Audio Engineering Society

Convention Paper

Presented at the 138th Convention
2015 May 7–10 Warsaw, Poland

A theoretical analysis of sound localisation, with application to amplitude panning

Dylan Menzies[1], Filippo Maria Fazi[1]

[1] *Institute of Sound and Vibration Research, University of Southampton, UK*

Correspondence should be addressed to Dylan Menzies (d.menzies@soton.ac.uk)

ABSTRACT

Below 700Hz sound fields can be approximated well over a region of space that encloses the human head, using the acoustic pressure and gradient. With this representation convenient expressions are found for the resulting Interaural Time Difference (ITD) and Interaural Level Difference (ILD). This formulation facilitates the investigation of various head-related phenomena of natural and synthesised fields. As an example, perceived image direction is related to head direction and the sound field description. This result is then applied to a general amplitude panning system, and can be used to create images that are stable with respect to head direction.

1. 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 gradient by

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

The approximation is good provided the wavelength is considerably larger than the region, and at least one of $P(\boldsymbol{x_0})$, $\nabla P(\boldsymbol{x_0})$ is not very small compared to higher order derivative terms. Below 700Hz a typical sound field region large enough to enclose the human head satisfies these conditions. For natural sound fields it would be very unusual for the first two terms to be highly suppressed relative to higher terms, and we do not consider this possibility here for synthetic fields. Further more at these wavelengths the head is approximately acoustically transparent, so the pressure signals at the ear positions when the listener is present, that is the binaural signals, closely match the corresponding pressures of the incident field. Therefore to a first order approximation the binaural signals at the right and

left ears are given by

$$P_R = P + \boldsymbol{r}_H \cdot \nabla P \quad (2)$$
$$P_L = P - \boldsymbol{r}_H \cdot \nabla P \ , \quad (3)$$

where P and ∇P are the pressure and gradient at the central point between the ears and $\boldsymbol{r}_H$ is the vector from the centre to the right ear. The gradient is related to particle velocity $\boldsymbol{V}$ by Euler's equation in the frequency domain,

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

using the positive frequency convention $p(\boldsymbol{x}, t) = P(\boldsymbol{x})e^{j\omega t}$ [8]. Z_0 is the free-field impedance and k is the wavenumber. In the following discussion only relative pressure phases are of interest, so it can be assumed without loss of generality that P is positive real valued. $\boldsymbol{V}$ can be written as a sum of real and imaginary vectors,

$$\boldsymbol{V} = \boldsymbol{V}_\Re + j\boldsymbol{V}_\Im \quad (5)$$

Hence,

$$P_R = P + kZ_0(\boldsymbol{r}_H \cdot \boldsymbol{V}_\Im - j\boldsymbol{r}_H \cdot \boldsymbol{V}_\Re) \quad (6)$$
$$P_L = P - kZ_0(\boldsymbol{r}_H \cdot \boldsymbol{V}_\Im - j\boldsymbol{r}_H \cdot \boldsymbol{V}_\Re) \quad (7)$$

The low frequency approximation condition can be written $kr_H << 1$.

Fig 1 shows P_R and P_L in the complex plane when only $\boldsymbol{V}_\Re$ is non-zero. There is then a phase difference between the binaural signals, but no amplitude difference. The phase difference ϕ_{RL} is greatest when the velocity vector $\boldsymbol{V}_\Re$ is along the inter-aural axis, since $|\boldsymbol{r}_H \cdot \boldsymbol{V}_\Re|$ is then at maximum. As the head rotates P_R and P_L move inwards and outwards along the dashed line.

Fig 2 shows the case where only $\boldsymbol{V}_\Im$ is non-zero. There is no phase difference between the two binaural signals, however they do differ in amplitude.

Figs 3,4 show more general examples where both $\boldsymbol{V}_\Im$ and $\boldsymbol{V}_\Re$ are non-zero. Fig 3 shows the case where $\boldsymbol{V}_\Im$ and $\boldsymbol{V}_\Re$ are in the same direction. Then

$$P_R = P + kZ_0 \boldsymbol{r}_H \cdot \hat{\boldsymbol{V}}(V_\Im - V_\Re j) \quad (8)$$
$$P_L = P - kZ_0 \boldsymbol{r}_H \cdot \hat{\boldsymbol{V}}(V_\Im - V_\Re j) \ , \quad (9)$$

