Proceedings of the Institute of Acoustics

IMPLEMENTATION OF DYNAMIC PANNING REPRODUCTION WITH ADAPTATION FOR HEAD ROTATION

Dylan Menzies [1] Institute of Sound and Vibration Research, University of Southampton
Filippo Maria Fazi Institute of Sound and Vibration Research, University of Southampton

ABSTRACT

Panning is a technique for spatial reproduction in which each source signal is weighted by gains and fed to the loudspeakers, producing a phantom image. The image is unstable to head rotation, that is the image appears to move even though the gains are fixed. This causes instability in the overall scene. We have shown how this error can be corrected. Further work has shown how the same methods can be used to create near-field images. These results are reviewed here. A working implementation is also described, along with an informal subjective assessment.

1 INTRODUCTION

When a listener's head rotates in the sound field produced by a panning system, images produced using static gains are not fixed, leading to instability in the overall image, for example see Fig. 1. In this example an image is produced by mixing a signal with equal gains into two loudspeakers. When the listener faces forward the image is in the same direction, but as the listener turns away from this direction the image direction shifts towards the listener.

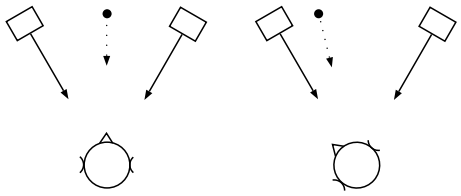

Figure 1: The black dot indicates the perceived direction of the image when two loudspeakers each have the same signal, for different head directions.

The effect is particularly evident in sound with low frequency content $< 1000Hz$. Higher frequencies are subject to different perceptual processes. The focus here is exclusively on low frequencies. Low frequency cues tend to dominate high frequency cues in composite sounds[1], it is expected that the method will work to some extent with broadband signals without further processing.

The panning stability problem was addressed by first calculating the localisation cues produced by a general low frequency sound field[2,3]. The Interaural Time Difference (ITD) and Interaural Level Difference (ILD) were found from the first order field approximation using free-field and spherical head scattering models. A simple formula was found relating the head orientation, image direction, and the sound field description given as a specific admittance vector, or equivalently as a Makita localisation vector. Dynamic stereo panning gains were then found that depend on the head orientation as well as image direction. Good agreement was found with experiment. Furthermore, allowing complex gains

[1] d.menzies@soton.ac.uk

gives the possibility of controlling ILD as well as ITD, leading to the creation of near-field images. An efficient formulation was found for this[4].

In the following sections this work is reviewed in more detail, including the related problem of listener position adaptation. An implementation is described, and some informal assessment given.

2 Sound field representation

A source free region of a sound field can be expanded as a series about any point in the region[5]. The first order approximation of the pressure P at a point $\boldsymbol{x}$, expanded about point $\boldsymbol{x_0}$, can be given in the frequency domain in terms of pressure and pressure gradient by

$$P(\boldsymbol{x}) \approx P(\boldsymbol{x_0}) + \nabla P(\boldsymbol{x_0}) \cdot (\boldsymbol{x} - \boldsymbol{x_0}) \tag{1}$$

The approximation is good provided the wavelength is considerably larger than $|\boldsymbol{x} - \boldsymbol{x_0}|$, and higher order derivative terms are small compared with $P(\boldsymbol{x_0})$, $\nabla P(\boldsymbol{x_0})$, which is usually the case when a source is not close. Below 700Hz a typical sound field region large enough to enclose the human head satisfies these conditions.

Although a listener's head is described as acoustically transparent at low frequencies, the binaural signals are not equal to the pressures at the corresponding locations in the incident field, due to the non-vanishing scattered field. Using a spherical model for the head, with ears at antipodal locations, a simple analytical approximation can be found for the binaural signals[5]. The free field ITD is $\frac{D}{c}\cos\theta$ where D is the ear separation, c is the sped of sound, and θ is the angle of the incident plane wave to the inter aural axis. The sphere-scattered ITD is found to have a limit $\frac{3D}{2c}\cos\theta$ at low frequency, and constant ILD of 1 (or 0 in dB)[6] . Furthermore in this approximation the resultant pressure field on the sphere, radius r, is equal to the incident plane wave field at a radius $\frac{3}{2}r$. Since any free field region can be approximated arbitrarily well by a plane wave expansion[7], this implies that for any incident field, including near source fields, the resultant pressure at the surface of the sphere is given by the incident field evaluated at $\frac{3}{2}r$, for low frequency.

