SURROUND SOUND WITHOUT REAR LOUDSPEAKERS: MULTICHANNEL COMPENSATED AMPLITUDE PANNING AND AMBISONICS

Dylan Menzies

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

Filippo Maria Fazi

Institute of Sound and Vibration Research,
University of Southampton

ABSTRACT

Conventional panning approaches for surround sound require loudspeakers to be distributed over the regions where images are needed. However in many listening situations it is not practical or desirable to place loudspeakers some positions, such as behind or above the listener. Compensated Amplitude Panning (CAP) is a method that adapts dynamically to the listener's head orientation to provide images in any direction, in the frequency range up to $\approx$ 1000 Hz using only 2 loudspeakers. CAP is extended here for more loudspeakers, which removes some limitations and provides additional benefits. The new CAP method is also compared with an Ambisonics approach that is adapted for surround sound without rear loudspeakers.

1. INTRODUCTION

Amplitude panning is a method for producing a spatial audio image in which 2 or more waves combine coherently at the listener position, each carrying the same signal but independent gains. For some choices of plane wave directions and gains the listener perceives an image, or phantom source, from a definite direction, a phenomena known as summing localisation [1]. The direction of the image can be varied continuously by varying the gains.

Below $\approx$ 1000Hz the perception of image direction is mainly determined by the Interaural Time Difference (ITD) cue. In this frequency range, a central stereo image, produced by panning with 2 loudspeakers, is unstable. If the listener faces straight ahead the image is also straight ahead. As the listener turns away from this direction the image moves in the direction of the listener, as illustrated in Fig. 1 [2, 3, 4]. A typical scene contains multiple

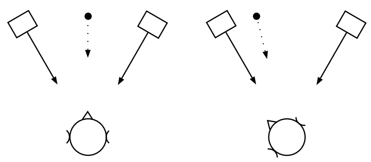

Figure 1: The black dot indicates the direction of the image when 2 loudspeakers each have the same signal, for different head directions.

images in different directions, so at any moment images that are not directly ahead of the listener or inline with a loudspeaker will be distorted. The distortion is greater when the angle between the loudspeakers, viewed from the listener, is increased. For example the listener can approach a stereo pair until the loudspeakers are 180° apart. In this position an image panned to the centre would be completely unstable. Producing consistent ITD cues when the head rotates, otherwise known as *dynamic ITD cues*, is important for localisation [1, 5, 6, 7].

The change in the panned image direction when the head is rotated is caused by the ITD cue not matching that of a static source for each head angle. Compensated Amplitude Panning (CAP), is an extension of conventional panning methods in which the ITD cues are corrected by modifying the gains to take account of the head orientation of the listener [8]. Tracking the listener accurately in real-time with low latency is a challenging requirement for this system. However suitable tracking technology is progressing very rapidly, driven by a wide range of applications.

CAP has been developed for 2 loudspeaker reproduction (Stereo-CAP). This produces more stable images than conventional stereo across the front stage. Further more, the method can produce images in any direction, because ITD is reproduced accurately in any case. Dynamic ITD cues generated by small head movements allow the resolution of front-back ambiguities, and elevation.

To cover the full bandwidth CAP can be combined with high frequency reproduction methods. CAP requires only 2 loudspeakers that are capable of driving the ITD frequency range, while the high frequency range can be driven using smaller and lighter loudspeakers, that are practical to use in higher numbers. Energy based panning, or *Vector Base Intensity Panning (VBIP)* [9] can be combined with Stereo-CAP to provide a very stable full bandwidth front stage. Stereo-CAP provides low frequency coverage elsewhere, which is useful for immersive ambience and reverberation. High frequency coverage can also be provided in all directions using *transaural cross-talk cancellation* [10, 11]. Cross-talk cancellation systems generally perform poorly at low frequencies because the inverse transfer function is then ill-conditioned. CAP can take over in this range, and has the advantage of not requiring calibration for the listener's head diameter.

An extension to Stereo-CAP for near-field images has been made by matching the low frequency ILD (Inter-aural Level Difference) to that of a near source. This is possible using complex panning gains realized with a 1st order filter [12].

