Imaging Transforms from the Ambisonic Toolkit: a new approach to sound-image composition

Joseph Anderson[1]

[1]School of Arts & New Media – University of Hull
Scarborough Campus – Scarborough, United Kingdom

j.anderson@hull.ac.uk

Abstract. *While Ambisonic imaging transforms have been known to the audio engineering community, these have largely remained unknown and unused in the world of electroacoustic and computer music composition. The transforms presented here give the artist a very powerful toolset of imaging manipulations, encouraging a sound-image composition paradigm. [re-write]*

1. Introduction

Soundfield transforms. Malham tutorial.

Menzies [Menzies, 1999] has advocated using dominance as a powerful spatial manipulation tool for the creation of electroacoustic music.

In combination with Gerzon's spreading filter [Gerzon, 1975a] applied to a point source (mono) input source, Menzies [Menzies, 1999] has advocated using the dominance transform to control the resulting spread and spatial immersion of an input sound. This procedure has been implemented Dylan's software package LAmb, which is described as "A System for Live Diffusion Using Ambisonics".

My uses. . .

[FRENCH ACCENTS!!!!]

2. Soundfield Transforms

Describe soundfield transforms here, as does Chapman.

$$B' = \mathbf{T}B \tag{1}$$

3. Rotations

All images and illustrations should be in black-and-white, or gray tones. The image resolution on paper should be about 600 dpi for black-and-white images, and 150–200 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

3.1. Rotate

$$\mathbf{R}_{Z,\theta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

3.2. Tilt

$$\mathbf{R}_{X,\phi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix} \tag{3}$$

3.3. Tumble

$$\mathbf{R}_{Y,\phi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & 0 & -\sin\phi \\ 0 & 0 & 1 & 0 \\ 0 & \sin\phi & 0 & \cos\phi \end{pmatrix} \tag{4}$$

4. Dominance and variants

In his report to the UK's National Research Development Corporation (NRDC) [Gerzon, 1975b] Gerzon introduced *dominance* as a useful and creative transform to apply to ambisonic soundfields. The initial applications proposed appear to focus on the needs of the audio engineer, and are discussed as if the engineer is working to post-process B-format concert recordings made with a soundfield microphone [Farrar, 1979]. In his NRDC report Gerzon describes dominance as a *width* control:

> [Which] permits the relative width of the front and back images to be varied (e.g. to modify the width of an orchestra or to emphasise the rear reverberation).... [P]lus an up/down width control that, for example, allows a below-horizontal orchestra and audience (caused by high-up microphones) to be made horizontal by narrowing the upward width.

Here Gerzon is describing the action of dominance[1], firstly on the X axis and then on the Z axis. The dominance transform on the X axis is represented in his report as

$$\mathbf{Z}_{X,\mu} = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}}\mu & 0 & 0 \\ \sqrt{2}\mu & 1 & 0 & \\ 0 & 0 & \sqrt{1-\mu^2} & 0 \\ 0 & 0 & 0 & \sqrt{1-\mu^2} \end{pmatrix} \tag{5}$$

where μ conveniently controls the amount of dominance applied. Valid values for μ are

$$-1 \leq \mu \leq 1 \tag{6}$$

When $\mu = 1$, maximum dominance in the forward direction is applied, with the resulting soundfield compressed to a point source (mono) image at front-centre. As well as spatial distortions, gains are also varied. The gain of what was previously encoded in the input B-format signal at front-centre is increased by $+6$ dB, and at back-centre gain is reduced to $-\infty$ dB. Elements on the Y–Z plane retain their previous gains. The astute observer will recognise this is equivalent to placing a single cardioid microphone in the soundfield

[1] Menzies [Menzies, 1999] has given a somewhat more detailed and precise description: Dominance has the effect of moving the directions towards or away from a special direction, the *direction of the dominance*. If the *dominance factor* is zero, then it [is] the identity. At maximum dominance the directions are all changed to the direction of dominance. Associated with the change of directions is a gain factor. For initial directions further from the dominance direction, the gain is relatively less than for closer directions. At maximum dominance opposed directions are eliminated completely by zero gain. Dominance has been used as a kind of zooming tool to give gain emphasis locally while maintaining B-format integrity.

aimed at front-centre, increasing the gain by $+6$ dB, and then re-encoding this mono signal into a new B-format signal as a point source image at front-centre ($\theta, \phi = 0°, 0°$).