With the above approximations the binaural signals at the right and left ears are given by

$$P_R = P + \bar{\boldsymbol{r}}_R \cdot \nabla P \tag{2}$$

$$P_L = P + \bar{\boldsymbol{r}}_L \cdot \nabla P \ , \tag{3}$$

where P and ∇P are the pressure and gradient at the central point between the ears and $\bar{\boldsymbol{r}}_R, \bar{\boldsymbol{r}}_L$ are related to the vectors from the centre to the right and left ears, $\boldsymbol{r}_R, \boldsymbol{r}_L$, by $\bar{\boldsymbol{r}}_R = \frac{3}{2}\boldsymbol{r}_R, \bar{\boldsymbol{r}}_L = \frac{3}{2}\boldsymbol{r}_L$. See Fig. 2.

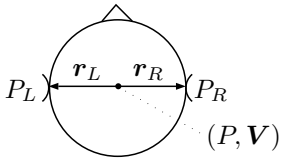

Figure 2: Sound field variables and vectors in relation to the listener's head.

The gradient is related to particle velocity $\boldsymbol{V}$ by Euler's equation in the frequency domain[8],

$$\nabla P = -jkZ_0\boldsymbol{V} \ , \tag{4}$$

using the positive frequency convention $P(\boldsymbol{x},t) = P(\boldsymbol{x})e^{j\omega t}$. Z_0 is the characteristic impedance and k is the wavenumber.

In the following discussion only relative phases of the binaural signals are of interest, so in order to simplify calculation the pressure and velocity are redefined by applying a phase rotation so that the pressure is positive real,

$$(P,\boldsymbol{V}) \rightarrow (P',\boldsymbol{V}') = (P,\boldsymbol{V})P^*/|P| = (|P|,\boldsymbol{V}_\Re + j\boldsymbol{V}_\Im) \tag{5}$$

with the phase rotated velocity, $\boldsymbol{V}'$, split into real and imaginary parts $\boldsymbol{V}_\Re, \boldsymbol{V}_\Im$.

Proceeding with phase rotated values of pressure and velocity and the binaural signals, with primes dropped, $P, P_L, P_R, \boldsymbol{V}_\Im, \boldsymbol{V}_\Re$, and using $\boldsymbol{r}_L = -\boldsymbol{r}_R$, the binaural signals can be written

$$P_R = P + kZ_0(\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Im - j\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Re) \tag{6}$$

$$P_L = P - kZ_0(\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Im - j\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Re) \tag{7}$$

The low frequency approximation condition is $kr_R << 1$.

Fig. 3 illustrates the general case using the complex plane. Both $\boldsymbol{V}_\Im$ and $\boldsymbol{V}_\Re$ are non-zero, and in different directions. As the listener's head rotates around a vertical axis P_R and P_L move around on opposite sides of an ellipse, because the terms $\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Im$ and $\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Re$ each vary as a sinusoidal function of angle, with a relative phase offset.

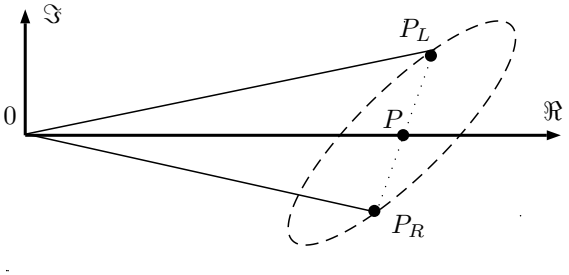

Figure 3: P_R and P_L in the complex plane for non zero and non aligned $\boldsymbol{V}_\Re$ and $\boldsymbol{V}_\Im$

3 Localisation cues and reproduction

ITD, also known as Interaural Phase Delay for a harmonic field[9], is equal to ϕ_{RL}/ω, where $\phi_{RL} = \arg(P_R/P_L)$ is the Interaural Phase Difference. ILD is given by $|P_R/P_L|$. Using the approximations (6,7) the ITD and ILD can be found from the pressure P and velocity $\boldsymbol{V}$.

Consider the problem of finding a reproduction field that produces the same perceived image as a plane wave target field. This is equivalent to finding a reproduction field with the same ITD and ILD as the target field, discounting other localisation cues. A plane wave reproduction field is a trivial solution. Other solutions are provided by matching pressure and velocity, since ITD and ILD depend only on pressure and velocity in the low frequency approximation. The reproduction field is not dependent on head orientation. The Ambisonic reproduction method employs this principle[10]. In the following more general solutions are identified, which are not generally independent of head orientation.

For the plane wave target the ILD is 0dB and

$$\boldsymbol{V}_\Re = \boldsymbol{V} = \frac{P}{Z_0}\hat{\boldsymbol{V}}, \;\; \boldsymbol{V}_\Im = 0 \tag{8}$$

where $\hat{\boldsymbol{V}}$ is the velocity direction unit vector. The corresponding diagram is shown in Fig. 4. The image direction unit vector is written $\hat{\boldsymbol{r}}_I = -\hat{\boldsymbol{V}}$.

