# A Complex Panning Method for Near-field Imaging

Dylan Menzies and Filippo Maria Fazi

***Abstract*—Conventional amplitude panning can be used to produce images of distant objects. A panning method is presented here that can also produce image cues for the near-field region, by the control of Inter-aural Level Difference cues in the low frequency range below 700Hz. A single first order filter is required for each image. The approach has grown from an adaptive panning method that corrects for the dependence of image direction on head orientation. A formulation is presented in a low frequency approximation. Cues are then calculated for various configurations using measured Head Related Transfer Functions. These results, and a listening test, confirm the ability of the method to control near-field cues, while also compensating for head rotation.**

***Index Terms*—IEEE, IEEEtran, journal, LaTeX, paper, template.**

## MATHEMATICAL SYMBOLS

| | |
|---|---|
| $k$ | wavenumber |
| $\omega$ | angular frequency |
| $Z_0$ | characteristic impedance of sound in air |
| $P$ | pressure of the incident free field at the head centre position |
| $P_L,\ P_R$ | resultant pressures at the ears |
| $\boldsymbol{V}$ | velocity vector of the incident free field at the head centre position |
| $\boldsymbol{V}_\Re$ and $\boldsymbol{V}_\Im$ | real and imaginary components of $\boldsymbol{V}$ |
| $\boldsymbol{r}_L,\ \boldsymbol{r}_R$ | displacement vectors from the head centre to each ear (bold type for vectors). |
| $\bar{\boldsymbol{r}}_R,\ \bar{\boldsymbol{r}}_L$ | $\bar{\boldsymbol{r}}_R = \frac{3}{2}\boldsymbol{r}_R,\ \ \bar{\boldsymbol{r}}_L = \frac{3}{2}\boldsymbol{r}_L$ |
| $\hat{\boldsymbol{r}}_I$ | direction to the image |
| $r_I$ | distance to image |
| $\boldsymbol{r}_V$ | Makita localisation vector ($\hat{\boldsymbol{r}}_V = -\hat{\boldsymbol{V}}$) |
| $\theta_{VN}, \theta_{IN}, \theta_{VI}$ | azimuth change from direction indicated by 1st subscript to 2nd. $V$ : $\boldsymbol{r}_V$, $N$ : nose, $I$ : image |
| $\theta_L$ | $2\theta_L$ is the stereo loudspeaker separation from the listener |
| $\theta_N$ | directed angle from the mid point between stereo loudspeakers to the direction the listener is facing in |
| $\theta_I$ | directed angle from the mid point between stereo loudspeakers to the image |
| $g_1, g_2$ | stereo panning gains |
| $\hat{\boldsymbol{r}}_1,\ \hat{\boldsymbol{r}}_2$ | direction vectors to the stereo loudspeakers from the listener |
| $h_{1L}, h_{1R}$ <br> $h_{2L}, h_{2R}$ | head related impulse responses from each loudspeaker (1,2) to each ear (L,R) |

## ABBREVIATIONS

| | |
|---|---|
| ITD | Inter-aural Time Difference |
| ILD | Inter-aural Level Difference |
| IPD | Inter-aural Phase Delay |
| VBAP | Vector Base Amplitude Panning |
| CAP | Compensated Amplitude Panning |
| KU100 | Neumann KU100 binaural microphone |
| KEMAR | KEMAR binaural microphone |
| HRIR | Head-Related Impulse Response |
| MAA | Minimum Audible Angle |

D. Menzies and F.M. Fazi are with the University of Southampton email: d.menzies@soton.ac.uk

## I. INTRODUCTION

In previous studies [1;2;3], the Inter-aural Time Difference (ITD) and Inter-aural Level Difference (ILD) localisation cues, were modelled for general low frequency sound fields. The model was then applied to the problem of panning reproduction, the creation of an image by the summation of coherent plane waves from multiple loudspeakers. If the head orientation is known, then the panning gains can be adjusted so that the resulting field produces the same localisation cues as a plane wave travelling from the target image direction. In this case the low frequency ILD is 0 dB. General stereo panning formula were found that include the listener's head orientation. Calculation with measured Head Related Transfer Functions predicts that this panning process produces stable images, for a wide range of configurations, making significant improvements over conventional panning processes that do not consider the head orientation. Furthermore, low frequency images can be generated in any direction, including behind, at the sides and above or below. This is because accurate ITD cues naturally lead to accurate dynamic ITD cues, which provide unambiguous imaging in all directions. The method is referred to as *Compensated Amplitude Panning* (CAP). A real-time implementation for the method has been constructed [1], including real-time head tracking to sense the position and orientation of the listener's head.

Brungart [4;5] performed a comprehensive set of listening experiments in anechoic conditions on the localisation of sources within 1 m from the listener, which he calls the *proximal region*. One conclusion is that the main cue for source proximity is ILD in the frequency region below 750 Hz. The contribution from the ILD as a distance cue reduces with increasing frequency, and becomes insignificant above 3 KHz. It appears that for high frequencies ILD is too saturated at a high value to allow the auditory system to discriminate source range, whereas the ILD variation at low frequencies is useable. Furthermore, range perception by ILD also depends on knowledge of head orientation. It may be that at low frequencies the ITD cue provides a more reliable and accurate source for orientation, which further enhances the effectiveness of ILD range cues in the low frequency region.

Studies have investigated the degrading effect of room reflections on ILD range cues from the direct signal [6;7]. These

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2827300, IEEE/ACM Transactions on Audio, Speech, and Language Processing



show that a precedence effect occurs such that the effect of reflections is largely suppressed. This is supported by common experience in which near-field imaging is largely unaffected by nearby surface reflections.

In this work the CAP formula are extended, by matching the localisation cues to those of a nearby point source, rather than a plane wave. By simultaneously controlling ILD and ITD in the low frequency region, the aim is to produce stable images with controllable direction and range in the near-field, while retaining the existing ITD correction depending on head orientation.

The remaining content is as follows. In Section II, the low frequency approximation for binaural signals is reviewed. These are used in Section III to find expressions for the localisation cues, ITD and ILD, in terms of a general field description and the listener's head orientation. In the previous work a relation was then found between a target image direction, the head orientation and the field description, by considering the cues for a plane wave travelling from the target image direction. Here this is extended by considering the cues for a near source positioned at the desired image location. The ILD for the near source varies with distance from the listener. In Section IV, extended stereo panning functions are derived, including 1st order filter components, by finding a panned field that produces the same cues that a real source would produce if it were located at the target image position. In Section V, the method is evaluated by calculating the cues using a spherical head model and measured HRTFs, focusing on the ILD cue for different head rotations. The direction cue provided by the ITD is also assessed for accuracy, since this may be reduced when compared with the existing CAP method for panning of far images. A supporting listening test is reported in Section VI. The results are discussed further in Section VII.

## II. Sound field representation

A source free sound field region can be expanded as a Taylor series about any point in the region[8]. 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 $\nabla P$ 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 distance $\|\boldsymbol{x} - \boldsymbol{x_0}\|$, 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 to $\boldsymbol{x_0}$. This condition can be written $kr << 1$ where $k = \frac{\omega}{c}$ is the wavenumber and $r = \boldsymbol{x} - \boldsymbol{x_0}$. Below around 700 Hz, a typical sound field region just large enough to enclose the human head satisfies these conditions.