Dominance can also be represented in terms of gain applied in the direction of dominance. Gerzon and Barton [Gerzon and Barton, 1992] have illustrated this form as follows

$$\mathbf{D}_{X,\lambda} = \begin{pmatrix} \frac{1}{2}(\lambda + \frac{1}{\lambda}) & \frac{1}{\sqrt{8}}(\lambda - \frac{1}{\lambda}) & 0 & 0 \\ \frac{1}{\sqrt{2}}(\lambda - \frac{1}{\lambda}) & \frac{1}{2}(\lambda + \frac{1}{\lambda}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7}$$

where λ controls dominance as gain applied at front-centre. If λ is defined as

$$\lambda = 10^{\frac{g}{20}} \tag{8}$$

g controls dominance as gain in dB applied at front-centre. The gain at back-center will be equal to $-g$ dB. Figure 1 illustrates gain and angular distortions applied to a B-format soundfield with varying values of g. This is the dominance variant named in the Ambisonic Toolkit as *dominance*.

These two forms of dominance have appeared throughout the literature named as both *dominance* and *zoom*.[2] Malham [Malham, 1990] has shown how these two variants may be regarded as equivalent, with further discussion to be found elsewhere in the literature [Daniel, 2001] [Hollerweger, 2006]. From a user point of view, however, varying μ gives a different experience of manipulating the soundfield than does varying g. For example, we've seen that for the dominance variant presented in (5), $\mu = 1$ gives maximum dominance, with the image collapsing to a point source image at front-centre. To achieve the equivalent spatial distortion with (7) requires $g = \infty$ dB. As you'd imagine, this is not particularly ideal. In practice, then, we'd expect that at least two forms of dominance to be desirable. Below we'll present two further variants which have been deemed to be musically useful,[3] along with other derived transforms developed to address features of dominance that may be regarded as less suitable in some circumstances.

4.1. Zoom

Both (5) and (7) have been described as both dominance and zoom, which can lead to some confusion. In practice, it appears that (5) is usually described as zoom. Within the Ambisonic Toolkit, we'll use this form, but instead apply dominance as a *dominance angle of distortion*, α.[4] The rationale for this being to supply a dominance variant that matches previous practice (under the name of zoom), yet provides slightly different, but musically useful, ergonomics.[5]

The ATK's zoom applies (5), but in terms of a distortion angle α, with valid values

$$-90° \leq \alpha \leq 90° \tag{9}$$

giving the user the option of directly considering dominance in terms of angular distortions. Daniel [Daniel, 2001] has conveniently shown how (5) transforms the angles (in the $X - Y$ plane) of an input sounfield in terms of μ:

$$cos\theta' = \frac{\mu + cos\theta}{1 + \mu cos\theta} \tag{10}$$

[2] Below we'll seek to address this naming issue further.

[3] And appear in the Ambisonic Toolkit.

[4] The form presented here varies from that presented in [Anderson, 2009b], which is derived from (7).

[5] Additionally we'll also see that further imaging transforms are supplied in terms of α.

The original encoded angle of incidence in the $X - Y$ plane is θ and the resulting transformed angle is θ'. We can consider what happens to the sound encoded at the hard-left of the image by setting $\theta = 90°$. Equation (10) then simplies to:

$$cos\theta' = \mu \tag{11}$$

Describing θ' in terms of the dominance angle of distortion, α

$$\theta' = 90° - \alpha \tag{12}$$

and with further simplification and substitution (5) becomes

$$\mathbf{Z}_{X,\alpha} = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}}\sin\alpha & 0 & 0 \\ \sqrt{2}\sin\alpha & 1 & 0 & 0 \\ 0 & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & \cos\alpha \end{pmatrix} \tag{13}$$

which is the transform the ATK names as zoom. Figure 2 illustrates gain and angular distortions applied to a B-format soundfield as zoom is applied in terms of α.