For a low frequency spherical head model, [8], the condition that the ITD and ILD cues match with the target plane wave can be formulated as

$$\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \boldsymbol{r}_V) = 0 \tag{1}$$

where $\hat{\boldsymbol{r}}_I$ is the direction of the image, $\hat{\boldsymbol{r}}_R$ is the inter-aural axis, and $\boldsymbol{r}_V$ is the Makita vector that represents the sound field at low frequencies [13]. If the field is produced by panning, the waves at the listener can be approximated as plane waves provided the listener is not so close to the loudspeakers that near-field cues are significant. In this case the Makita vector is given by

$$\boldsymbol{r}_V = \frac{\sum g_i \hat{\boldsymbol{r}}_i}{\sum g_i} \tag{2}$$

where g_i are the gains of the source signal *at the listener*, and $\hat{\boldsymbol{r}}_i$ are the direction vectors of the loudspeakers relative to the listener [8]. The gains at the loudspeakers are compensated for the variable distance to the loudspeakers. Since the wave amplitude falls by $1/r$ the compensated loudspeaker gains are $r_i g_i$. Also delays are introduced to the loudspeaker feeds so that the signals at the listener are in phase. These compensations depend on accurate knowledge of the ambient speed of sound, as well as the distances.

Combining (1) and (2), and normalising the total gain, which determines the overall level, leads to expressions for Stereo-CAP gains,

$$g_1 = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_2)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \quad g_2 = \frac{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_I - \hat{\boldsymbol{r}}_1)}{\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_2 - \hat{\boldsymbol{r}}_1)} \tag{3}$$

These panning laws were tested objectively by calculating the resulting cues at different frequencies for a KEMAR dummy head [8]. The perceived directional error was then calculated and found to be within a Minimum Audible Angle (MMA) [14] for a wide range of target images and head orientations. Subjective tests were carried out to evaluate the stability of images in all directions. Dynamic head tracking was used to allow natural unrestricted listening. The tests showed that images between loudspeakers were improved, and further more steady images could now be created in directions away from the loudspeakers.

It is helpful to visualise the 3-dimensional vectors in the solution. Fig. 2 shows a plan view of these vectors. This is called a *Makita diagram* here since each point on this diagram corresponds to a value of $\boldsymbol{r}_V$, rather than a point in 3-dimensional space. The

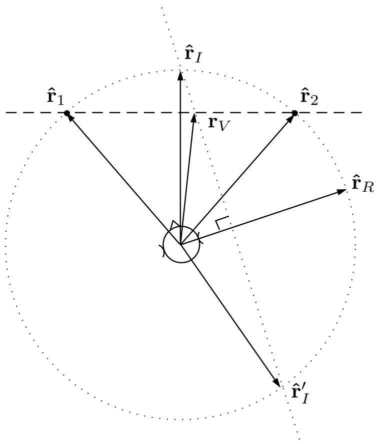

Figure 2: Makita diagram for Stereo CAP, in plan view, for a listener facing towards left of centre of the stereo array. The Makita vector is to the right of centre in order to keep the image central. Shown are loudspeaker directions $\hat{\boldsymbol{r}}_1$, $\hat{\boldsymbol{r}}_2$ the inter-aural direction $\hat{\boldsymbol{r}}_R$, image direction $\hat{\boldsymbol{r}}_I$ and Makita vector $\boldsymbol{r}_V$

dotted circle is a cross section through a sphere of radius 1. A point $\boldsymbol{r}_V$ on the circle or sphere corresponds to a plane wave, such as that from a distant loudspeaker or source. The dotted line represents a plane perpendicular to the page containing all the values of $\boldsymbol{r}_V$ of sound fields that produce an image $\hat{\boldsymbol{r}}_I$. The image is not unique, since there is a circle of consistent images, where the plane intersects with the sphere, the *cone of confusion*. The dashed line shows the values of $\boldsymbol{r}_V$ that can be produced by panning using the 2 loudspeakers. Where the plane and line cross is the single value of $\boldsymbol{r}_V$ that can produce the image using stereo panning. The method is valid whatever the direction of the image, even if it is behind or above.

The panning gains are positive for values of $\boldsymbol{r}_V$ between $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_2$. Outside this region, one of the gains is negative, and there is cancellation of the pressure at the listener. The cancellation implies the sum of gain magnitudes $\sum |g_i|$ is greater than the sum of gains $\sum g_i$. Since the reproduction error due to each gain generally accumulates, then for given $\sum g_i$ the total error increases as the sum of gain magnitudes $\sum |g_i|$, and degree of cancellation. Reproduction error is due to inaccuracies in the head model, the audio hardware, and the tracking of the listener and loudspeakers.

If the listener faces towards the side, the plane and line become close to parallel, and the denominators vanish. The gains become large and polarised and the error increases. The common gain in the denominators can be limited, however this will reduce the perceived image level.