The binaural signals are not equal to the pressures at the corresponding locations in the incident field, even at low frequency, due to the non-vanishing scattered field at the head surface. Using a spherical model for the head, with ears at antipodal locations, an analytical approximation can be found for the binaural signals[9;10;8;11]. The ITD is found to have a low frequency limit value of $\frac{3D}{2c}\cos\theta$, and constant ILD of 1 (or 0 in dB)[12], where $D$ is the ear separation, $c$ is the speed of sound, and $\theta$ is the angle separating the direction of the incident plane wave and the inter-aural axis. The free-field ITD is the ITD measured at the ear locations in the absence of the head, and is $\frac{D}{c}\cos\theta$. The resultant pressure field on the sphere, radius $r = D/2$, is equal to the pressure of the incident free-field plane wave on the corresponding sphere at radius $\frac{3}{2}r$ [12]. Since any field in a free-field region can be approximated arbitrarily well by a plane wave expansion[13], then for any incident field at low frequency, including near source fields, the resultant pressure at the surface of a rigid sphere of radius $r$ can be approximated by the incident field, evaluated at radius $\frac{3}{2}r$, provided it is homogenous up to this radius.

With the above considerations applied to a general first order field, the binaural signals at the right and left ears can be approximated as

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

where $P$ and $\nabla P$ are the pressure and gradient of the incident field at the central point between the ears, in the centre of the head, see Fig. 1. $\bar{\boldsymbol{r}}_R = \frac{3}{2}\boldsymbol{r}_R$ and $\bar{\boldsymbol{r}}_L = \frac{3}{2}\boldsymbol{r}_L$, where $\boldsymbol{r}_R, \boldsymbol{r}_L$ are displacement vectors from the centre to the right and left ears. The gradient is related to particle velocity $\boldsymbol{V}$ by Euler's

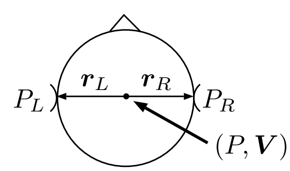



Fig. 1: Sound field variables and vectors in relation to the listener's head.

equation in the frequency domain[14],

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

using the positive frequency convention so that the actual pressure field is $P(\boldsymbol{x}, t) = \Re(P(\boldsymbol{x})e^{j\omega t})$, where $\omega$ is the angular frequency. $Z_0$ is the characteristic impedance of sound in air[8]. The low frequency approximation condition is $kr_R << 1$.

In the following discussion only the relative phases of the binaural signals are of interest, so without loss of generality and in order to simplify calculation the pressure and velocity phases are rotated by the same amount so that pressure is real valued,

$$(P, \boldsymbol{V}) \to \frac{P^*}{|P|}(P, \boldsymbol{V}) \tag{5}$$

where $P^*$ is the complex conjugate of $P$. In the following it is assumed $P$, $\boldsymbol{V}$ and derived quantities $P_L$, $P_R$ have been phase rotated. The real and imaginary vector components of $\boldsymbol{V}$ are written $\boldsymbol{V}_\Re$ and $\boldsymbol{V}_\Im$ where $\boldsymbol{V} = \boldsymbol{V}_\Re + j\boldsymbol{V}_\Im$. Using $\boldsymbol{r}_L = -\boldsymbol{r}_R$ and Euler's equation (4), the binaural signals, (2) and (3), can be written as

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

Fig. 2 illustrates the general case using the complex plane. Both $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$ are non-zero, and in different directions. As the listener's head rotates around any axis, $P_R$ and $P_L$ move around on opposite sides of an ellipse, shown with the dashed line. This is because $\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_{\Re} = \bar{\boldsymbol{r}}_R \cdot \bar{\boldsymbol{V}}_{\Re}$ and $\bar{\boldsymbol{r}}_R \cdot \boldsymbol{V}_{\Im} = \bar{\boldsymbol{r}}_R \cdot \bar{\boldsymbol{V}}_{\Im}$, where $\bar{\boldsymbol{V}}_{\Re}$ and $\bar{\boldsymbol{V}}_{\Im}$ are $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$ projected in any plane to which the rotation axis is normal. $\bar{\boldsymbol{r}}_R \cdot \bar{\boldsymbol{V}}_{\Re}$ and $\bar{\boldsymbol{r}}_R \cdot \bar{\boldsymbol{V}}_{\Im}$ each vary as sinusoids of the angle $\boldsymbol{r}_R$ with any reference direction in the plane, possibly with different amplitudes and phases, resulting in an ellipse.

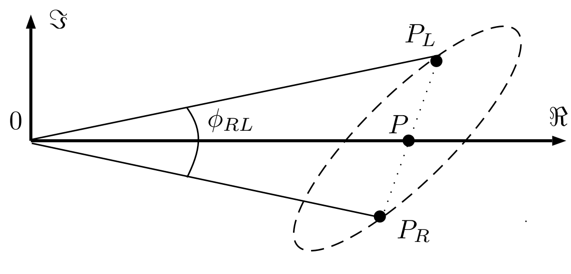



Fig. 2: $P_R$ and $P_L$ in the complex plane for non zero and non-aligned $\boldsymbol{V}_{\Re}$ and $\boldsymbol{V}_{\Im}$

An omnidirectional point source is used to model a source in the near-field region of the listener. The velocity and pressure components are calculated as follows. The pressure field of a point source is

$$P = \frac{A}{r_S} e^{-jkr_S} \tag{8}$$

where $A$ is the complex amplitude of the source and $r_S$ is the distance from source. The gradient is

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

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

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

The complex velocity components are then

$$\boldsymbol{V}_{\Re} = \frac{P}{Z_0}\hat{\boldsymbol{r}}_{\boldsymbol{S}} \tag{11}$$

$$\boldsymbol{V}_{\Im} = -\frac{P}{Z_0 kr}\hat{\boldsymbol{r}}_{\boldsymbol{S}} \tag{12}$$

Note that $\hat{\boldsymbol{V}}_{\Re} = -\hat{\boldsymbol{V}}_{\Im} = \hat{\boldsymbol{r}}_S$, the direction from the source to the listener. If the source is distant then as noted earlier the linear approximation is valid provided $kr_R << 1$. However this is not sufficient if the source comes close. For example the approximation can fail completely when $r_S = r_R$ and improves as $r_S$ increases. When the source is on the inter-aural axis and in the limit of low frequency it is straight forward to show that the relative error of the linear approximation for $P$ is $-(r_R/r_S)^2$. This is $-1/16$ when $r_S = 4r_R$, ie when the source is approximately 30cm from the head centre. The final effect of the approximation on reproduced cues will be become clearer later when the complex panning method is evaluated for various $r_S$.

## III. LOCALISATION CUES AND REPRODUCTION

The localisation cues can be calculated from the binaural signals that were approximated in the previous section. For a single frequency $\omega$, the ITD cue is equal to $\phi_{RL}/\omega$ (also called the Inter-aural Phase Delay (IPD) in this context), where $\phi_{RL} = \arg(P_R/P_L)$ is the inter-aural phase difference[15], shown in Fig. 2. The ILD cue 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 now the problem of finding, for any target point source, all possible fields that evoke the same perceived image as this target. This is equivalent to finding fields giving rise to the same ITD and ILD as the target, disregarding other localisation cues. A trivial solution is given by the point source field itself. Another solution is provided by matching pressure and velocity, since in the low frequency approximation the pressure and velocity determine the ITD and ILD. The Ambisonic reproduction method employs this principle[16].

