SPATIAL REPRODUCTION OF NEAR SOURCES AT LOW FREQUENCY USING ADAPTIVE PANNING

Dylan Menzies
Institute of Sound and Vibration Research, University of Southampton, UK
d.menzies@soton.ac.uk

Filippo Maria Fazi
Institute of Sound and Vibration Research, University of Southampton, UK

ABSTRACT

Panned images are unstable to head rotation. In previous work it was shown how the panning gains can be compensated using the tracked head orientation so that the Interaural Time Difference is controlled and images are stabilised. We now extend this method for near point images, for which the Interaural Level Difference must also be considered. The panning gains are no longer frequency independent, but can be approximated efficiently using a 1st order filter.

1 INTRODUCTION

In previous works[1,2] we estimated the localisation cues produced by a general low frequency sound field. The Interaural Time Difference (ITD) and Interaural Level Difference (ILD) were found from first order field approximation with a free-field scattering model. This framework was applied to the problem of panning reproduction. When the listener's head rotates in the sound field produced by a panning system, images produced using static gains are not fixed, for example see Fig. 1, leading to instability in the overall image. A simple formula was found relating the head orientation, image direction, and the field description given as a specific admittance, or equivalently as a Makita localisation vector. Dynamic panning gains were then found that depend on the head orientation as well as image direction, and good agreement was found with experiment.
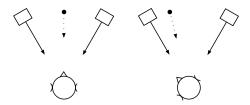

Figure 1: The black dot indicates the direction of the image when two loudspeakers each have the same signal, for different head directions.

The analysis here introduces a more accurate model for the binaural signals, using a spherical head scattering model. Surprisingly, and fortunately, the results are unaffected.The panning sources and images were previously assumed to be distant, with the individual panning fields phase aligned at the listener. This causes ITD to be zero. The main aim of this paper is to this paper is to generalise the analysis for near images, for which ITD is no longer zero. It is then hoped to create convincing images between the source loudspeakers and the listener, by modifying the panning system.

In related work we have also shown how to compensate a panning system for listener location,

creating a dynamic sweetspot, with motion parallax for distance cues[3]. Combining this process with rotation compensation provides stable images while giving the listener full freedom of movement.

We proceed by reviewing previous analysis, and extend this towards a solution for near-field panning.

2 Sound field representation

A source free region of a sound field can be expanded as a series about any point in the region[4]. 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 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 the region, 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[4]. 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)[5] .

Furthermore the resultant pressure field is equal to the incident field at a radius $\frac{3}{2}r$, where r is the sphere radius, by projection through the centre. Since any free field region can be approximated arbitrarily well by a plane wave expansion, this implies that for any incident field 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

$$P_R \quad = \quad P + \bar{\boldsymbol{r}}_R \cdot \nabla P \tag{2}$$
$$P_L \quad = \quad P + \bar{\boldsymbol{r}}_L \cdot \nabla P \quad , \tag{3}$$

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.
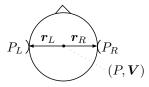

Figure 2: Sound field variables and vectors in relation to the listener's head.

The gradient is related to particle velocity $\boldsymbol{V}$ by Euler's equation in the frequency domain[6],

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

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 pressure phases 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,

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

with the phase rotated velocity, $\boldsymbol{V}'$, decomposed 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 can be written $kr_R << 1$.

Figs 3 shows the general case where both $\boldsymbol{V}_{\Im}$ and $\boldsymbol{V}_{\Re}$ are non-zero, and in different directions. As the listener's head rotates P_R and P_L move around on opposite sides of an ellipse.

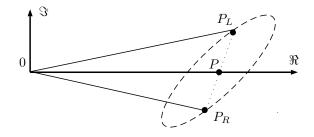Figure 3: P_R and P_L in the complex plane for non zero and non aligned $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$

2.1 Point source field representation

The pressure field of a point source is

$$P = \frac{A}{r}e^{-jkr} \tag{8}$$

where r is the distance from source. The gradient is the

$$\nabla P = -P(\frac{1}{r} + jk)\hat{\boldsymbol{r}} \tag{9}$$

where $\hat{\boldsymbol{r}}$ is the unit vector in the direction from the source to the measurement position. Using (4)

$$\boldsymbol{V} = \frac{P}{Z_0}(1 - \frac{j}{kr})\hat{\boldsymbol{r}} \tag{10}$$

The complex velocity components are then

$$\boldsymbol{V}_{\Re} = \frac{P}{Z_0}\hat{\boldsymbol{r}} = -\frac{P}{Z_0}\hat{\boldsymbol{r}}_I \tag{11}$$
$$\boldsymbol{V}_{\Im} = -\frac{P}{Z_0 kr}\hat{\boldsymbol{r}} = \frac{P}{Z_0 kr_I}\hat{\boldsymbol{r}}_I \tag{12}$$

where $\hat{\boldsymbol{r}}_I = -\hat{\boldsymbol{r}}$ is the unit vector in the direction from the measurement point, where the listener is, to the source.