4.1.1. Balance

We've seen Gerzon describe the dominance (zoom) transform as being akin to a width transform.[6] Applying zoom in the form of (13) on the Y axis results in a transform having a similar result to the well known stereo *balance* transform. The ATK's balance is shown here

$$\mathbf{Z}_{Y,\alpha} = \begin{pmatrix} 1 & 0 & \frac{1}{\sqrt{2}}\sin\alpha & 0 \\ 0 & \cos\alpha & 0 & 0 \\ \sqrt{2}\sin\alpha & 0 & 1 & 0 \\ 0 & 0 & 0 & \cos\alpha \end{pmatrix} \tag{14}$$

with valid values

$$-90° \leq \alpha \leq 90° \tag{15}$$

Distortion angle α controls the displacement of both front-centre and back-center, where $\alpha = 90°$ pushes the image to hard-left, $\alpha = -90°$ to hard-right, and $\alpha = 0°$ leaves the image untouched. Figure 3 illustrates the results the effect of balance with positive values of α.

4.2. Focus

Daniel [Daniel, 2001] has described a transform he names as *focus*:

> ... implemented by Véronique Larcher for [a] work [of] spatial sound composition [by] Cécile le Prado.[7] The visual analogue approximate [to] the effect of a flashlight scanning the darkness.[8]

To realise this effect he advises using (5), setting $\mu = 0$. The results are as described previously in the introduction to section 4.

The ATK takes a slightly different approach, preferring for imaging transforms to be continuously variable. Instead we take the zoom transform of (13) and normalise gain

[6] The author has described a number of classic two-channel stereo transforms in [Anderson, 2009a].

[7] Name the piece or pieces here. Emailed Véronique with the question.

[8] Translation: http://translate.google.co.uk

in the direction zoom is applied, so that at maximum zoom, $\alpha = 90°$, the gain is 0 dB rather than $+6$ dB as in (13). For zoom, gain in the direction of applied dominance is

$$1 + \sin\alpha \tag{16}$$

If we wish our gain normalisation to behave well for both positive and negative values of α, so that gain in normalised in the direction focus is applied, we'll need to use

$$1 + \sin|\alpha| \tag{17}$$

Normalising (13) by (17) results in ATK's focus

$$\mathbf{F}_{X,\alpha} = \begin{pmatrix} \frac{1}{1+\sin|\alpha|} & \frac{1}{\sqrt{2}}\left(\frac{\sin\alpha}{1+\sin|\alpha|}\right) & 0 & 0 \\ \sqrt{2}\left(\frac{\sin\alpha}{1+\sin|\alpha|}\right) & \frac{1}{1+\sin|\alpha|} & 0 & 0 \\ 0 & 0 & \frac{\cos\alpha}{1+\sin|\alpha|} & 0 \\ 0 & 0 & 0 & \frac{\cos\alpha}{1+\sin|\alpha|} \end{pmatrix} \tag{18}$$

A value of $\alpha = 90°$ gives a focused image at front-centre, where a value of $\alpha = -90°$ generates focus at back-centre. For positive values of α gain at front-centre is stabilised at 0 dB. (Figure 4 illustrates.) The same holds true for gain at back-centre for negative values.

5. Combining transforms

It should not be surprising to note that any Ambisonic transform may follow another. That is, we may process a B-format signal with one imaging transform, resulting in a new B-format signal, and then continue to process the result with any number of of subsequent transforms. Such a procedure can produce a number of useful outcomes. As an example, a cascade of two transforms may be represented as

$$B' = \mathbf{T}_1 B \tag{19}$$

$$B'' = \mathbf{T}_2 B' \tag{20}$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ are two imaging transforms. This is equivalent to

$$B'' = (\mathbf{T}_2\mathbf{T}_1)B \tag{21}$$

Generalising to a cascade of N imaging transforms we get

$$B' = (\mathbf{T}_N \cdots \mathbf{T}_2\mathbf{T}_1)B \tag{22}$$

Then, for convenience, we can define a new imaging transform which combines this cascade of operations into a single matrix

$$\mathbf{T}_{1\cdots N} = \mathbf{T}_N \cdots \mathbf{T}_2\mathbf{T}_1 \tag{23}$$

We'll see this process of combination gives us the ability to develop new imaging matrices as well to 'direct' the axial transforms shown above in section 4.