There remains the possibility of pressure and velocity solutions that give the correct ITD and ILD but do no match the target field pressure and velocity. To find these first observe that the ITD and ILD depend on, and determine, the shape (but not the size) of the triangle $\triangle 0 P_L P_R$ in Fig. 2. So if two given fields have triangles with the same shape, then they produce the same image, according to the cues considered.

So, given a desired image, consider a target point source field that would produce this image, with velocity and pressure components $\boldsymbol{V}_{\Re}$, $\boldsymbol{V}_{\Im}$ and $P$, described by (11) and (12). Let $\tilde{\boldsymbol{V}}_{\Re}$, $\tilde{\boldsymbol{V}}_{\Im}$ and $\tilde{P}$ be the components of a general field with ITD and ILD that match those of the target field. The shape of the triangle shown in Fig. 2 is determined by two ratios, $\Re(P_L - P)/P$ and $\Im(P_L - P)/P$ (equally $P_R$ could be used). Evaluating these separately for the general field and the target point source field, by substituting for $\boldsymbol{V}_{\Re}$, $\boldsymbol{V}_{\Im}$ and $P$ in (6) and (7) gives,

$$\frac{\Im(P_L - P)}{P} = \begin{cases} k\bar{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{V}}_{\Re}, & \text{target field} \\ kZ_0\bar{\boldsymbol{r}}_R \cdot \tilde{\boldsymbol{V}}_{\Re}/\tilde{P}, & \text{general field} \end{cases} \tag{13}$$

$$\frac{\Re(P_L - P)}{P} = \begin{cases} \bar{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{V}}_{\Re}/r, & \text{target field} \\ -kZ_0\bar{\boldsymbol{r}}_R \cdot \tilde{\boldsymbol{V}}_{\Im}/\tilde{P}, & \text{general field} \end{cases} \tag{14}$$

Equating the two parts for each ratio gives two conditions for matching the cues,

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Re(\boldsymbol{r}_V) - \hat{\boldsymbol{r}}_I) = 0 \tag{15}$$

$$\hat{\boldsymbol{r}}_{\boldsymbol{R}} \cdot (\Im(\boldsymbol{r}_V) + \frac{\hat{\boldsymbol{r}}_I}{kr_I}) = 0 \tag{16}$$

where $\hat{\boldsymbol{r}}_I$ and $\boldsymbol{r}_V$ are defined for convenience as follows: $\hat{\boldsymbol{r}}_I$ is the direction from the listener to the image, which is opposite to the direction of the real part of the target velocity,

$$\hat{\boldsymbol{r}}_I = -\hat{\boldsymbol{V}}_{\Re} \tag{17}$$

$r_I = r$, the distance between the listener and the image. $\boldsymbol{r}_V$ describes the general field, and coincides with the *Makita Localisation Vector* given by Gerzon[16], and extended for complex values,

$$\boldsymbol{r}_V = -\tilde{\boldsymbol{V}} Z_0/\tilde{P} \tag{18}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2827300, IEEE/ACM Transactions on Audio, Speech, and Language Processing



$\boldsymbol{r}_R$ has been replaced with the unit vector in the same direction, $\hat{\boldsymbol{r}}_{\boldsymbol{R}}$. The size of the listener's head represented by $r_R$ is not needed for finding solutions for the general field.

It is apparent from (15) and (16) that, given an image location described by $\hat{\boldsymbol{r}}_I$ and $r_I$, and a head orientation, which determines $\hat{\boldsymbol{r}}_R$, there is a continuous set of $\boldsymbol{r}_V$ and corresponding fields that produce the desired image. The vector diagram in Fig. 3 shows a projection from above the listener. Possible values of $\Re(\boldsymbol{r}_V)$ and $\Im(\boldsymbol{r}_V)$ lie on the two planes normal to the page and passing through the two straight dotted lines. The full set of consistent values for $\hat{\boldsymbol{r}}_{\boldsymbol{I}}$ forms a circle around the inter-aural axis in the plane for $\Re(\boldsymbol{r}_V)$, commonly known as the *cone of confusion*. The effect of head orientation on image direction was discussed in detail previously[1].

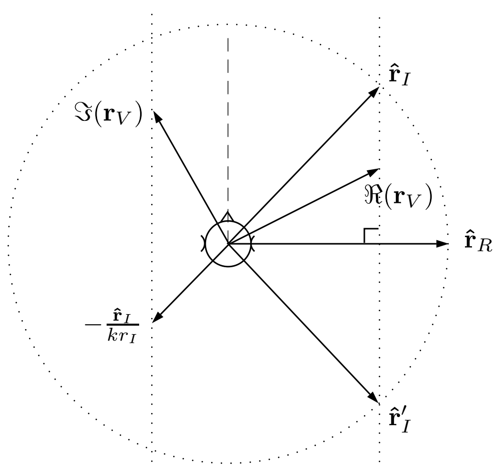



Fig. 3: Geometric relationship between the inter-aural direction $\hat{\boldsymbol{r}}_{\boldsymbol{R}}$, image direction $\hat{\boldsymbol{r}}_{\boldsymbol{I}}$ and the real and imaginary parts of the Makita vector $\boldsymbol{r}_{\boldsymbol{V}}$

So, in conclusion for this section, necessary and sufficient conditions have been found, (15) and (16), relating the head orientation, 1st order field description, and image location. These will be used in the next section to design an complex panning technique for loudspeakers.

## IV. Complex Panning

Amplitude panning is a spatial audio reproduction method in which several loudspeakers produce plane waves converging at the listener position in phase. For each image, the signals driving the loudspeakers are produced by multiplying the source signals by a set of real-valued gains.

In the following, $g_i$ is defined as the gain for the $i$-th loudspeaker. Also, $\boldsymbol{V}_i$ and $P_i$ are defined as the velocity and pressure at the central head position, and $\hat{\boldsymbol{r}}_i$ is the direction vector from the centre to the $i$-th loudspeaker. $\hat{\boldsymbol{r}}_i = -\hat{\boldsymbol{V}}_i$. Possibly additional delay compensation is required for some driving signals to compensate for different values $\boldsymbol{r}_i$, so that the waves all arrive in phase. The pressure and velocity at the listener are then given by the sum of the contributions from each loudspeaker, so that

$$\boldsymbol{r}_V = -Z_0\frac{\boldsymbol{V}}{P} = -Z_0\frac{\sum \boldsymbol{V}_i}{\sum P_i} = -\frac{\sum P_i\hat{\boldsymbol{V}}_i}{\sum P_i} = -\frac{\sum g_i\hat{\boldsymbol{V}}_i}{\sum g_i} = \frac{\sum g_i\hat{\boldsymbol{r}}_i}{\sum g_i} \tag{19}$$

This is in agreement with the original definition of Makita Localisation Vector in terms of panning gains[16]. In this section, complex values are allowed for the gains and $\boldsymbol{r}_V$, in order to control ILD in addition to ITD. The term *complex panning* is used to distinguish from sound field reproduction methods, such as Ambisonics that may also use complex valued gains.