Introducing another loudspeaker between the existing pair would introduce more freedom for controlling $\boldsymbol{r}_V$,and the singular case can be avoided. Solutions for more than 2 loudspeakers are developed in the remainder of this article.

2. SOLUTIONS WITH MORE THAN 2 LOUDSPEAKERS

A Makita diagram with 3 loudspeakers is illustrated Fig. 3. Provided the loudspeakers direction vectors are distinct, then the producible values of $\boldsymbol{r}_V$ cover a plane containing $\hat{\boldsymbol{r}}_1, \hat{\boldsymbol{r}}_2, \hat{\boldsymbol{r}}_3$. The corresponding gains are positive for $\boldsymbol{r}_V$ in the triangular region inside these points, the *convex hull* of the points, and at least one gain is negative for each point outside. Two image examples are shown, each with a head superimposed to show the head orientation in order to simplify the picture in Fig. 2. The image direction and head orientation define the plane of permitted $\boldsymbol{r}_V$ values indicated by the dotted line. If the dotted and dashed planes intersect then

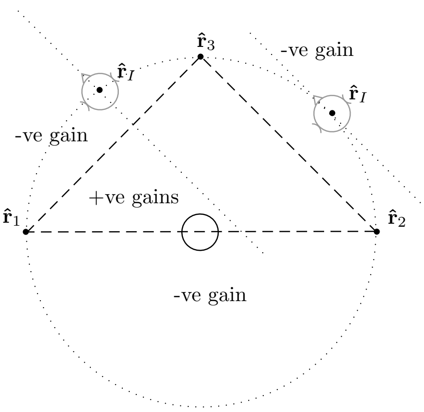

Figure 3: Makita diagram for CAP with 3 loudspeakers, in plan view. Shown are loudspeaker directions $\hat{\boldsymbol{r}}_1$, $\hat{\boldsymbol{r}}_2$, $\hat{\boldsymbol{r}}_3$, and two images $\hat{\boldsymbol{r}}_I$, each with associated head orientations.

there are possible solutions along the line of intersection. There are no solutions only when the planes are parallel and separated, which only happens when the inter-aural axis is perpendicular to the loudspeaker plane, ie when one ear is pointing directly up. Different strategies can be considered for selecting from the possible solutions:

Localised energy : It is natural to try and localise loudspeaker energy in the directions where images are. In high frequency panning this reduces image spread, and makes images more compatible for multiple listeners is different locations. In the low frequency ITD range image spread is perceived much less, provided the cues are consistent, because the cues only contain directional information. The image on the left side in Fig. 3 has a localised solution where the dotted line crosses the dashed line between $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_3$. The gain is zero for the other loudspeaker $g_2 = 0$. This is similar to a pairwise panning arrangement. However for the image on the right side there are no positive solutions. Solutions are possible with negative gain and either $g_2 = 0$ or $g_3 = 0$, but they are not localised to the target image. To move continuously between these solutions when the head rotates requires non-zero gain from all loudspeakers.

Least radiated energy: The energy radiated, $\sum r_i^2 g_i^2$, drives room reverberance that interferes with the direct signal at the listener. Reducing this energy reduces interference, and also the maximum power required from the loudspeakers. Although the precedence effect mitigates the localisation error caused by reverberance, it is desirable to minimise the reverberance because of its overall effect. A minimum energy solution will generally be spread over all the available loudspeakers. However, as explained above, spreading is not a primary concern in the low frequency ITD range.

Least direct energy: CAP may produce gains with opposite sign, and cancellation of pressures at the listener. As with the case of Stereo-CAP, cancellation implies the sum of gain magnitudes $\sum |g_i|$ is greater than $\sum g_i$, and the total reproduction error is increased. The energy sum $\sum g_i^2$ provides a measure of total error that captures the incoherent addition of errors, and is convenient to optimise. Minimising this quantity will minimise the reproduction error due to the direct signal. The solutions for least radiated energy and least cancellation error could be combined to give partial weight to each strategy. These solutions are the same when the distances r_i are equal. Note that $r_V >> 1$ implies cancellation and $\sum g_i^2 >> 1$, however $\sum g_i^2 >> 1$ is also possible for $r_V = 1$, for example in the case of Ambisonics.

Ambisonic: If the image direction $\hat{\boldsymbol{r}}_I$ is restricted to the plane containing the loudspeaker directions, then there is a solution $\boldsymbol{r}_V = \hat{\boldsymbol{r}}_I$ that is independent of head orientation. This is equivalent to Ambisonic panning based on mode matching of the sound field to first order [3, 15]. The low frequency cues depend only on the first order approximation. It is unusual to consider mode matching for full surround without loudspeakers behind the listener. Mathematically this is possible, but it is not immediately clear how well conditioned it is, and how much direct energy is needed.