3 Localisation cues and reproduction

ITD, also known as Interaural Phase Delay for a harmonic field[7], is equal to ϕ_{RL}/ω, where $\phi_{RL} = \arg(P_R/P_L)$ is the Interaural Phase Difference. ILD is given by $|P_R/P_L|$. For a given pair (ITD, ILD) there is only one possible shape of the form $0 \to P_L \to P \to P_R$ in the diagram Fig. 3 (ignoring the ellipse). This can be seen by first constructing the lines through $0 \to P_L$ and $0 \to P_R$ for the given ITD, and then choosing P_L and finding P_R to satisfy the given ILD. The point at P is then midway between P_L and P_R.

Hence if we wish to reproduce the same image, that is the same ITD and ILD, using a synthetic field as would be produced by a natural source field then the corresponding diagrams must be similar. The shape of each diagram can be specified with two quantities describing the ratio of lengths in the diagram,

$$\frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Re}{P} , \quad \frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Im}{P} \tag{13}$$

For a point source using (11) we have

$$\frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Re}{P} = -\frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{r}_I}{Z_0} \tag{14}$$

$$\frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_\Im}{P} = \frac{\bar{\boldsymbol{r}}_R \cdot \boldsymbol{r}_I}{k r_I Z_0} \tag{15}$$

These equations can be reinterpreted es constraints of similarity between a point source field and a general field $(P, \boldsymbol{V})$. Rearranging, the constraints can be written

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\frac{Z_0}{P} \boldsymbol{V}_\Re + \hat{\boldsymbol{r}}_I) = 0 \tag{16}$$

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\frac{Z_0 k r_I}{P} \boldsymbol{V}_\Im - \hat{\boldsymbol{r}}_I) = 0 \tag{17}$$

The first constraint was arrived at in our previous studies on distant image reproduction. The second constraint in the form (15) was also satisfied since then $\boldsymbol{V}_\Im = 0$ and r_I is infinite.

4 Panning

Amplitude panning is a spatial audio reproduction method in which several loudspeakers are assumed to 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

$$Z_0 \frac{\boldsymbol{V}}{P} = Z_0 \frac{\sum \boldsymbol{V}_i}{\sum P_i} = \frac{\sum P_i \hat{\boldsymbol{V}}_i}{\sum P_i} = \frac{\sum g_i \hat{\boldsymbol{V}}_i}{\sum g_i} = -\boldsymbol{r}_V \tag{18}$$

where $\{g_i\}$ are gains applied to a source signal to provide the feeds for loudspeakers that are equidistant from the listener. $\boldsymbol{r}_V$ is the *Makita localisation vector*,

$$\boldsymbol{r}_V = \frac{\sum g_i \hat{\boldsymbol{r}}_i}{\sum g_i} \tag{19}$$

where $\hat{\boldsymbol{r}}_i = -\hat{\boldsymbol{V}}_i$ are the unit vectors to the loudspeakers. In its original form $\boldsymbol{r}_V$ is defined with real valued gains. However this can be extended by allowing complex valued gains, corresponding to to plane waves arriving at the listener with relative phase offsets. $\boldsymbol{r}_V$ is then complex valued. The

constraints (16) can then be written

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Re(\boldsymbol{r}_V) - \hat{\boldsymbol{r}}_I) \quad = \quad 0 \tag{20}$$

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Im(\boldsymbol{r}_V) + \frac{\hat{\boldsymbol{r}}_I}{k r_I}) \quad = \quad 0 \tag{21}$$

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

$$g_1 + g_2 = 1 \tag{22}$$

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 (22). The new constraint allows the constraints (20, 21) to be written in terms of separate sets of variables $(\Re(g_1), \Re(g_2))$ and $(\Im(g_1), \Im(g_2))$. The constraint (22) implies

$$\Re(g_1) + \Re(g_2) \quad = \quad 1 \tag{23}$$

$$\Im(g_1) + \Im(g_2) \quad = \quad 0 \tag{24}$$

hence from the definition (19),

$$\Re(\boldsymbol{r}_V) = \Re(g_1 \hat{\boldsymbol{r}}_1 + g_2 \hat{\boldsymbol{r}}_2) = \Re(g_1)(\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2) + \hat{\boldsymbol{r}}_2 \tag{25}$$

$$\Im(\boldsymbol{r}_V) = \Im(g_1 \hat{\boldsymbol{r}}_1 + g_2 \hat{\boldsymbol{r}}_2) = \Im(g_1)(\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2) \tag{26}$$

Substituting (25) in (20) gives the real gains, which were calculated in previous work[2] . The solution fails when the loudspeakers are symmetrically at the side of the listener and $\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2) = 0$.

$$\Re(g_1) \quad = \quad \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \tag{27}$$

$$\Re(g_2) \quad = \quad \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_1)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_2 - \hat{\boldsymbol{r}}_1)} \tag{28}$$

The imaginary gains are similarly found by substituting (26) in (21),

$$\Im(g_1) = -\Im(g_2) = \frac{\hat{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{r}}_I}{k r_I \hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \tag{29}$$

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 Reproduction for very near images

The approximation discussed in the previous section is valid for long wavelengths where the contribution to the binaural signals from higher field derivatives is small compared to the contribution

pressure and gradient a the centre. This is useful because ILD is perceptually significant up to 1-2m from the listener and the ears are approximately 0.07 m from the centre of the head.