The near-field panning problem is to find panning gains $\{g_i\}$ that produce a given image direction $\hat{\boldsymbol{r}}_I$ and range $r_I$ for a given head direction $\hat{\boldsymbol{r}}_R$. A solution is first sought for two loudspeakers, since this is the most restrictive case. To simplify calculation the following additional constraint is imposed,

$$g_1 + g_2 = 1 \tag{20}$$

This does not restrict the solution search in practice since any unconstrained solution $(\tilde{g}_1, \tilde{g}_2)$, is a multiple of a constrained solution $(g_1, g_2) = (\tilde{g}_1, \tilde{g}_2)/(\tilde{g}_1 + \tilde{g}_2)$. (20) also normalises the pressure $P$ at the listener to a constant value, so that the perceived low frequency volume can be controlled using an additional scaling factor applied to the gains.

The new constraint (20) allows (15) and (16) to be written in terms of separate sets of variables $(\Re(g_1), \Re(g_2))$ and $(\Im(g_1), \Im(g_2))$. Splitting the constraint into real and imaginary parts,

$$\Re(g_1) + \Re(g_2) = 1 \tag{21}$$
$$\Im(g_1) + \Im(g_2) = 0 \tag{22}$$

hence from (19),

$$\Re(\boldsymbol{r}_V) = \Re(g_1\hat{\boldsymbol{r}}_{\boldsymbol{1}} + g_2\hat{\boldsymbol{r}}_{\boldsymbol{2}}) = \Re(g_1)(\hat{\boldsymbol{r}}_{\boldsymbol{1}} - \hat{\boldsymbol{r}}_{\boldsymbol{2}}) + \hat{\boldsymbol{r}}_{\boldsymbol{2}} \tag{23}$$
$$\Im(\boldsymbol{r}_V) = \Im(g_1\hat{\boldsymbol{r}}_{\boldsymbol{1}} + g_2\hat{\boldsymbol{r}}_{\boldsymbol{2}}) = \Im(g_1)(\hat{\boldsymbol{r}}_{\boldsymbol{1}} - \hat{\boldsymbol{r}}_{\boldsymbol{2}}) \tag{24}$$

Substituting (23) in (15) gives the real gains, which were calculated before[2] .

$$\Re(g_1) = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \tag{25}$$
$$\Re(g_2) = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_1)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_2 - \hat{\boldsymbol{r}}_1)} \tag{26}$$

The imaginary gains are similarly found by substituting (24) in (16),

$$\Im(g_1) = -\Im(g_2) = -\frac{\hat{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{r}}_I}{kr_I\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \tag{27}$$

The common factor in the denominators, $\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)$, indicates that the panning process fails when the loudspeakers are on the left or right sides, opposite the inter aural axis. The image instability in this configuration is well known[16;17], and was one of the motivations in the early development of Ambisonics. The formulation shows explicitly there is no solution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2827300, IEEE/ACM Transactions on Audio, Speech, and Language Processing



The imaginary part of the gains, $j\Im(g_i)$ are frequency dependent and proportional to $1/(jk)$, and so represent an integrating filter. This can be implemented in a real-time system with a simple one-pole low-pass filter with transfer function[18]

$$H(z) = \frac{b}{1 + az^{-1}} \tag{28}$$

By choosing $a = -1 + \epsilon$ for small $\epsilon > 0$ the cutoff frequency can be adjusted to the low end of the ILD localisation range $\approx 150Hz$. The phase response then rises linearly from near $-\pi/2$ at the cutoff frequency up to 0 at Nyquist, however the region of interest up to $\approx$ 700 Hz is well below Nyquist with typical sampling rates $\approx$ 40000 Hz, and so the phase is approximately $-\pi/2$ there. $b$ is then adjusted so that the amplitude response is close to $1/k$ in the region of interest.

The loudspeaker gains are calculated without consideration for the overall distance attenuation from image source to listener. The gain of the free-field pressure at the centre-head position to the source signal is proportional to the sum of loudspeakers gains $g_1 + g_2$. Since the gains were derived assuming $g_1 + g_2 = 1$, this implies the pressure gain is constant. Therefore to vary the loudspeaker gains to account for source distance an additional factor of $1/r_I$ should be introduced to the source prior to panning. Low frequency diffraction is included in the model so we expect that spectral changes are approximated, including a relative increase in bass level at the ipsilateral ear as the distance of the source to the listener reduces, described by Brungart[4]. It would be possible to further equalise the source signals, depending on image distance, prior to panning, in order to improve the overall binaural response. This is not developed in further detail here.

For high frequencies the normalisation used for low frequencies is not valid. An energy based model is more appropriate[19] with normalisation

$$\Re(g_1)^2 + \Re(g_2)^2 = 1 \tag{29}$$

Before applying this normalisation the common factor $\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)$ in (25), (26) and (27) can be discarded. To simplify the system, the energy normalisation can be applied across the whole spectrum, as in VBAP, although equalisation error will result.

## V. SIMULATIONS

In this section, the complex panning formulae (25), (26), (27), are tested by comparing the cues produced by panning with those for a source located at the target image location, using a range of target distances. This process is applied first using the spherical head model, and then using Head Related Impulse Responses (HRIRs). The HRIRs used are from the Technical University, Cologne[20;21], which were measured with a Neumann KU100 for several source distances.

There are many listening configurations that can be tested. The focus here is on the calculation of ILD for various target image distances, and head angles relative to the image. ITD is also calculated, to check if this is adversely affected by image distance control. A more comprehensive investigation of ITD and image error for far image targets is provided in a previous article[1].

Responses for the spherical head model are calculated directly at the low frequency limit, denoted by 0 Hz, based on the observations of Section II. Loudspeaker sources are modelled as plane waves, while the target source is modelled as a point source. In each case the response at the point $\boldsymbol{r}$ is calculated by evaluating the freefield at $\frac{3}{2}\boldsymbol{r}$. The frequency responses for the measured HRIRs are found by evaluating the z-transform $H(z)$ of each impulse on the unit circle, $H(e^{j\omega})$, for each required angular frequency $\omega$ (using *freqz* in Matlab). For the loudspeaker responses the HRIR used was for a source distance of 3.25 m, as before.

The binaural responses $H_L, H_R$ for the panning simulation are calculated by mixing the individual responses using the complex gains $g_1$ and $g_2$, from (25), (26) and (27),

$$H_L = g_1 H_{1L} + g_2 H_{2L} \tag{30}$$

$$H_R = g_1 H_{1R} + g_2 H_{2R} \tag{31}$$

When using measured HRIRs, $H_L$ and $H_R$ accurately predict the responses that would be measured in a real panning experiment using the binaural microphone. The only assumption is linearity. Second order effects from mutual loudspeaker scattering can be ignored.

The binaural phase difference is $\phi_{RL} = \arg(H_R/H_L)$. If necessary the phase is unwrapped beyond $\pm\pi$ to accommodate the spatial aliasing at the top end of the ITD range $700 - 1500Hz$. The ITD and IPD at this frequency are then given by $\phi_{RL}/\omega$.