2.1. Least energy solution

From the above discussion, the most useful solutions for general images are for the least radiated energy and the least direct energy. These solutions can be found analytically. This is shown first for the least radiated energy case. The least direct energy solution is then a special case of this.

Substituting (2) in (1) and multiplying by $\sum g_i$ gives the constraint

$$\sum g_i \,(\hat{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{r}}_i) = \hat{\boldsymbol{r}}_R \cdot \hat{\boldsymbol{r}}_I \tag{4}$$

The summation range for the index i is omitted here and in the following. A second condition is needed to fix the level of the perceived image to a non-zero value, without which the gains would be minimized to zero. This is achieved by specifying the the incident pressure at the listener, which ensures the binaural signals will match those of a planewave with the same incident pressure. For a normalised level,

$$\sum g_i = 1 \tag{5}$$

The 2 constraints (4) and (5) can be combined to produce an alternative for constraint (4),

$$\sum g_i \,\alpha_i = 0\;,\;\; \alpha_i = \;\hat{\boldsymbol{r}}_R \cdot (\hat{\boldsymbol{r}}_i - \hat{\boldsymbol{r}}_I) \tag{6}$$

where α_i is defined here for convenience. Using constraints (5) and (6) simplifies the gain formulae that will be derived. The least energy problem can be stated by minimising the total energy radiated by the loudspeakers,

$$\text{argmin}_{\{g_i\}} \sum (r_i g_i)^2 \tag{7}$$

subject to the previous constraints (6) and (5). This function and the conditions are smooth, so a closed solution is sought using Lagrange multipliers. The Lagrangian is

$$\mathcal{L} = \sum (r_i g_i)^2 - \lambda_1 \sum g_i \alpha_i - \lambda_2 (\sum g_i - 1) \tag{8}$$

with multipliers λ_1, λ_2. Setting partial derivatives by the unknown parameters to zero, $\partial\mathcal{L}/\partial g_i = 0, \partial\mathcal{L}/\partial\lambda_1 = 0, \partial\mathcal{L}/\partial\lambda_2 = 0$, produces $n + 2$ constraints, including the original 2 constraints, where n is the number of loudspeakers.

$$2r_i^2 g_i - \lambda_1 \alpha_i - \lambda_2 \;=\; 0\;,\; i = 1 \mathinner{..} n \tag{9}$$

$$\sum g_i \alpha_i \;=\; 0 \tag{10}$$

$$\sum g_i \;=\; 1 \tag{11}$$

From (9) the gains can be written

$$g_i = \frac{\lambda_1 \alpha_i + \lambda_2}{2r_i^2} \tag{12}$$

Substituting the gains into (10),

$$\sum \frac{\lambda_1 \alpha_i + \lambda_2}{2r_i^2} \alpha_i = 0$$

$$\lambda_1 \sum \frac{\alpha_i^2}{r_i^2} + \lambda_2 \sum \frac{\alpha_i}{r_i^2} = 0$$

$$\lambda_1 \gamma + \lambda_2 \beta = 0 \tag{13}$$

where $\beta = \sum \frac{\alpha_i}{r_i^2}$ and $\gamma = \sum \frac{\alpha_i^2}{r_i^2}$ are defined for convenience. Substituting the gains into (11),

$$\sum \frac{\lambda_1 \alpha_i + \lambda_2}{r_i^2} = 2$$

$$\lambda_1 \beta + \eta \lambda_2 = 2 \tag{14}$$

where $\eta = \sum \frac{1}{r_i^2}$. (13) and (14) can be solved simultaneously to find λ_1 and λ_2,

$$\lambda_1 = \frac{-2\beta}{\gamma\eta - \beta^2} \tag{15}$$

$$\lambda_2 = \frac{2\gamma}{\gamma\eta - \beta^2} \tag{16}$$

The resulting optimal gains are found by substituting into (12),

$$g_i = \frac{\gamma - \beta\alpha_i}{r_i^2(\gamma\eta - \beta^2)} \tag{17}$$

These gains are inexpensive to evaluate, which allows them to be updated frequently when the listener moves. The compensated loudspeaker gains are $r_i g_i$. A global gain factor can be added to set the reproduction level. The least direct energy solution can be found by setting all the loudspeaker distances $r_i = 1$. The least energy solution using 2 loudspeakers has to be identical to Stereo-CAP, because there can be only one solution. This can also be checked algebraically by simplifying (17) for the case $n = 2$. Like Stereo-CAP, it is possible to extend the least energy solution for near-field images, although this is not shown here.