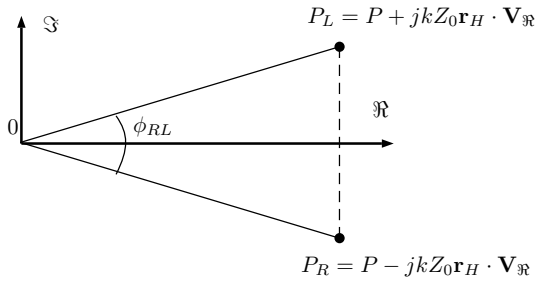

Fig. 1: P_R and P_L in the complex plane for non-zero $V_\Re$

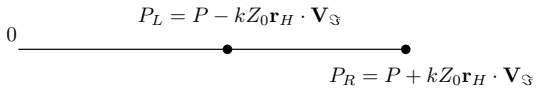

Fig. 2: P_R and P_L in the complex plane for non-zero $V_\Im$

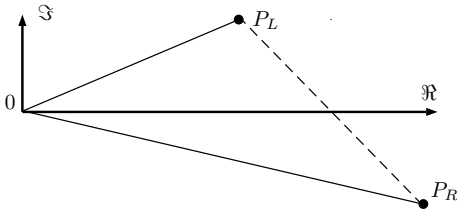

Fig. 3: P_R and P_L in the complex plane for aligned $V_\Re$ and $V_\Im$

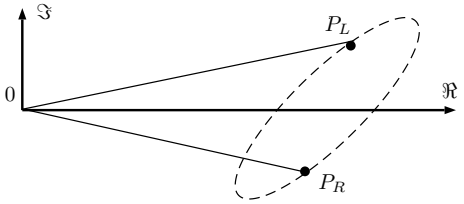

Fig. 4: P_R and P_L in the complex plane for non-aligned $V_\Re$ and $V_\Im$

where the direction norm $\hat{\boldsymbol{V}} = \hat{\boldsymbol{V}}_{\Re} = \hat{\boldsymbol{V}}_{\Im}$, so P_R and P_L lie on a the same straight line for all head directions determined by $\boldsymbol{r}_H$. Where $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$ are in different directions then P_R and P_L move around an ellipse as the head rotates, as shown in Fig 4.

2. LOCALIZATION CUES

There are two types of sound localisation cues operating in the range below 700Hz, Interaural Time Difference (ITD) and Interaural Level Difference (ILD) [1].

For a harmonic field like that in (6,7) the ITD is also called the Interaural Phase Delay, and is ϕ_{RL}/ω, where $\phi_{RL} = \arg(P_R/P_L)$ is the Interaural Phase Difference. For purely real $\boldsymbol{V}$, i.e. $\boldsymbol{V}_{\Im} = 0$, P and $\boldsymbol{V}$ are in phase, and P_R and P_L are of equal amplitude, but differ in phase except when the listener is pointing in the direction $\boldsymbol{V}_{\Re}$. Then

$$\tan(\phi_{RL}/2) = \frac{kZ_0 \boldsymbol{r}_H \cdot \boldsymbol{V}_{\Re}}{P} \tag{10}$$

For $kr_H << 1$ ITD is frequency independent [1], $\phi_{RL}/\omega \approx Z_0 \boldsymbol{r}_H \cdot \boldsymbol{V}/(cP)$. For example for a plane wave, $P = Z_0 V$.

Interaural Level Difference (ILD) is the amplitude ratio $|P_R/P_L|$. For purely imaginary $\boldsymbol{V}$, i.e. $\boldsymbol{V}_{\Re} = 0$, then P_R and P_L are in phase or anti-phase, but differ in amplitude except when the listener is pointing in the direction $\boldsymbol{V}_{\Im}$. The ITD is then zero and the ILD is given by

$$\left|\frac{P_R}{P_L}\right| = \left|\frac{P + kZ_0 \boldsymbol{r}_H \cdot \boldsymbol{V}_{\Im}}{P - kZ_0 \boldsymbol{r}_H \cdot \boldsymbol{V}_{\Im}}\right| \tag{11}$$

In general $\boldsymbol{V}_{\Im}$ and $\boldsymbol{V}_{\Re}$ are both non zero. If the vectors point in different directions then the cues are inconsistent for a single source. The difference between the vectors could be used in constructing an objective measure of image quality. ILD at low frequencies is only consistent with sources that are near compared to the ear separation $2r$. Controlling both $\boldsymbol{V}_{\Im}$ and $\boldsymbol{V}_{\Re}$ it may be possible to create consistent near-field cues.

$\boldsymbol{V}/P$ can be viewed as a local complex vector admittance. For real values of $\boldsymbol{V}/P$ the incident field appears locally as a plane wave whose wavelength is longer or shorter than of a free-field plane wave of the same frequency, and the ITDs are correspondingly greater or smaller.