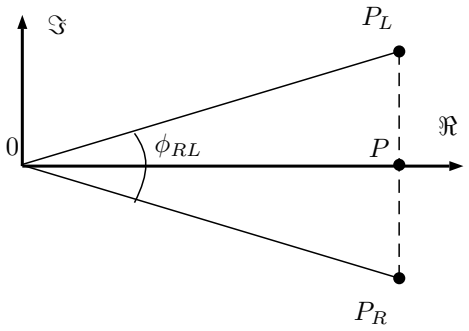

Figure 4: P_R and P_L in the complex plane for $\boldsymbol{V}_\Re > 0$ and $\boldsymbol{V}_\Im = 0$

Now consider any reproduction field such that $\boldsymbol{V}_\Im = 0$ in order to satisfy ILD = 0dB. Phase coherent panning is an example field of this type. The field is described by a diagram like Fig. (4). The ITD of the reproduced field will match the ITD of the target field only if shape of the respective triangles $0 \to P_L \to P_R$ are the same. This is the case if the ratio $|P_L - P|/P$ is the same for target and reproduction fields. A condition can then be derived[2] relating variables describing the reproduction field $P, \boldsymbol{V}_\Re$, the head orientation $\hat{\boldsymbol{r}}_{\boldsymbol{R}}$, and the target image direction $\hat{\boldsymbol{r}}_I$,

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\hat{\boldsymbol{r}}_I - \boldsymbol{r}_V) = 0 \tag{9}$$

where $\boldsymbol{r}_V = -\boldsymbol{V}_\Re Z_0/P$ is defined for convenience. For a plane wave field $\boldsymbol{r}_V$ coincides with the direction to the source. $P/V_\Re$ can be viewed as the local specific impedance of the field. (9) can be visualised as a vector diagram, which is shown here as a projection from above the listener.

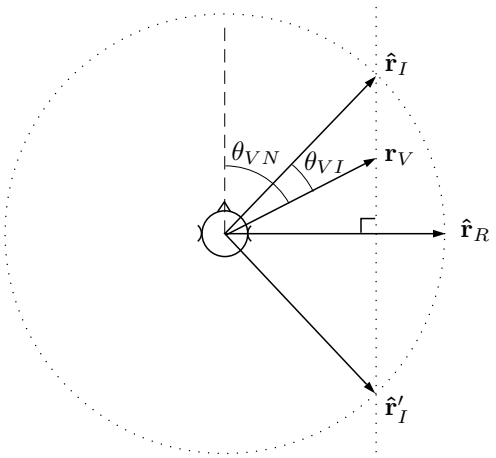

Figure 5: Geometric relationship between vectors $\hat{\boldsymbol{r}}_{\boldsymbol{R}}, \hat{\boldsymbol{r}}_{\boldsymbol{I}}, \boldsymbol{r}_V$

In Fig. 5 shows an example where the listener is not facing the image ($\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot \hat{\boldsymbol{r}}_I \neq 0$). There is a continuum of possible reproduction fields $\boldsymbol{r}_{\boldsymbol{V}}$ satisfying (9), which lie on a plane normal to $\boldsymbol{r}_R$, represented by the dotted line shown. $\boldsymbol{r}_{\boldsymbol{V}}$ can have any length, lying inside or outside the unit circle. Conversely given any field $\boldsymbol{r}_{\boldsymbol{V}}$, a consistent image can lie anywhere on a cone about $\hat{\boldsymbol{r}}_{\boldsymbol{R}}$, the *cone of confusion*. High frequency cues and dynamic cues further restrict the image. Fig. 6 shows the case with the listener facing the image. $\boldsymbol{r}_V$ is then constrained to the direction $\boldsymbol{r}_I$ but can have any length.

Restricting $\hat{\boldsymbol{r}}_{\boldsymbol{R}}, \hat{\boldsymbol{r}}_{\boldsymbol{I}}, \boldsymbol{r}_{\boldsymbol{V}}$ to the horizontal plane, (9) can be used to find the effect of different values of r_V on image direction[2,3]. In Fig. 7 the angle θ_{VI} between field vector $\boldsymbol{r}_{\boldsymbol{V}}$ and the image direction $\hat{\boldsymbol{r}}_{\boldsymbol{I}}$, is plotted against the angle θ_{VN} between the forward head direction and the field. Each plot has constant

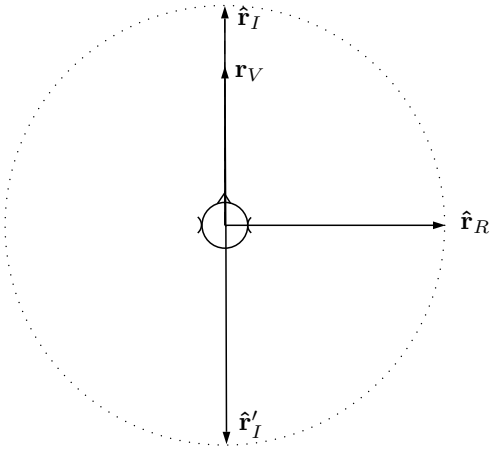

Figure 6: Geometric relationship between vectors $\hat{\boldsymbol{r}}_{\boldsymbol{R}}, \hat{\boldsymbol{r}}_{\boldsymbol{I}}, \boldsymbol{r}_V$