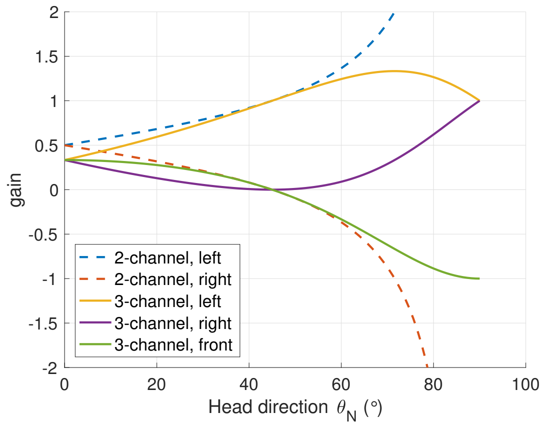

Figure 4: Gains for Stereo-CAP and 3-way CAP for an image at 180° azimuth, and a range of head directions.

The plots shown in Fig. 4 compare the gains produced by the Stereo-CAP system with the least energy 3-way CAP system. Head direction is varied, and the image is directly behind. The Stereo loudspeakers are directly to the left and right. The 3-way system has loudspeakers in these positions and an extra one directly in front, the same as Fig. 3. When the listener turns to the side the Stereo-CAP gains become large, whereas the 3-way CAP gains have magnitudes similar to the total gain $\sum g_i = 1$.

Adding a 4th loudspeaker that is not coplanar with the others, for example above the front loudspeaker in the example shown in Fig. 3, increases the space of $\boldsymbol{r}_V$ that can be produced by panning, from a plane to the whole 3-dimensional Makita space. The panning gains are all positive for points inside the convex hull described by the 4 loudspeaker direction vectors, and at least one gain is negative for each point outside this region. The intersection of the whole space with the plane described by the ITD constraint is always non empty, so there are no singular configurations.

The multichannel solution can be used with any number of loudspeakers. While an advantage of the CAP system is that it requires only a few loudspeakers, more loudspeakers can be added to progressively reduce the radiated energy. Effectively this is beam forming focused on the listener.

The subjective performance of least energy CAP with more than 2 loudspeakers can be inferred from the objective and subjective results for the 2-channel case [8]. These results show that an upper bound for the subjective localisation error can be given that depends only on the total gain energy $\sum g_i^2$. From this the given 3-channel case the total energy is sufficiently low, across all common configurations of image and listener, so that the inferred error is within an MMA. This also implies reverberant interference is at least as low as the 2-channel test cases, for which reverberance could be heard but did not affect image localisation.

2.2. Ambisonic solution

In the mode-matched Ambisonic approach, the aim is to produce an image by reproducing the associated sound field. To produce an accurate low frequency ITD cue it is enough to reproduce pressure and velocity, forming the 1st order of approximation. The first order problem can be written in terms of the variables used in this article by combining (2) and (5) into a single matrix equation,

$$\begin{bmatrix} 1 & 1 & 1 \,.. \\ \hat{\boldsymbol{r}}_1 & \hat{\boldsymbol{r}}_2 & \hat{\boldsymbol{r}}_3 \,.. \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ . \\ . \end{bmatrix} = \begin{bmatrix} 1 \\ \boldsymbol{r}_V \end{bmatrix} \tag{18}$$

Or, abreviated,

$$\mathbf{R}\mathbf{g} = \mathbf{s} \tag{19}$$

The least energy solution, where it exists, is given using the pseudoinverse

$$\mathbf{g} = \mathbf{R}^{+}\mathbf{s} \tag{20}$$

$\mathbf{R}^{+}$ is the Ambisonic decoding matrix. For the example shown in Fi. 4, with 3 loudspeakers and an image behind,

$$\mathbf{s} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{21}$$

$$\mathbf{R}^{+} = \begin{bmatrix} 1/2 & -1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \tag{22}$$

The ordering of loudspeakers here is left, right then centre. Although there are no rear loudspeakers, the target image can be produced without excessive gains or cancellation in this case. However if the listener position is set further back, so that the loudspeakers are separated by smaller angles relative to the listener, then the gains for rear images increase rapidly in size and there is more cancellation. For example if the left and right loudspeakers are positioned closer at $-30°, +30°$ then the gains producing a rear image are

$$\mathbf{g} = \begin{bmatrix} 7.46 \\ 7.46 \\ -13.93 \end{bmatrix} \tag{23}$$

Using CAP the gains in this case are small when the listener is facing forward,

$$\mathbf{g} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \tag{24}$$

Assuming equal loudspeaker distances, the total energy radiated by the Ambisonic array is 916 times greater than that for CAP. The CAP gain magnitudes generally increase smoothly as the listener turns their head to the side, and are equal to the Ambisonic gains when the listener faces directly to the sides.