$\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$ are directly related to the active and reactive intensity vectors $\boldsymbol{I}_a$, $\boldsymbol{I}_r$.

$$\boldsymbol{V}_{\Re} = \frac{1}{|P|}\Re(P\boldsymbol{V}^*) = \frac{2}{|P|}\boldsymbol{I}_a \tag{12}$$

$$\boldsymbol{V}_{\Im} = \frac{1}{|P|}\Im(P\boldsymbol{V}^*) = \frac{2}{|P|}\boldsymbol{I}_r \tag{13}$$

The formula are valid for P and $\boldsymbol{V}$ both complex valued, with $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$ defined as before by first rotating P and $\boldsymbol{V}$ in the complex plane so that P is positive real valued.

$\boldsymbol{I}_a$ is frequently used as a measure of localisation, however $\boldsymbol{I}_r$ is generally overlooked even though it can have just as much impact. The use of $\boldsymbol{I}_r$ in tandem with $\boldsymbol{I}_a$ should therefor be encouraged.

3. IMAGE DIRECTION

Synthesized sound fields can produce virtual images whose direction varies significantly depending on the listener's head orientation. For example in Fig. 5 an image is generated by feeding the same signal to two loudspeakers. The image appears centrally if the listener faces it but if the listener faces away the image direction is shifted towards the listener's new direction.

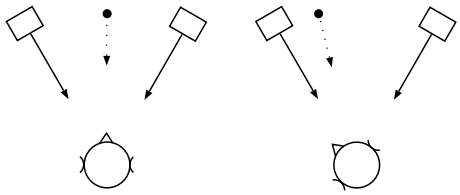

Fig. 5: Showing the image when two loudspeakers each have the same signal, for different head directions.

This effect can cause unwanted perceptual instability. Pulkki investigated this for the case of panning between two loudspeakers [6], and created compensated panning functions by fitting to subjective measurements for particular angular loudspeaker separations. In the following, the off-median image direction is predicted from first principles using the sound field description.

Assuming $\boldsymbol{V}_{\Re} \neq 0$ and $\boldsymbol{V}_{\Im} = 0$, and discounting other factors such as high frequency spectral, visual and dynamic cues, a listener in the field will perceive an image in a direction based only on the current ITD. The key observation in this section is that we can reasonably suppose that the listener processes the ITD information under the assumption that it *originates from a plane wave*, because there is no contrary cue, even though the field may be more general as we consider here. The perceived direction should therefor be that of a plane wave with same ITD as the incident field described by P and ∇P. From (10) the interaural phase difference ϕ_{RL} then satisfies

$$\tan(\phi/2) = \frac{kZ_0 \boldsymbol{r}_H \cdot \boldsymbol{V}}{P} = \frac{kZ_0 \boldsymbol{r}_H \cdot \boldsymbol{V_I}}{P_I} \qquad (14)$$

where $\boldsymbol{V_I}$ and P_I describe a plane wave travelling from the perceived image direction. Since $P_I = Z_0 V_I$

$$\frac{Z_0 \boldsymbol{r}_H \cdot \boldsymbol{V}}{P} = \boldsymbol{r}_H \cdot \hat{\boldsymbol{V}}_I \ , \qquad (15)$$

where $\hat{\boldsymbol{V}}_{\boldsymbol{I}}$ is the normalised vector in the direction of $\boldsymbol{V_I}$.

And so,

$$\hat{\boldsymbol{r}}_H \cdot (\alpha \hat{\boldsymbol{V}} + \hat{\boldsymbol{r}}_I) = 0 \qquad (16)$$

where the relative admittance $\alpha = Z_0 V/P$, and the direction norm to the image $\hat{\boldsymbol{r}}_I = -\hat{\boldsymbol{V}}_I$. Note that $\hat{\boldsymbol{r}}_H$ could instead be replaced with the reverse vector pointing to the left ear. (16) can also be written

$$\alpha \cos\theta_{HV} = \cos\theta_{HI} \qquad (17)$$

where θ_{HV} and θ_{HI} are the angles between $\hat{\boldsymbol{r}}_H$ and $-\hat{\boldsymbol{V}}$, and $\hat{\boldsymbol{r}}_H$ and $\hat{\boldsymbol{r}}_I$ respectively.

So given the head direction and the field description, and provided $\alpha \cos\theta_{HV} = \alpha \hat{\boldsymbol{r}}_H \cdot \hat{\boldsymbol{V}} < 1$ then the image direction $\hat{\boldsymbol{r}}_I$ has possible values on a *cone of confusion* about $\hat{\boldsymbol{r}}_H$ with angle θ_{HI}. Otherwise the most plausible value for $\hat{\boldsymbol{r}}_I$ is $\hat{\boldsymbol{r}}_H$, as this gives the closest ITD agreement. It is striking that $\hat{\boldsymbol{r}}_I$ is independent of the listener head dimensions, and any individual mapping between ITD and source direction. The relationships in (16,17) are illustrated by cross-section in Fig. 6. The cone of confusion extends out of the page.