$r_V = |\boldsymbol{r}_V|$. θ_S will be explained in the next section. For $r_V = 1.0$ the pressure and velocity match that of a plane wave field, and no deviation in image direction occurs as the head rotates. For $r_V < 1$ P increases relative to V. As the listener turns away from $\boldsymbol{r_V}$ the image also moves away from $\boldsymbol{r_V}$ in the direction the head is moving, but more slowly. The maximum separation of the image from the listener, θ_{NI}, is less than 90^o. Conversely when $r_V > 1$ P decreases relative to V. As the listener turns away from $\boldsymbol{r_V}$ the image moves away from $\boldsymbol{r_V}$ in the opposite direction the head is moving, as shown by the dashed plots. In this case the maximum separation of the image from the listener θ_{NI} reaches 90^o for $\theta_{VN} < 90^o$. Above this value there is no consistent image, because the ITD is too large. This is why the dashed lines extend less than $\theta_{VN} = 90^o$.

4 Panning with compensation for head rotation

Amplitude panning is a spatial audio reproduction method in which several loudspeakers produce plane waves converging at the listener in phase. The pressure and velocity at the listener are given by the sum of the pressure and velocity of these waves. Then

$$-\boldsymbol{r}_V = Z_0 \frac{\boldsymbol{V}}{P} = Z_0 \frac{\sum \boldsymbol{V}_i}{\sum P_i} = \frac{\sum P_i \hat{\boldsymbol{V}}_i}{\sum P_i} = \frac{\sum g_i \hat{\boldsymbol{V}}_i}{\sum g_i} = -\frac{\sum g_i \hat{\boldsymbol{r}}_i}{\sum g_i} \tag{10}$$

where $\{g_i\}$ are gains applied to a source signal to provide the feeds for loudspeakers that are equidistant from the listener, and $\hat{\boldsymbol{r}}_i = -\hat{\boldsymbol{V}}_i$ are the unit vectors to the loudspeakers.

From (10) we see that $\boldsymbol{r}_V$ coincides with the definition of the *Makita localisation vector*[10],

$$\boldsymbol{r}_V = \frac{\sum g_i \hat{\boldsymbol{r}}_i}{\sum g_i} \tag{11}$$

Returning to the case where the listener faces the image, shown in Fig. 6, we see that the image direction $\hat{\boldsymbol{r}}_{\boldsymbol{I}}$ is the same direction as $\sum g_i \hat{\boldsymbol{r}}_i$. This is the defining characteristic for stereo tangent panning, and VBAP panning[11]. The adaptive panning described here is a generalisation of these methods.

In Fig. 7 we can interpret θ_{VI} as the image direction error when panning with the stereo tangent law for the case of loudspeakers separated by $2\theta_S$ with equal panning gains. For wider loudspeaker separations the error increases more rapidly as the listener turns away from the central direction.