However when the source is close to the ear compared with the head radius the higher derivatives become significant. (6) and (7) are no longer accurate for describing the target source field (the binaural signal becomes infinite when the source is at an ear). On the other hand these equations remain valid when applied to the panned field, which consists of the summation of two plane waves (the total pressure and velocity can only be zero if the two plane waves are from the same direction but with opposite phase). The approach of the previous section can be modified by giving the ILD and ITD according to a more accurate model for the target. According to the argument in Section 3 ILD and ITD are sufficient to define the shape of Fig. 3, and the ratios in (13). Applying the dot product of $\hat{r}_R$ to (25) and (26) allows the left side to be written in terms of the ratios from (13).

$$\hat{r}_R \cdot \Re(r_V) = \Re(g_1)\hat{r}_R \cdot (\hat{r}_1 - \hat{r}_2) + \hat{r}_R \cdot \hat{r}_2 \tag{30}$$
$$\hat{r}_R \cdot \Im(r_V) = \Im(g_1)\hat{r}_R \cdot (\hat{r}_1 - \hat{r}_2) \tag{31}$$

Rearranging for $\Re(g_1)$ and $\Im(g_1)$, and applying (23) provides the solution for the panned image.

The left hand sides of (30,31) can be found from the ITD and ILD as follows (We do not give a model for ITD and ILD here, which could be based on simulation or measurement). ITD and ILD are functions of k. Fig. 4 illustrates the variables now used. The length P in Fig. 3 is represented here with length 1 for simplicity.
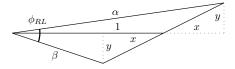

Figure 4: Illustrates variables used in geometric calculation of ratios from ITD and ILD.

The values to be determined are

$$x = kZ_0\bar{r}_R \cdot V_\Im/P = k\bar{r}_R \cdot \Im(r_V) \tag{32}$$
$$y = -kZ_0\bar{r}_R \cdot V_\Re/P = -k\bar{r}_R \cdot \Re(r_V) \tag{33}$$

The definition of ILD is

$$\text{ILD} = \frac{\alpha}{\beta} \tag{34}$$

Trignometric identities give

$$\alpha^2 = (1+x)^2 + y^2 \tag{35}$$
$$\beta^2 = (1-x)^2 + y^2 \tag{36}$$
$$4(x^2+y^2) = \alpha^2 + \beta^2 - 2\alpha\beta\cos(\phi_{RL}) \tag{37}$$

Where $\phi_{RL} = \omega\text{ITD}$. Adding (35),(36) provides an expression in α, β for (x^2+y^2) which can be substituted into (37). Using (34) the result can be written as a quadratic equation for α or β. x is then given by subtracting (35),(36), and y then e.g. from (35),

$$x = \frac{1}{4}(\alpha^2 - \beta^2) \tag{38}$$
$$y = \sqrt{\alpha^2 - (1+x)^2} \tag{39}$$

The quantities required to calculate the gains in (30, 31) are then $\hat{r}_R \cdot \Re(r_V) = -y/k, \ \ \hat{r}_R \cdot \Im(r_V) = x/k$. This will generally lead to the real and imaginary parts of the gains that do not have a simple dependence on k, and so will be more complex to implement as a filter compared with the integrating filter identified in Section 4.

6   Conclusion

It has been shown how the previous work reproducing panned images with compensation for head rotation can be extended for near images. An efficient formulation has been found that is valid provided the image is not very near.

While the ITD cue is limited to low frequencies, the ILD cue is not, and interference is expected from conflicting high frequency ILD. Transaural methods can be considered for the control of high frequencies. The advantage of the current approach for low frequencies compared with such methods is that we do not attempt to control the ears individually, which is inherently difficult at low frequencies. Also the head size r is not required in the calculations here, whereas in the dynamic transaural approach it is required both in the binaural encoding stage and the transaural reproduction of the encoded binaural signals.

The approach described here has already been shown effective experimentally for far images. Further experimental work will be needed to investigate the effectiveness of the near field formulation.

Acknowledgment

References

[1] 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.

[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] 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.

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

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

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

[7] J. Blauert, "Spatial hearing," 1983.