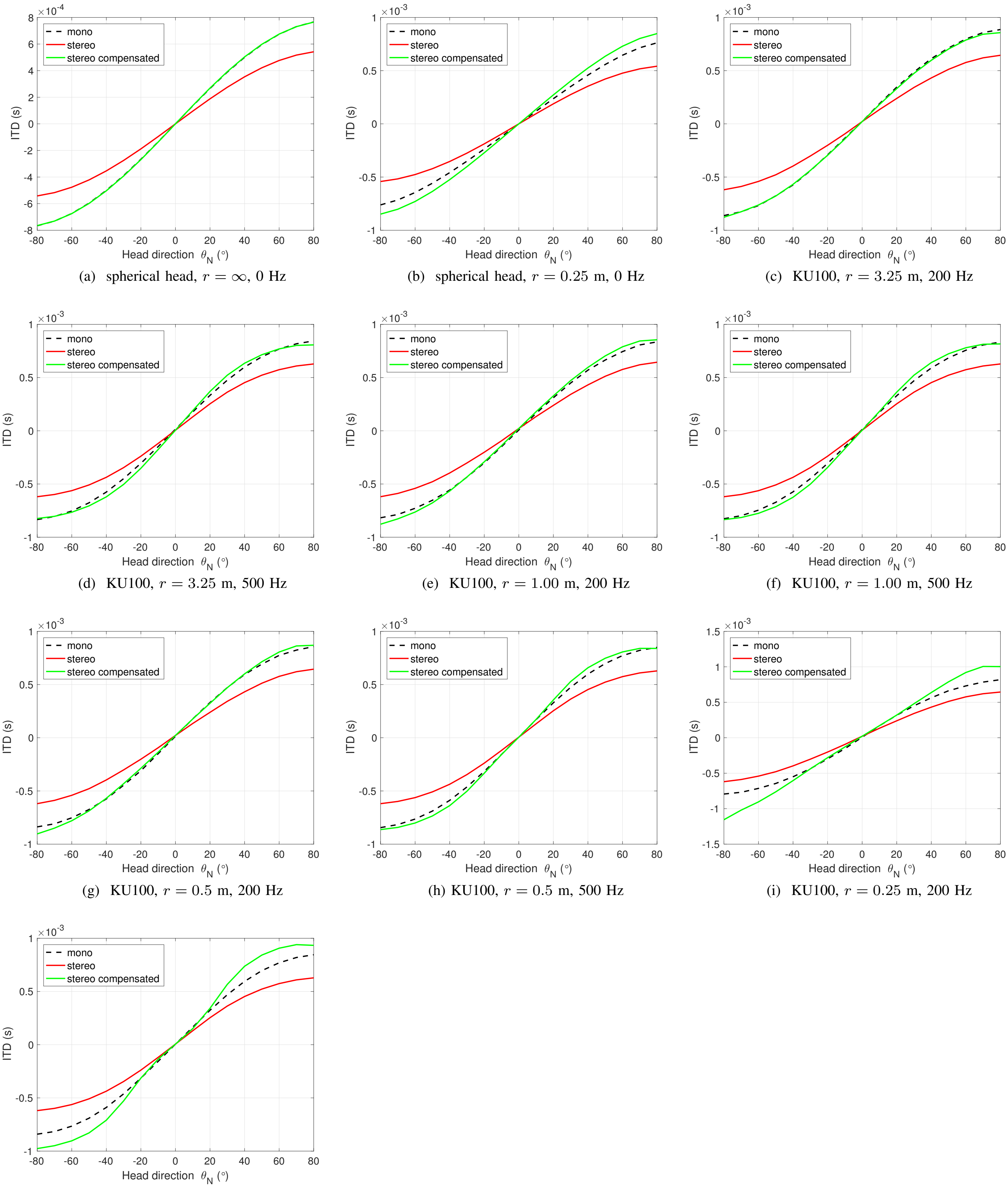
In the simulations the following parameters are specified: the angle $\theta_L$ of each loudspeaker from the centre direction, the target image angle $\theta_I$ from the centre in the horizontal plane, the target image distance from the listener $r_I$, and the frequency. Unless otherwise stated the target image is central, $\theta_I = 0$, and the loudspeakers are positioned symmetrically with $\theta_L = 45°$. The ITD and ILD cues are plotted against $\theta_N$ which is the change in horizontal angle from the central direction midway between the loudspeakers to the direction the listener is facing. The geometrical parameters are illustrated in Fig. 6.

Each plot shows three graphs. *mono* shows the ITD and ILD for a single source in the target image direction. The response is based on the relative angle between the listener's head and the target image. *stereo* shows the ITD and ILD when the image is panned non-adaptively. The static gains are given by the case where the head points towards the target, which is equivalent to tangent law panning. The ITD in this case is expected to be near 0 s when the listener is facing the target image direction. The ILD should be 0 dB across the range for distant loudspeakers because the field at the listener is fixed and $\boldsymbol{r_V}$ is real valued. *stereo compensated* shows the ITD and ILD when the image is panned adaptively, with $\hat{\boldsymbol{r}}_R$ varied according to the listener direction indicated by the horizontal axis.

For a given head direction $\theta_N$ the angular error of the uncompensated or compensated stereo images is given by $\theta_E = \theta_N - \theta'_N$, where $\theta'_N$ is the head angle which gives the same ITD on the mono graph, as on the stereo graph in question. This can be read of the plots directly as a horizontal segment between the mono and stereo graphs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2827300, IEEE/ACM Transactions on Audio, Speech, and Language Processing






(a) spherical head, $r = \infty$, 0 Hz

(b) spherical head, $r = 0.25$ m, 0 Hz

(c) KU100, $r = 3.25$ m, 200 Hz

(d) KU100, $r = 3.25$ m, 500 Hz

(e) KU100, $r = 1.00$ m, 200 Hz

(f) KU100, $r = 1.00$ m, 500 Hz

(g) KU100, $r = 0.5$ m, 200 Hz

(h) KU100, $r = 0.5$ m, 500 Hz

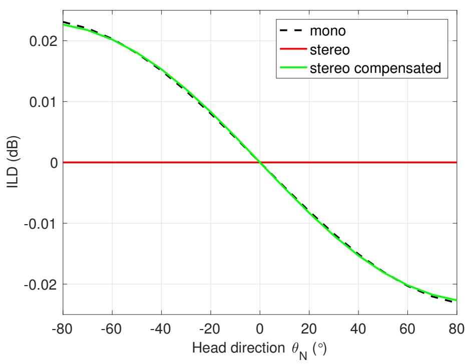
(i) KU100, $r = 0.25$ m, 200 Hz

(j) KU100, $r = 0.25$ m, 500 Hz

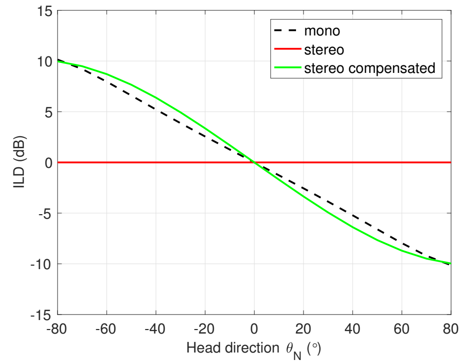
Fig. 4: Simulated ITD vs head direction, for the spherical head model and measured head Neumann KU100. Target image $\theta_I = 0^o$, loudspeaker angles: $\theta_L = 45^o$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2827300, IEEE/ACM Transactions on Audio, Speech, and Language Processing