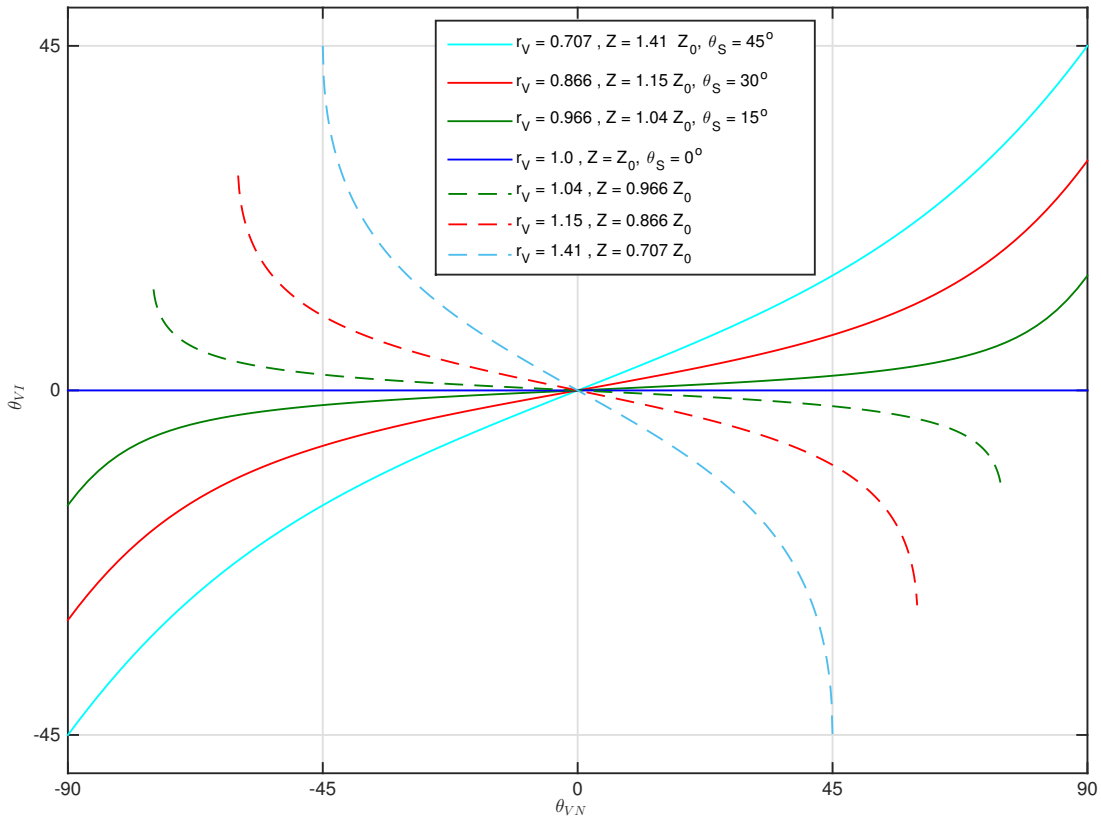

Figure 7: The image shift relative to the velocity θ_{VI} is plotted as a function of the angle between head direction and the velocity vector θ_{VN} for some values of r_V.

The constraints can be satisfied using stereo, that is two loudspeaker reproduction. To simplify the calculation we impose the additional constraint

$$g_1 + g_2 = 1 \tag{12}$$

This does not restrict the solution search in practice since given any solution, a solution satisfying the additional constraint exists by dividing gains by $(g_1 + g_2)$. Therefore all solutions are easily found from solutions satisfying (12).

Substituting the Makita vector (11), with 2 plane waves, in (9) and using (12) gives the gains[2] .

$$g_1 = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \tag{13}$$

$$g_2 = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_1)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_2 - \hat{\boldsymbol{r}}_1)} \tag{14}$$

There is no solution when the loudspeakers are symmetrically at the side of the listener, $\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2) = 0$.

The solution can be visualised using a vector diagram. The example shown in Fig. 8 is for centrally panned image with a listener facing the image. Fig. 9 shows how $\boldsymbol{r}_V$ must move to the right to

compensate for head movement to the left in order to keep the image central. Possible values of $\boldsymbol{r}_V$ that can be produced by panning between the two loudspeakers lie on the dashed line between $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_2$. Values of $\boldsymbol{r}_V$ which produce the desired images $\hat{\boldsymbol{r}}_I$ or $\hat{\boldsymbol{r}}'_I$ lie on the dotted line. The intersection of the two lines gives the one value of $\boldsymbol{r}_V$ that produces the desired image by panning.

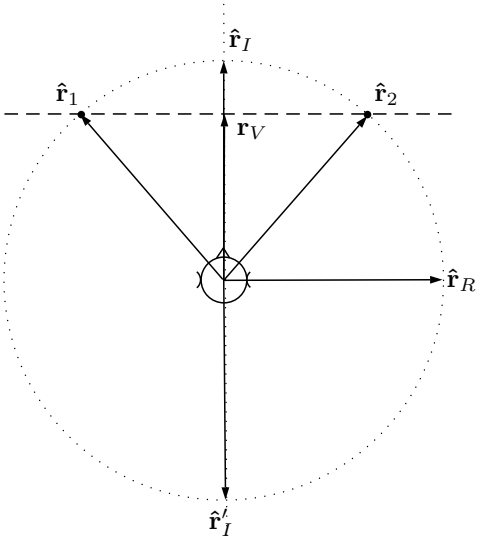

Figure 8: Vector diagram showing how a panned image is related to the field description and the head orientation. The listener is directly facing a centrally panned image.

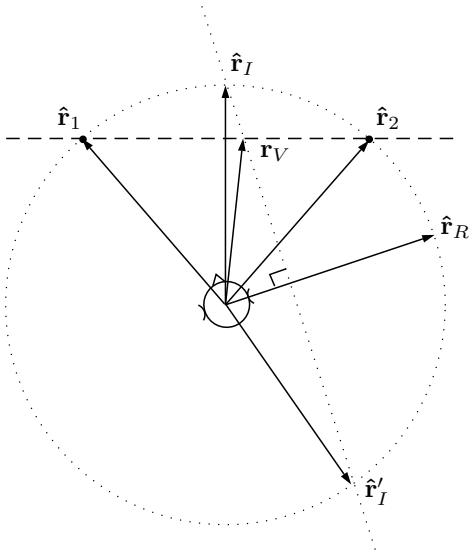

Figure 9: As the listener turns to the left the field description vector $\boldsymbol{r}_V$ moves to the right

Pulkki has investigated the panning of off median images[12]. In an experiment subjects adjusted the panning gains for two loudspeakers separated by 60^o so that the image aligned with real reference sources for various head directions. The gain ratio g_1/g_2 was recorded. The measured ratio can be compared with that predicted by the model presented here. Dividing (13) by (14) gives

$$\frac{g_1}{g_2} = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_I)} = \frac{\sin\theta_{NI} - \sin\theta_{N2}}{\sin\theta_{N1} - \sin\theta_{NI}} \tag{15}$$

where the angles $\theta_{NI}, \theta_{N1}, \theta_{N2}$ are between the listener facing direction to the image direction, and the loudspeaker directions respectively.