[Show additions!!]

[Also, show push before press, as it is simpler]

5.1. Asymmetry

Just as the balance transform is found in two-channel stereo imagers, *asymmetry*[9] has appeared in two-channel imaging tools. The asymmetry transform allows the balance and orientation of image elements to be altered, while retaining front-centre at front-centre. We can bring the asymmetry transform to B-format by following the procedure outlined above.

The algorithm consists of balance (14) followed by rotation (2), with the rotation re-centering the displacement of front-centre caused by balance. The asymmetry transform can be represented as

$$B'' = (\mathbf{R}_{Z,\theta}\mathbf{Z}_{Y,-\alpha})B \tag{24}$$

In line with the Ambisonic convention of positive angles representing anti-clockwise rotations we've chosen to invert the sign on α. Then for the rotation, setting

$$\theta = \alpha \tag{25}$$

'un-does' the rotation to front-centre caused by balance. The asymmetry transform is then

$$\mathbf{A}_{YZ,\alpha} = \begin{pmatrix} 1 & 0 & -\frac{1}{\sqrt{2}}\sin\alpha & 0 \\ \sqrt{2}\sin^2\alpha & \cos^2\alpha & -\sin\alpha & 0 \\ -\sqrt{2}\cos\alpha\sin\alpha & \cos\alpha\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & \cos\alpha \end{pmatrix} \tag{26}$$

with the resulting image transformation illustrated in Figure 5.

5.2. Press

The *press* and *push* transforms have been developed as alternatives to the focus transform. As has been described above, at $\alpha = 90^\circ$ the focus transform results in a point source image in the direction the transform is applied, with the resulting directional response as 'cardioid'. Reviewing Figure 4 one sees elements at back-centre drop out of the image with a gain of $-\infty$ dB. If our goal is to re-image a B-format signal, yet retain all elements and their relative gains, then the focus transform may not be ideal. Instead, the goal is to develop a transform that becomes a point source with an omni-directional response when $\alpha = 90^\circ$, thus retaining all elements of the input B-format signal, but with all elements pressed or pushed together.

To develop press, we'll introduce one further dominance variant, a version of (7), but with dominance specified in terms of distortion angle α rather than gain. Cotterell [Cotterell, 2002] has conveniently shown how (7) transforms the angles of an input sounfield in terms of λ:

$$cos\theta' cos\phi' = \frac{\lambda^2 - 1 + (\lambda^2+1)cos\theta cos\phi}{\lambda^2 + 1 + (\lambda^2 - 1)cos\theta cos\phi} \tag{27}$$

$$sin\theta' cos\phi' = \frac{2\lambda sin\theta cos\phi}{\lambda^2 + 1 + (\lambda^2 - 1)cos\theta cos\phi} \tag{28}$$

$$sin\phi' = \frac{2\lambda sin\phi}{\lambda^2 + 1 + (\lambda^2 - 1)cos\theta cos\phi} \tag{29}$$

[9] Described in [Anderson, 2009b]

The original encoded angles of incidence are θ, ϕ with the resulting transformed angles as θ', ϕ'. We can consider what happens to the sound encoded at the hard-left of the image by setting $\theta = 90°$ and $\phi = 0$. Equations (27), (28) and (29) then reduce to:

$$cos\theta' = \frac{\lambda^2 - 1}{\lambda^2 + 1} \tag{30}$$

$$sin\theta' = \frac{2\lambda}{\lambda^2 + 1} \tag{31}$$

Describing θ' in terms of the dominance angle of distortion α

$$\theta' = 90° - \alpha \tag{32}$$

and with further simplification and substitution (7) becomes

$$\mathbf{D}_{X,\alpha} = \begin{pmatrix} \sec\alpha & \frac{1}{\sqrt{2}}\tan\alpha & 0 & 0 \\ \sqrt{2}\tan\alpha & \sec\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{33}$$

This form of dominance becomes useful for two reasons: firstly, as dominance is expressed as an angle of distortion, we can follow by a rotation to 'un-do' angular displacements as we've done to develop asymmetry above; secondly, the gain of elements at $90°$ to the dominance axis (in this case, the X axis) remain at 0 dB through all values of α.