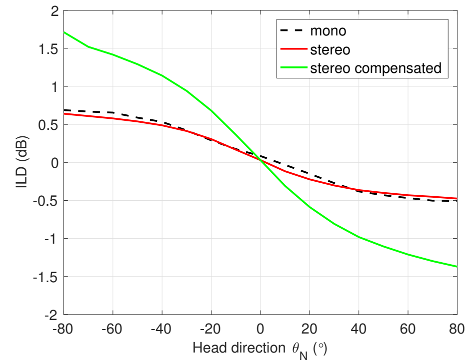




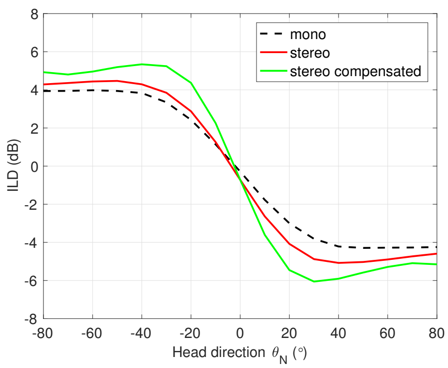

(a) spherical, $r = \infty$, 0 Hz



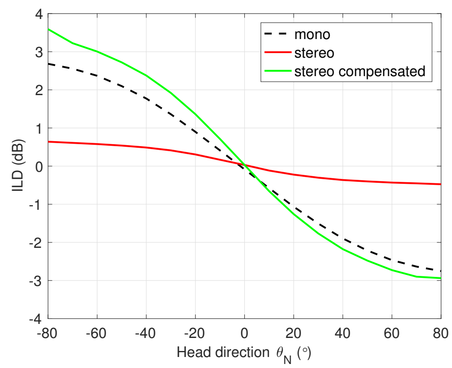

(b) spherical, $r = 0.25$ m, 0 Hz



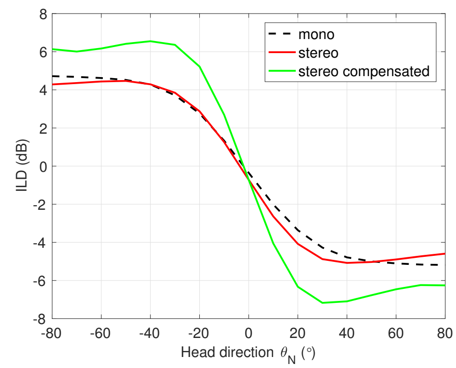

(c) KU100, $r = 3.25$ m, 200 Hz



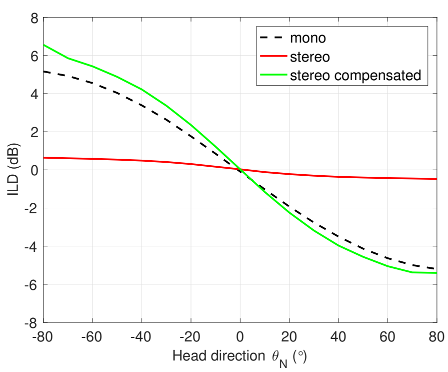

(d) KU100, $r = 3.25$ m, 500 Hz



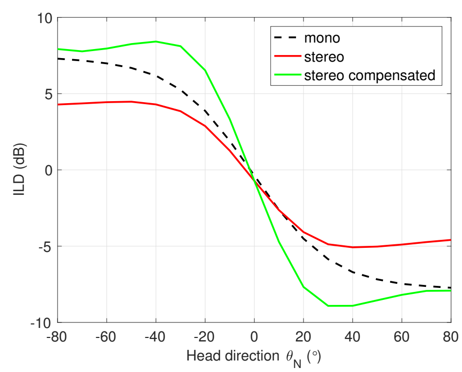

(e) KU100, $r = 1.00$ m, 200 Hz



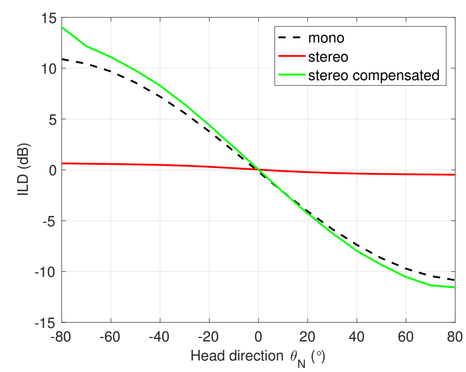

(f) KU100, $r = 1.00$ m, 500 Hz



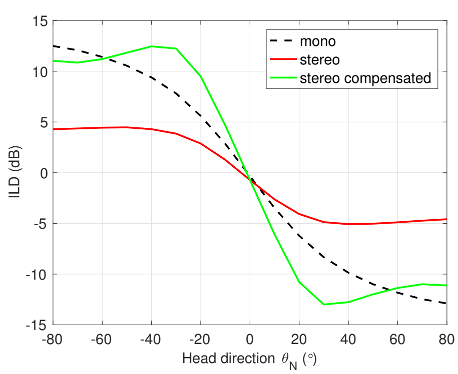

(g) KU100, $r = 0.5$ m, 200 Hz




(h) KU100, $r = 0.5$ m, 500 Hz




(i) KU100, $r = 0.25$ m, 200 Hz




(j) KU100, $r = 0.25$ m, 500 Hz

Fig. 5: Simulated ILD vs head direction, for the spherical head model and measured head Neumann KU100. Target image $\theta_I = 0^o$, loudspeaker angles: $\theta_L = 45^o$.

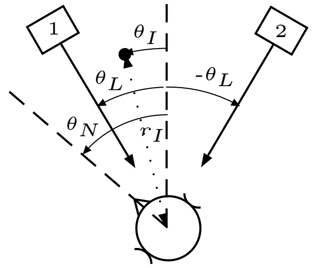



Fig. 6: Geometry for simulated stereo listening. Shown are $\theta_L$ the angle from the centre line to each loudspeaker, $\theta_I$ the angle to the image, $r_I$ the target image distance, and $\theta_N$ the angle the listener faces in. In this example all variables are positive.

The directional accuracy of the adapted panning is assessed by first looking at where the uncompensated error is significant, where it exceeds the Minimum Audible Angle (MAA)[22]. The uncompensated error for a given head direction is read by taking the horizontal angle gap between the solid red (light) line at the value of the head direction, and the the black line. MAA is fairly constant over the ITD frequency range, up to 1000 Hz. For lateral horizontal image displacements of 0°, 30°, 60° and 75°, representative MAAs in the ITD range are respectively 1°, 1.5°, 3°,7°. MAA is measured under static head conditions, which suppresses the ITD localisation process. Variants of MAA under moving source and self moving head conditions show greater sensitivity to angular changes,[23;24]. Even these may not capture the full sensitivity of the ITD localisation process. The comparisons are made with the more conservative MMA parameter.