Adding a 4th loudspeaker that is not coplanar with the others, for example above the centre loudspeaker in the example shown in Fig. 3, allows gains to be produced for any image direction, using the Ambisonic method. Comparatively high gains are required when the loudspeakers are positioned more closely, as for the 3 loudspeaker case.

3. CONCLUSION

Using 2 loudspeakers and with full 6-degrees-of-freedom head tracking, position and orientation, it was previously shown possible to create low frequency images in any direction, although excessive gain is required for some listener orientations . Here it was shown that with 3 loudspeakers all images directions can be reproduced with moderate gain except for a small range of orientations that are practically unimportant. Alternatively, taking an Ambisonic approach with position tracking, 3 frontal loudspeakers can reproduce horizontal images, and 4 loudspeakers can reproduce images in any 3D direction. Ambisonics does not require orientation tracking. As loudspeaker separation is reduced Ambisonics suffers from rapidly increasing gains and cancellation for forward head directions and rear images, whereas in this case CAP gains are lower and remain low as separation is reduced. The overall CAP energy can be reduced further by increasing the number of loudspeakers. In the light of the objective and subjective results for the 2-channel case, the multichannel CAP gains mean the localisation accuracy is within an MMA for all common configurations for the test array considered.

The most recent real-time implementation of the multichannel CAP system is based on an extensive and flexible C++ / Python framework for spatial sound rendering, called the *Versatile Interactive Software Rendering framework (VISR)*. It is planned to make this publicly available in due course.

4. ACKNOWLEDGMENT

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.

5. REFERENCES

[1] Jens Blauert, *Spatial hearing*, Cambridge, MA: MIT Press, 1997.

[2] Benjamin Bernfeld, “Attempts for better understanding of the directional stereophonic listening mechanism,” in *Audio Engineering Society Convention 44*, March 1973, number C-4.

[3] Michael Anthony Gerzon, “General metatheory of auditory localisation,” in *92nd Audio Engineering Society Convention, Vienna*, 1992, number 3306.

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

[5] Hans Wallach, “On sound localization,” *J. Acoust. Soc. Am*, vol. 10, pp. 270–274, 1939.

[6] Hans Wallach, “The role of head movements and vestibular and visual cues in sound localization.,” *Journal of Experimental Psychology*, vol. 27, no. 4, pp. 339, 1940.

[7] Bosun Xie and Dan Rao, “Analysis and experiment on summing localization of two loudspeakers in the median plane,” in *Audio Engineering Society Convention 139*. Audio Engineering Society, 2015.

[8] Dylan Menzies, Marcos F. Simon Galvez, and Filippo Maria Fazi, “A low frequency panning method with compensation for head rotation,” *IEEE Trans. Audio, Speech, Language Processing*, vol. 26, no. 2, February 2018.

[9] Jean-Marie Pernaux, Patrick Boussard, and Jean-Marc Jot, “Virtual sound source positioning and mixing in 5.1 implementation on the real-time system genesis,” in *Proc. Conf. Digital Audio Effects (DAFx-98)*. Citeseer, 1998, pp. 76–80.

[10] Bishnu S. Atal and Manfred R Schroeder, “Apparent sound source translator,” Feb. 22 1966, US Patent 3,236,949.

[11] Ole Kirkeby, Philip A. Nelson, and Hareo Hamada, “Virtual source imaging using the stereo dipole,” in *Audio Engineering Society Convention 103*, Sep 1997.

[12] Dylan Menzies and Filippo Maria Fazi, “Spatial reproduction of near sources at low frequency using adaptive panning,” in *Proc. TecniAcustica, Valencia*, October 2015.

[13] Y Makita, “On the directional localization of sound in the stereophonic sound field,” *E.B.U Review*, vol. A, no. 73, pp. 102–108, 1962.

[14] Allen William Mills, “On the minimum audible angle,” *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[15] J. Daniel, “Spatial sound encoding including near field effect,” in *Proc. AES 23nd International Conference, Helsinger, Denmark*, 2003.