The algorithm we'll use will be very similar as to that seen for asymmetry in (24), and can be seen as applying asymmetry using the dominance of (33) in both the positive and negative directions, taking half the resulting signal from positive asymmetry and half from negative. This can be represented as

$$B'' = \frac{1}{2}[(\mathbf{R}_{Z,\theta}\mathbf{D}_{Y,-\alpha})B + (\mathbf{R}_{Z,-\theta}\mathbf{D}_{Y,\alpha})B] \tag{34}$$

Moving B gives

$$B'' = \frac{1}{2}[\mathbf{R}_{Z,\theta}\mathbf{D}_{Y,-\alpha} + \mathbf{R}_{Z,-\theta}\mathbf{D}_{Y,\alpha}]B \tag{35}$$

With our desired imaging transform being

$$\mathbf{P}_{X,\alpha} = \frac{1}{2}[\mathbf{R}_{Z,\theta}\mathbf{D}_{Y,-\alpha} + \mathbf{R}_{Z,-\theta}\mathbf{D}_{Y,\alpha}] \tag{36}$$

Without going through the details of simplifing (36), the resulting imaging matrix is

$$\mathbf{P}_{X,\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{2}\sin^2\alpha & \cos^2\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & \cos\alpha \end{pmatrix} \tag{37}$$

However, we'll make one 'practical' modification to (37) so that press behaves well, as does focus, for positive and negative values of α. Our modification results in

$$\mathbf{P}_{X,\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{2}\sin|\alpha|\sin\alpha & \cos^2\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & \cos\alpha \end{pmatrix} \tag{38}$$

which allows the image to be pressed to front-centre with positive α and to back-centre with negative α. Figure 6 illustrates the application of press to a B-format signal.

5.3. Push

$$\mathbf{U}_{X,\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{2}\sin|\alpha|\sin\alpha & \cos^2\alpha & 0 & 0 \\ 0 & 0 & \cos^2\alpha & 0 \\ 0 & 0 & 0 & \cos^2\alpha \end{pmatrix} \tag{39}$$

6. Directing the transforms

Rotate, etc.

7. Valid?

8. HOA

9. Conclusion

Say some stuff here. . .

10. References

Bibliographic references must be unambiguous and uniform. We recommend

References

Anderson, J. (2009a). Classic stereo transforms—a review. In Pampin, J., editor, *Spatial Composition*, chapter 3. QU.

Anderson, J. (2009b). Introducing. . . the ambisonic toolkit. Graz. Ambisonics Symposium 2009.

Cotterell, P. (2002). *On the Theory of the Second-Order Soundfield Mircophone*. PhD thesis, University of Reading.

Daniel, J. (2001). *Representation de champs acoustiques, application a la transmission et a la reproduction de scenes sonores complexes dans un contexte multimedia*. PhD thesis, Universite Paris 6.

Farrar, K. (1979). Soundfield microphone: design and development of microophone and control unit. *Wireless World*, pages 48–50 (Oct.) 99–103 (Nov.).

Gerzon, M. A. (1975a). Artificial reverberation and spreader devices. Ambisonic Technology Report 4, NRDC.

Gerzon, M. A. (1975b). Panpot and soundfield controlls. Ambisonic Technology Report 3, NRDC.

Gerzon, M. A. and Barton, G. J. (1992). Ambisonic decoders for hdtv. In *Preprint 3345 of the 92nd Audio Preprint 3345 of the 92nd Audio Engineering Society Convention*.

Hollerweger, F. (2006). *Periphonic Sound Spatialization in Multi-User Virtual Environments*. PhD thesis, Institute of Electronic Music and Acoustics (IEM) and Center for Research in Electronic Art Technology (CREATE).

Malham, D. G. (1990). Ambisonics: A technique for low cost, high precision, three dimensional sound diffusion. ICMC Conference, pages 118–120, Glasgow. International Computer Music Association.