If the listener were facing in the direction $-\hat{\boldsymbol{V}}$ then they would perceive an image in the median plane.

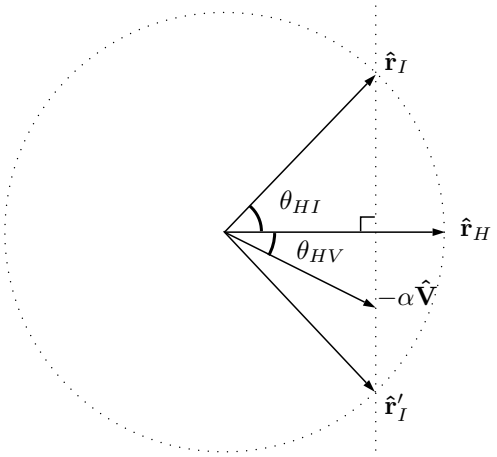

Fig. 6: Spatial vector diagram showing the relationship between the possible image directions $\hat{\boldsymbol{r}}_I$ and $\hat{\boldsymbol{r}}'_I$, inter-aural direction $\hat{\boldsymbol{r}}_H$, and field description vector $-\alpha \boldsymbol{V}$. The circle has unit radius.

For the case where all vectors all restricted to the horizontal plane, the shift in perceived image direction caused by a general listener head direction given by $\hat{\boldsymbol{r}}_H$, is $\theta_{IV} = \theta_{HV} - \theta_{HI}$.

If the field is static and the shift is non-zero, then a static image cannot be reproduced. However by tracking the head direction it is possible to modify the field so that the image is static and stable. In the next section the control of the field using panning is described.

For a field containing a mixture of frequencies, α may be constant or vary over frequency. If it varies then the image will be spread over a range of directions.

4. APPLICATION TO 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. So

$$\alpha = Z_0 \frac{|\sum \boldsymbol{V}_i|}{\sum P_i} = \frac{|\sum P_i \hat{\boldsymbol{V}}_i|}{\sum P_i} = \frac{|\sum g_i \hat{\boldsymbol{V}}_i|}{\sum g_i} \qquad (18)$$

where $\{g_i\}$ are gains applied to a source signal to provide loudspeaker feeds. The *Makita localisation*

vector [2–4] is defined by

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

where $\hat{\boldsymbol{r}}_i = -\hat{\boldsymbol{V}}_i$ are the direction norms to the loudspeakers. (16) can then be written

$$\hat{\boldsymbol{r}}_H \cdot (\hat{\boldsymbol{r}}_I - \boldsymbol{r}_V) = 0 \tag{20}$$

$\hat{\boldsymbol{r}}_V = -\hat{\boldsymbol{V}}$ and $\alpha = r_V = |\boldsymbol{r}_V|$.

According to ambisonic panning a plane wave is reproduced in a local region at the listener's head, that is $r_V = 1$, by using positive and negative gains in loudspeakers that surround the listener, ideally at regular intervals. If the listener moves, tracking is required so that the sweetspot region can be moved with the listener. Using the (20) the cues for a plane wave can reproduced using only loudspeakers within one half of the space around the listener. This could provide greater flexibility in some applications. Tracking can also be used to achieve correct parallax according to target source distance, as discussed in [7].

5. CONCLUSION

In the low frequency region ITD and ILD have a simple representation in terms of the complex pressure and velocity components. This was used here to derive a compact formula relating image and head directions with the representation. The approach can be used to find panning gains for a desired image direction given the listener's head direction, giving significant improvement over panning based on the tangent law. Initial indication is that this agrees well with experiment. This and other applications of the pressure-velocity representation of ITD and ILD will be further investigated in the future.

6. ACKNOWLEDGEMENTS

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.

7. REFERENCES

[1] Jens Blauert. Spatial hearing, 1983.

[2] M. A. Gerzon. Practical periphony: The reproduction of full-sphere sound. In *Proc. Audio Engineering Society 65th Convention*, 1980.

[3] M. A. Gerzon. General metatheory of auditory localisation. In *92nd Audio Engineering Society Convention, Vienna*, 1992.

[4] Y Makita. On the directional localization of sound in the stereophonic sound field. *E.B.U Review*, A(73):102–108, 1962.

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

[6] Ville Pulkki. Compensating displacement of amplitude-panned virtual sources. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Jun 2002.

[7] M. Simon, D. Menzies, F.M. 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.

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