The test results from[12] are reproduced in Fig 10, with annotations added. Four different broadband sound sources were tested, shown at the bottom of the figure. The angles shown are relative to the central direction between the loudspeakers. The reference sources were located in one of three directions shown as Ref -15^o, Ref 0^o, Ref +15^o. The dashed green lines show the gain ratios required to pan the image to the reference direction with the listener facing the reference direction, the same as standard tangent law panning. The solid red lines indicate the gain ratio predicted using (15). There is broad agreement given the variability in the measured data, particularly for ratios less than 5dB magnitude. At larger dB magnitudes the scale is less sensitive to change of angle, which magnifies the error in dB. Also the high frequency component of broadband sounds will modify the cue expected from a purely low frequency sound.

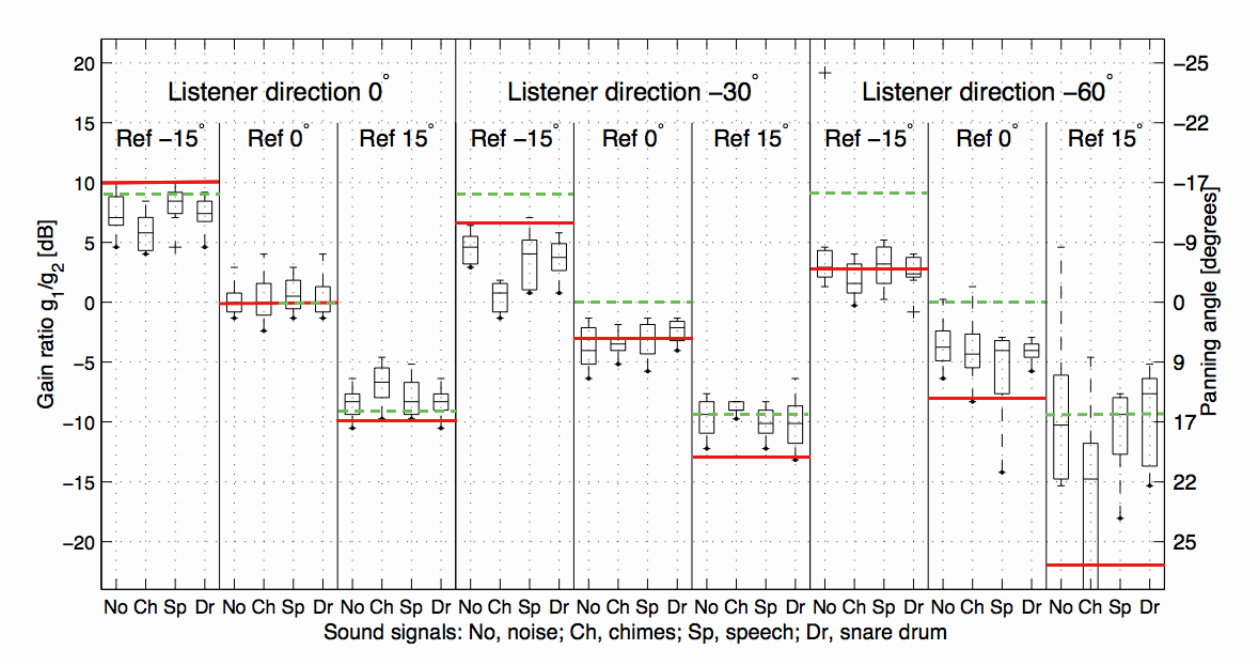

Figure 10: Results taken from Pulkki[12] with annotations added

Pulkki found compensation formula that fit the data for specific loudspeaker separations. (15) generates corrections efficiently for any speaker separation. This is especially useful where the listener may move and the angle subtended to the loudspeaker changes.

4.1 Near-field imaging

The previous section considered a plane wave source $\boldsymbol{V}_\Im = 0$. For a point source this value is non-zero,

$$\boldsymbol{V}_\Im = \frac{P}{Z_0 k r_I} \hat{\boldsymbol{r}}_I \tag{16}$$

where r_I is the distance to the desired image[4]. The ITD is then non zero. Provided the source is not very close to an ear, $r_I >> D/2$ then the first order approximation for the binaural signals is reasonably accurate. In practice this is useful because the perception of ITD extends for sources out to a radius of about 1.5m, compared with $D/2 \approx 0.07m$.