Menzies, D. (1999). *New Electronic Performance Instruments for Electroacoustic Music*. PhD thesis, University of York.

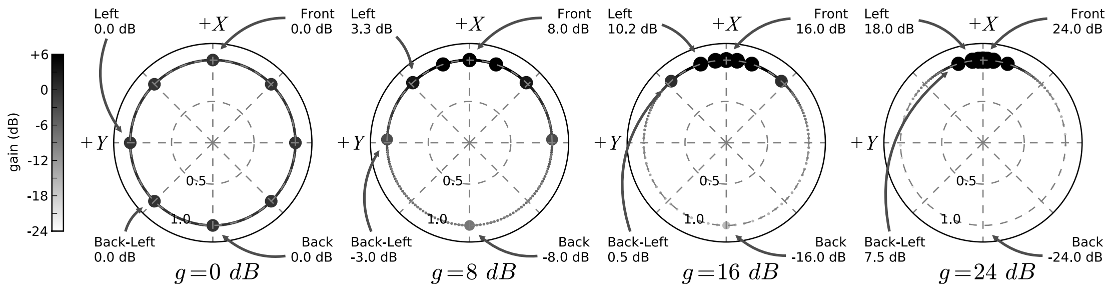

Figure 1: Dominance

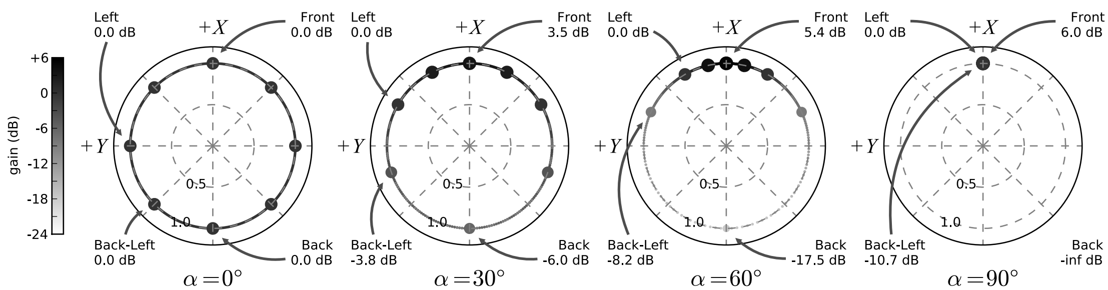

Figure 2: Zoom

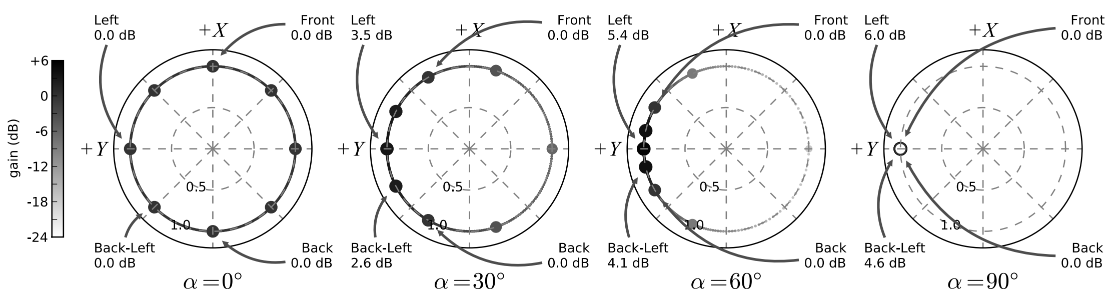