Compared with directional localisation there has been relatively little research to quantify the resolution of near-field distance localisation and its relation to low frequency ILD. Some studies have investigated the just noticeable difference (JND) for ILD[25;26]. Hafter reports the ILD JND using a low frequency click train, with energy from 0.1-2K Hz, across the ILD range 0 to 24 dB and found JNDs were typically in the range 1-2 dB with slight increase towards higher ILD.

The first set of plots check that the complex panning method behaves as expected using the spherical head in the low frequency limit. A distant target $r_I = \infty$ m is simulated using adaptive panning without distance control. For a distant target the compensated ITD and ILD should match the target, which is observed, see Figs. 4a, 5a. In both cases the uncompensated stereo case deviates significantly from the target over much of the head rotation range. The uncompensated stereo has constant 0 dB ILD, as expected.

For a target at 0.25 m the compensated ITD and ILD match the target less well than before, but still within the ITD MMA and ILD JND, see Figs. 4b, 5b. This is expected because compared with the velocity, the higher derivatives of the field become significant for a near source, even in the low frequency limit.

Using the measured HRIRs, plots were made for image distances 3.25 m, 1.0 m, 0.5 m, 0.25 m, for frequencies 200 Hz, 500 Hz. Note the ILD and ITD scales are variable, in order to display the best resolution in each case. The HRIRs used for the loudspeakers are for a distance of 3.25 m, the longest available, but less than $\infty$ m assumed by the panning method.

Figs. 4c, 5c, 4d, 5d show the case for $r_I = 3.25$ m. The compensated stereo ITD matches well at both frequencies. The ILD is within a JND over the whole angular range, for both frequencies (in Fig. 4c the error looks worse because the scale is different in each plot) . However the ILD is clearly worsened by the proximity of the loudspeakers. This can be mostly corrected by changing the target distance to $r_I = \infty$ m. Adaptation does not improve the ILD, but is needed so that the ITD is corrected.

For $r_I = 1.0$ m, Figs. 4e, 5e, 4f, 5f, the compensated ITD and ILD are sufficiently accurate, according to JND, although the ILD accuracy at 500 Hz is marginal. The ILD at 200 Hz clearly benefits from compensation. There is a similar picture for $r_I = 0.5$ m, Figs. 4g, 5g, 4h, 5h.

For $r_I = 0.25$ m, Figs. 4i, 5i, 4j, 5j the compensated ITD is accurate over at least the range ±30°, for both frequencies. The compensated ILD is accurate over nearly the whole range at 200 Hz. At 500 Hz there are sections where the error exceeds the JND, although the ILD is improved by compensation over the whole range.

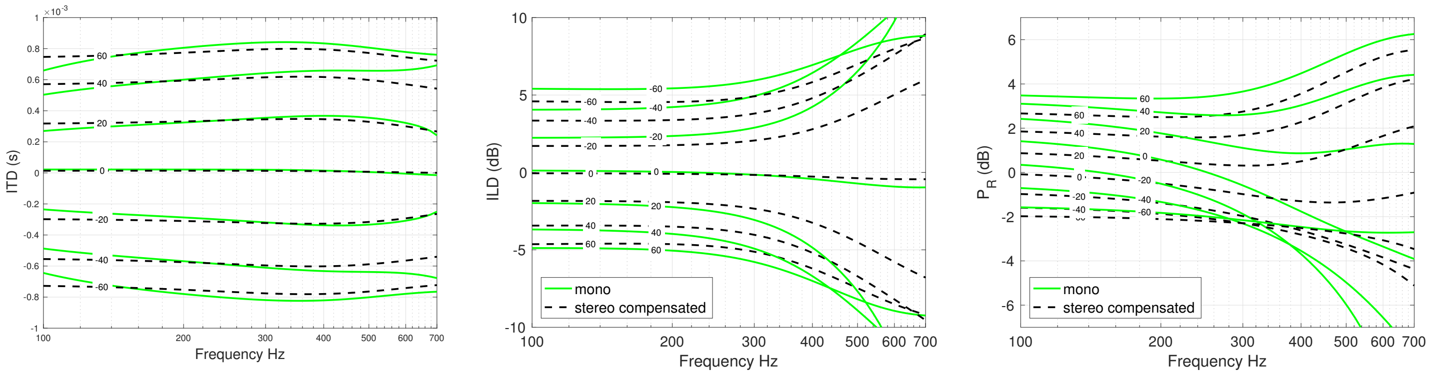
The previous results are evaluated for two frequencies. In the following, the mono and stereo compensated cases are compared across a frequency range of 100 Hz ? 700 Hz, for several head positions. In each case the target image is straight ahead, $\theta_I = 0^o$, at a range $r = 0.5$m. The HRIRs used for the stereo loudspeakers have $r = 3.25$m, the longest range available.

Fig. 7a shows that the ITD of the compensated reproduction matches the mono source very well over the whole frequency range, according to the criteria already established in this section. The compensated ITD even follows the deviations of the mono source ITD, although in the low frequency model the ITD is constant. The compensated ILD shown in Fig. 7b is within JND below 400 Hz, then the error increases towards 700 Hz. The slight elevation in the ILD at lower frequencies can be explained by the stereo loudspeakers being closer than infinity, which is the range assumed by the CAP method. The ILD asymmetry displayed across the head rotations would suggest errors in the original HRIR measurement process, since the binaural head itself is manufactured to be precisely symmetrical. This highlights the importance of precise measurement in this kind of study.

The right side binaural pressure signal is shown in Fig. 7c.

## VI. Listening Test

A real-time stereo system was constructed that implements the near-field panning method in an experimental setup, by extending an existing system[2], and including the filter described, (28). Fig. 8 shows a schematic of the overall system. The listener's head position is tracked with a Microsoft Kinect, and the orientation using a wireless Sparkfun Razor IMU device attached to a cap. Image objects can be placed anywhere in space, and the listener is free to move relative to these and




(a) (b) (c)

Fig. 7: Binaural parameters plotted against frequency for different head angles $\theta_N$ (indicated). a) ITD, b) ILD, c) $P_R$ (pressure at right ear). Simulations using measurements from the Neumann KU100 binaural microphone. Target image $\theta_I = 0^o$, $r = 0.5$m, loudspeaker angles: $\theta_L = 45^o$. Mono reproduction (dashed) and compensated stereo reproduction are shown.

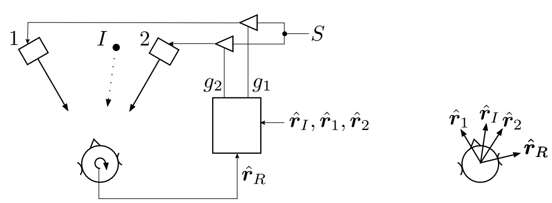



Fig. 8: Overview of a system implementing CAP for stereo, including adaptation for head rotation, shown for one object image. The signal from the head track is shown, and the input audio signal $S$.

listen from different positions. In informal listening it was immediately clear that the system greatly improves the perception of near images. A listening experiment was conducted with 10 subjects. All had prior experience with audio testing, reported normal hearing, and were involved as students or professionals with audio. The stereo loudspeakers were set at standing head height, and 2m apart. Two conditions were presented, with and without the near-field gain component present (27). In each condition a moving image was presented moving slowly from left to right and back, 1m in front of the loudspeakers, and with 4m travel. A repeating tuba sound was used with 99% energy below 1000 Hz, and 50% of energy below 280 Hz. The listening conditions are made as natural as possible, with source and listener movement, so that the ILD cue is represented in a complete context, and the comparison with existing range cues can be made clearly.