Equating the shapes of triangles of the form shown in Fig. 3 for target and and panned fields produces two conditions acting on the real and imaginary parts of the Makita vector.

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Re(\boldsymbol{r}_V) - \hat{\boldsymbol{r}}_I) = 0 \quad (17)$$

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Im(\boldsymbol{r}_V) + \frac{\hat{\boldsymbol{r}}_I}{kr_I}) = 0 \quad (18)$$

The first condition is the same as for the plane wave. These more complex constraints can still be satisfied by stereo panning gains. The additional constraint $g_1 + g_2 = 1$ is applied as before to simplify calculation. In this case it also separates the equations so that each acts either on real or imaginary parts of the gains. The resulting solution[13] is

$$\Re(g_1) = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \quad (19)$$

$$\Re(g_2) = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_1)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_2 - \hat{\boldsymbol{r}}_1)} \quad (20)$$

$$\Im(g_1) = -\Im(g_2) = \frac{\hat{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{r}}_I}{kr_I \hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \quad (21)$$

We see that the imaginary part of the gains, $j\Im(g_i)$ are frequency dependent and proportional to $1/(jk)$, constituting an integrating filter. This can be approximated with a simple one pole low pass filter with sufficiently low cutoff frequency. The phase response of this filter is not constant up to Nyquist, however the region of interest up to $1000Hz$ is well below Nyquist with typical sampling rates $\approx 40000Hz$, and so is approximately constant in this region.

5 Head translation

Panned images are unstable to head translation, as well as rotation. If the listener moves towards a loudspeaker the image *collapses* to that loudspeaker. We addressed the problem previously[14]. The gains are adjusted so the velocity at the listener is directed away from the desired image direction, see Fig 11. Then the listener facing in the desired image direction will have 0 ITD, and therefore perceive an image in this direction. The variable signal path from each loudspeaker should be taken into account, including compensation for distance attenuation and delay, and loudspeaker directivity. For a simple point source loudspeaker model the compensated gain g is gr where r is distance between listener and loudspeaker. The compensating delay introduced must be causal since the system is interactive, and is $(r_{max} - r)/c$ where r_{max} is the maximum value of r for any loudspeaker. These effects are fixed and equal in the simple symmetric stereo listening case, and so do not then need to be taken into account.

The adaptation for head rotation described in previous sections can readily be extended for positional adaptation simply by updating the vectors $\hat{\boldsymbol{r}}_{\boldsymbol{I}}, \hat{\boldsymbol{r}}_{\boldsymbol{R}}, \hat{\boldsymbol{r}}_{\boldsymbol{1}}, \hat{\boldsymbol{r}}_{\boldsymbol{2}}$ as the listener moves. Compensation is required for the variable signal path from each loudspeaker, as described above. The panning described in[14] is a special case of rotational adaptation with the listener facing the image, that is when $\hat{\boldsymbol{r}}_{\boldsymbol{I}} \cdot \hat{\boldsymbol{r}}_{\boldsymbol{R}} = 0$.

6 Implementation and informal evaluation

A symmetrical panning system with adaptation for head rotation was implemented using the Cycling74 Max/MSP framework. An overview is shown in Fig. 12. The head orientation was initially sensed using a Polhemus Fastrak magnetic tracking device. Once the system was working reliably this was

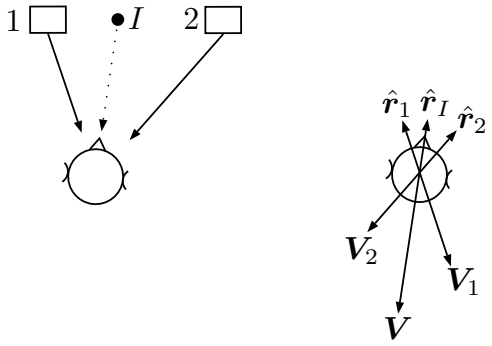

Figure 11: A translated listener with velocity $\boldsymbol{V}$ adjusted so that an image appears directly ahead.

replaced with a Microsoft Kinect based video tracking tracking system, which was previously developed for listener tracking for adaptive panning with translation[14]. Video tracking has the advantage of being non-invasive for the listener.

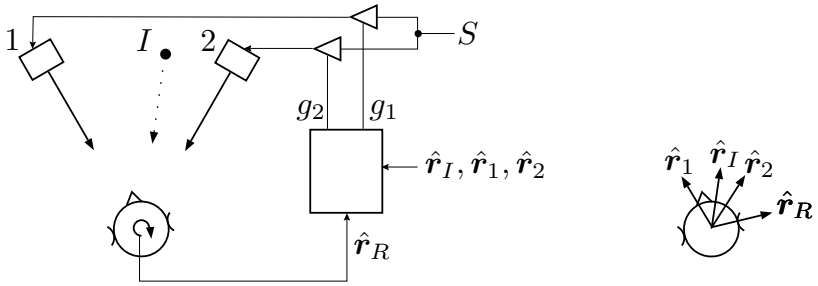

Figure 12: Panning with adaptation for head rotation, shown for one object image.