Figure 3: Balance

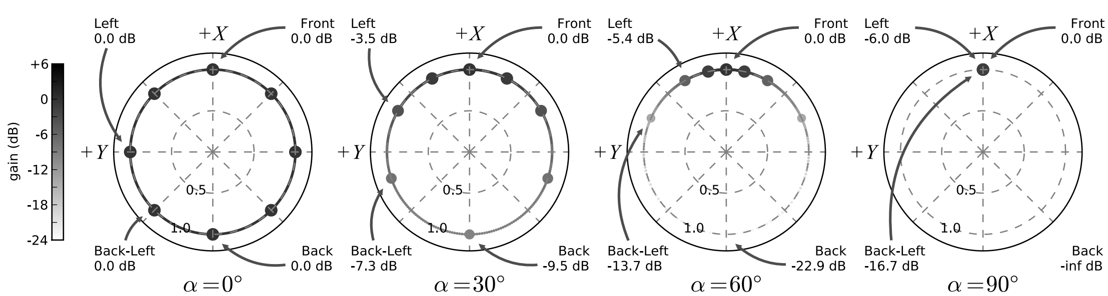

Figure 4: Focus

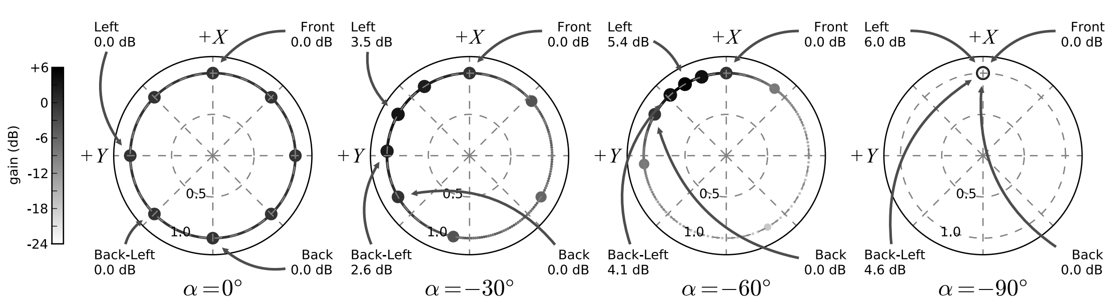

Figure 5: Asymmetry

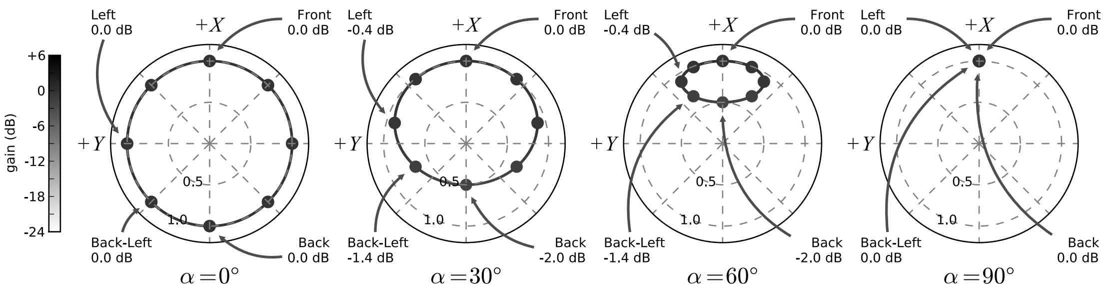

Figure 6: Press

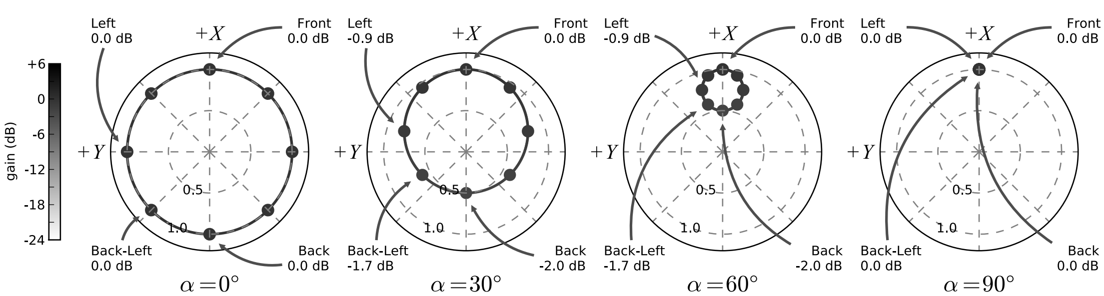

Figure 7: Push