Each subject initially stood 2m in front of the loudspeakers and is then free to move around and observe the moving image. The subject was asked two questions: first to select which condition provided the most plausible reproduction of a source moving in the target trajectory, and second to select which condition gave the best impression of source proximity. For both questions, the near-field enabled system was selected in every case. Using a binomial significance test this yields a p-value $p = 0.01$ for both questions, which strongly suggests that the near-field algorithm is improving the near-field reproduction. A typical comment by a subject was 'Wow, that passed right by me'.

## VII. Discussion

The earlier stereo CAP system[1] was able to provide simple range cues, including level change and motion parallax that depend on the relative motion of the user and the sources. ILD range cues produced using the new method were found to make near images much more plausible. The computational cost is only slightly greater than that for the CAP method for far-field images, as the added filter requires at most two multiplies.

The simulation results show that the complex panner performs well within just noticeable thresholds, for both ILD and ITD cues, up to 200 Hz, and for target image distances as close as 0.25 m. The spherical head ILD results for 0.25 m are surprisingly good considering the approximation used is not exact in this case. At 500Hz the ILD cues are less accurate, but not confusing. According to prior studies[4,5] the distance discrimination that can be provided by the mono ILD cue at 500 Hz is significantly reduced compared with 200 Hz. Evidence for this can be seen here by comparing the difference in maximum mono ILD at 3.25 m and 1 m, for 200 Hz and 500 Hz.

The algorithm is based on the spherical head model in the low frequency limit. This could possibly be refined in some ways without greatly increasing complexity or computational cost. For example, the model itself could be improved for example by allowing for ears that are off centre, or approximating more closely the response at higher frequencies. For a more complex model it may however be difficult to find a tractable solution for the gains. Correct equalisation is ensured in the low frequency limit by the pressure condition (20). The equalisation deviates slightly at at higher frequencies. Although acceptable, this could be improved by adding compensating equalisation after the source input stage.

The loudspeakers are assumed to be distant. It was found that near loudspeakers can be compensated partly by increasing the image distance. A more systematic compensation process maybe possible by extending the conditions. The overall ILD match could be improved by warping the image distance control $r_I$, without distorting the ITD significantly. It may be useful to exaggerate the ILD cue, for example for special effect.

## VIII. Summary

A complex panning method was described, based on the reproduction of low frequency cues for image direction and range. The ITD and ILD cues are simultaneously controlled according to the directions of the image and loudspeakers, and the orientation of the listener's head, without knowledge of head size. Simulations with measured HRIRs show that, considering frequencies representing the core ILD frequency range, stable images can be produced up to 0.25 m range. The initial stereo implementation is being extended to a flexible multi-channel framework, as part of a flexible object-based system.

## IX. Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) "S3A" Programme Grant EP/L000539/1, and the BBC Audio Research Partnership. No new data was created in this work.

## References

[1] D. Menzies and F. M. Fazi, "A low frequency panning method with compensation for head rotation," *IEEE Trans. Audio, Speech, Language Processing*, 2017 submitted.

[2] ——, "A theoretical analysis of sound localisation, with application to amplitude panning," in *Proc. AES 138th Convention, Warsaw*, May 2015.

[3] F. M. 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. S. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1101–1109, 2001.

[5] ——, "Auditory localization of nearby sources. iii. stimulus effects," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3589–3602, 1999.

[6] H. Gaskell, "The precedence effect," *Hearing research*, vol. 12, no. 3, pp. 277–303, 1983.

[7] R. M. Dizon and H. S. Colburn, "The influence of spectral, temporal, and interaural stimulus variations on the precedence effect," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2947–2964, 2006.

[8] P. Morse and K. Ingard, *Theoretical Acoustics*. New York, NY: McGraw-Hill, 1968.

[9] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

[10] R. S. Woodworth and H. Schlosberg, *Experimental psychology*. Oxford and IBH Publishing, 1954.

[11] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[12] B. Xie, *Head-related transfer function and virtual auditory display*. Plantation, FL: J Ross, 2013.

[13] F. M. Fazi, M. Noisternig, and O. Warusfel, "Representation of sound fields for audio recording and reproduction," *Acoustics 2012 Nantes*, 2012.

[14] E. Williams, *Fourier Acoustics: sound radiation and nearfield acoustical holography*. Cambridge, MA: Academic Press, 1999.

[15] J. Blauert, *Spatial hearing*. Cambridge, MA: MIT Press, 1997.

[16] M. A. Gerzon, "General metatheory of auditory localisation," in *92nd Audio Engineering Society Convention, Vienna*, no. 3306, 1992.

[17] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc*, vol. 25, no. 4, pp. 196–200, 1977. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=3376

[18] J. O. Smith, *Introduction to digital filters: with audio applications*. Julius Smith, 2008, vol. 2.

[19] 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

[20] J. M. Arend, A. Neidhardt, and C. Pörschmann, "Measurement and perceptual evaluation of a spherical near-field hrtf set," in *Proc. 29th Tonmeistertagung - VDT International Convention, 2016*, 2016.

[21] B. Bernschütz, "A spherical far field hrir/hrtf compilation of the neumann ku 100," in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*, 2013, p. 29.

[22] A. W. Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[23] D. R. Perrott and K. Marlborough, "Minimum audible movement angle: marking the end points of the path traveled by a moving sound source," *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1773–1775, 1989.

[24] W. O. Brimijoin and M. A. Akeroyd, "The moving minimum audible angle is smaller during self motion than during source motion," *Frontiers in neuroscience*, vol. 8, no. 273, 2014.

[25] E. R. Hafter, R. H. Dye, J. M. Neutzel, and H. Aronow, "Difference thresholds for interaural intensity," *The Journal of the Acoustical Society of America*, vol. 61, no. 3, pp. 829–834, 1977.

[26] R. Hershkowitz and N. Durlach, "Interaural time and amplitude jnds for a 500-hz tone," *The Journal of the Acoustical Society of America*, vol. 46, no. 6B, pp. 1464–1467, 1969.

**Dylan Menzies** Dr Dylan Menzies is a Senior Research Fellow in the Institute of Sound and Vibration, at the University of Southampton. Areas of interest include spatial audio synthesis and reproduction, sound synthesis for virtual environments, and musical synthesis and interfaces. He holds a PhD in Electronics from the University of York, an MA in Mathematics from Cambridge University, and has worked as a research engineer for several companies including Sony Professional Audio.

**Filippo Maria Fazi** Filippo Maria Fazi graduated in Mechanical Engineering from the University of Brescia (Italy) in 2005. He obtained his PhD in acoustics from the Institute of Sound and Vibration Research (ISVR) of the University of Southampton, UK, in 2010, with a thesis on sound field reproduction. In the same year, he was awarded a research fellowship by the Royal Academy of Engineering and by the Engineering and Physical Sciences Research Council. He is currently an Associate Professor at the University of Southampton. Dr Fazi s research interests acoustical inverse problems, multi-channel systems, virtual acoustics, microphone and loudspeaker arrays.