The angle between the loudspeakers can be specified. Vectors are represented with cartesian coordinates. The loudspeaker signal gains are calculated using (13), (14), for each reproduced object image. The denominators can be pre-calculated since they are the same for all images. Whenever the head changes direction the vector $\hat{\boldsymbol{r}}_R$ is reevaluated from the azimuthal angle (head rotation is assumed to be about the vertical axis).

In informal listening the effect of rotation adaptation is immediately clear. The images becomes more stable and this increases the plausibility of the images as real sources. This is particularly true of source signals with significant content below 1000Hz. It becomes evident that in normal stereo listening we become accustomed to ignoring the image distortions caused by head rotation in order to focus on the most plausible aspects of the reproduction.

If the head is moved towards the limit, pointing parallel to the line connecting the loudspeakers, the image becomes weaker losing bass frequencies, then eventually becoming unstable. This is because r_V becomes > 1 and the gains become negative with respect to each other. P becomes comparable with $P_L - P$, which means P_L increases significantly with k, causing apparent loss of bass. This could be addressed by applying compensating equalisation to the source signal before panning, but this is not explored here.

Controlling the rear image direction $\hat{\boldsymbol{r}}_I'$ in Fig. 9 results in the perception of a rear image even though there are no loudspeakers in that direction. This striking effect occurs because the way ITD changes with head movement is only consistent with the rear image. The effect is strengthened by low pass filtering in order to reduce the opposing effect of high frequency cues. This filtering is consistent with head related transfer functions, and therefore not unnatural. The effect was first described in 1940

by Wallach,[15]. Here it is naturally integrated into the process for rotation adaptation. The adaptation stabilises the rear images as well as the frontal images.

7 Conclusion

Linearising the field at low frequencies leads to some compact and practical expressions for panning gain with adaptation for head rotation. The same framework provides a simple method for reproducing near sources, also with rotation adaptation.

The dominance of the ITD cue over other cues operating at higher frequency can explain why this method works well with broadband source signals, The situation with ILD is not currently so clear. ILD cues also operate at higher frequencies where the panning process is no longer so predictable.

The relationship of these methods with transaural reproduction is interesting. The advantage of the current panning approach for low frequencies compared with such methods is that we do not attempt to control the ears independently, which is inherently difficult at low frequencies. As a result the head size r is not required in the panning calculations, whereas in the dynamic transaural approach it is required both in the binaural encoding stage and the transaural reproduction of the encoded binaural signals. A combination of object based panning methods at lower frequency and transaural methods at higher frequency could be appropriate.

Acknowledgment

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. No new data were created during this study.

References

[1] F. L. Wightman and D. J. Kistler, “The dominant role of low-frequency interaural time differences in sound localization,” *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992.

[2] D. Menzies and F. Fazi, “A theoretical analysis of sound localisation, with application to amplitude panning,” in *Proc. AES 138th Convention, Warsaw*, May 2015.

[3] F. Fazi and D. Menzies, “Estimation of the stability of a virtual sound source using a microphone array,” in *Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence*, July 2015.

[4] D. Menzies and F. Fazi, “Spatial reproduction of near sources at low frequency using adaptive panning,” in *Proc. TecniAcustica, Valencia*, 2015.

[5] P. Morse and K. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968.

[6] B. Xie, *Head-related transfer function and virtual auditory display*. J Ross, 2013.

[7] F. Filippo, M. Noisternig, O. Warusfel *et al.*, “Representation of sound fields for audio recording and reproduction,” *Acoustics 2012 Nantes*, 2012.

[8] E. Williams, *Fourier Acoustics: sound radiation and nearfield acoustical holography*. Elsevier, 1999.

[9] J. Blauert, “Spatial hearing,” 1983.

[10] M. A. Gerzon, “General metatheory of auditory localisation,” in *92nd Audio Engineering Society Convention, Vienna*, 1992.

[11] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7853

[12] ——, “Compensating displacement of amplitude-panned virtual sources,” in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Jun 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11138

[13] D. Menzies, “Spatial reproduction of near sources at low frequency using adaptive panning,” in *In Proc. TechniAcustica 2015, Valencia*, October 2015.

[14] M. Simon, D. Menzies, F. Fazi, T. de Campos, and A. Hilton, “A listener position adaptive stereo system for object based reproduction,” in *Proc. AES 138th Convention, Warsaw*, May 2015.

[15] H. Wallach, “On sound localization,” *J. Acoust. Soc. Am*, vol. 10, pp. 270–274, 1940.