\cos(4\,\theta) - 0.06146\sin(4\,\theta)$$

The same information is presented below in a form that may be more useful to those who wish to paste the coefficients from an electronic copy of this paper.

```
feed[C] = .10492+.33223*cos(theta)+.26500*cos(2*theta)
        +.16902*cos(3*theta)+.5978e-1*cos(4*theta),
feed[Lf] = .16656+.24162*cos(theta)+.27215*sin(theta)
        -.5322e-1*cos(2*theta)+.22189*sin(2*theta)
        -.8418e-1*cos(3*theta)+.5939e-1*sin(3*theta)
        -.6994e-1*cos(4*theta)+.8435e-1*sin(4*theta),
feed[Rf] = .16656+.24162*cos(theta)-.27215*sin(theta)
        -.5322e-1*cos(2*theta)-.22189*sin(2*theta)
        -.8418e-1*cos(3*theta)-.5939e-1*sin(3*theta)
        -.6994e-1*cos(4*theta)-.8435e-1*sin(4*theta)
feed[Lb] = .35579-.35965*cos(theta)+.42548*sin(theta)
        -.6361e-1*cos(2*theta)-.11778*sin(2*theta)
        +.12e-3*cos(3*theta)-.4692e-1*sin(3*theta)
        +.2722e-1*cos(4*theta)-.6146e-1*sin(4*theta),
feed[Rb] = .35579-.35965*cos(theta)-.42548*sin(theta)
        -.6361e-1*cos(2*theta)+.11778*sin(2*theta)
        +.12e-3*cos(3*theta)+.4692e-1*sin(3*theta)
        +.2722e-1*cos(4*theta)+.6146e-1*sin(4*theta).
```