STEREOPHONIC UPMIXING TO B-FORMAT

Haydon Cardew

MSc Audio Engineering at Derby University

Aims and Objectives

Aim

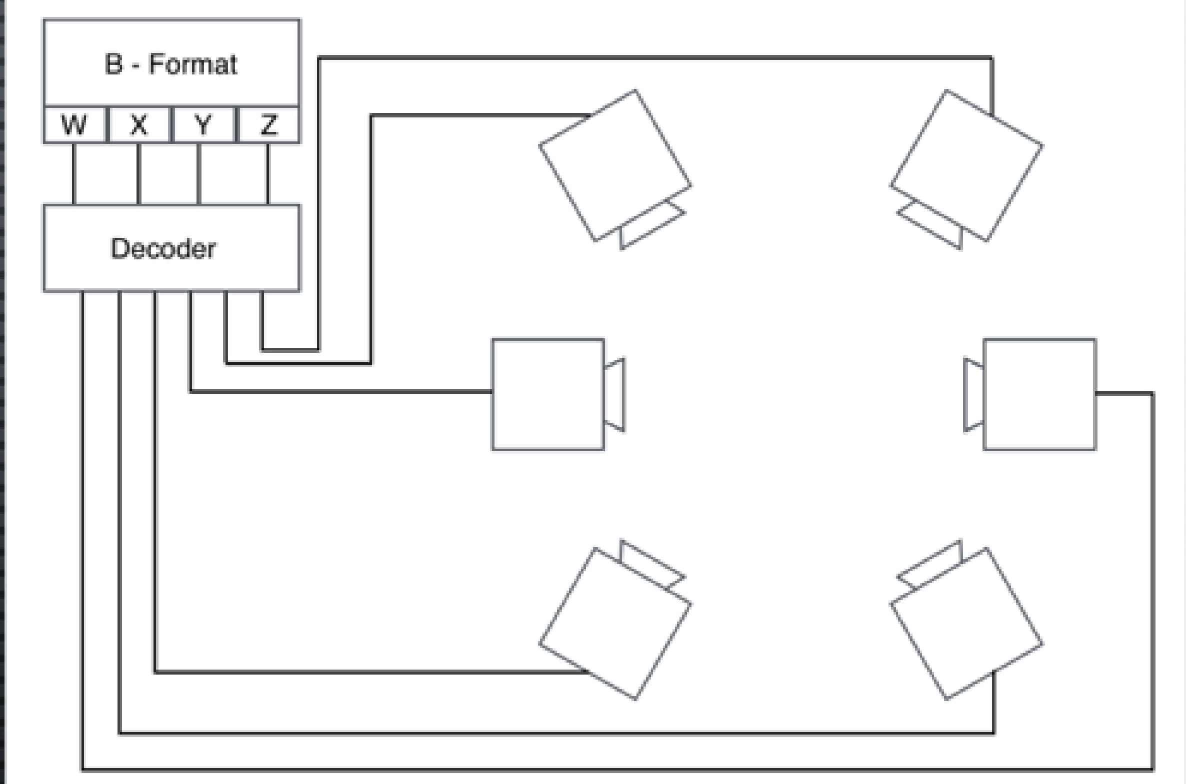
Create a stereo to B-format up-mix algorithm.

Objectives

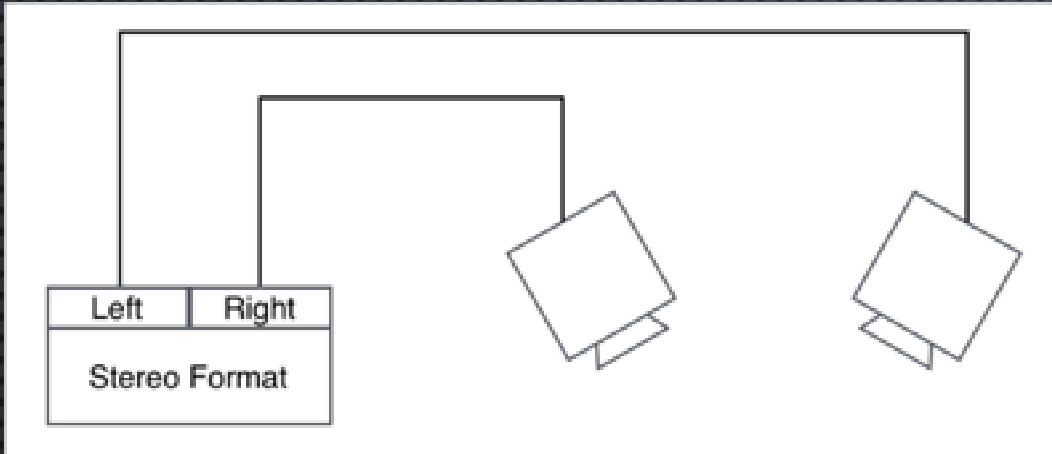
- Use conventional, commercial stereophonic audio.
- Use an appropriate source separation method to divide the audio into constituent parts.
- Compute a horizontal angle (azimuth) of the extracted parts.
- Encode extracted audio into B-format.

A Closer Look At The Formats

Ambisonic



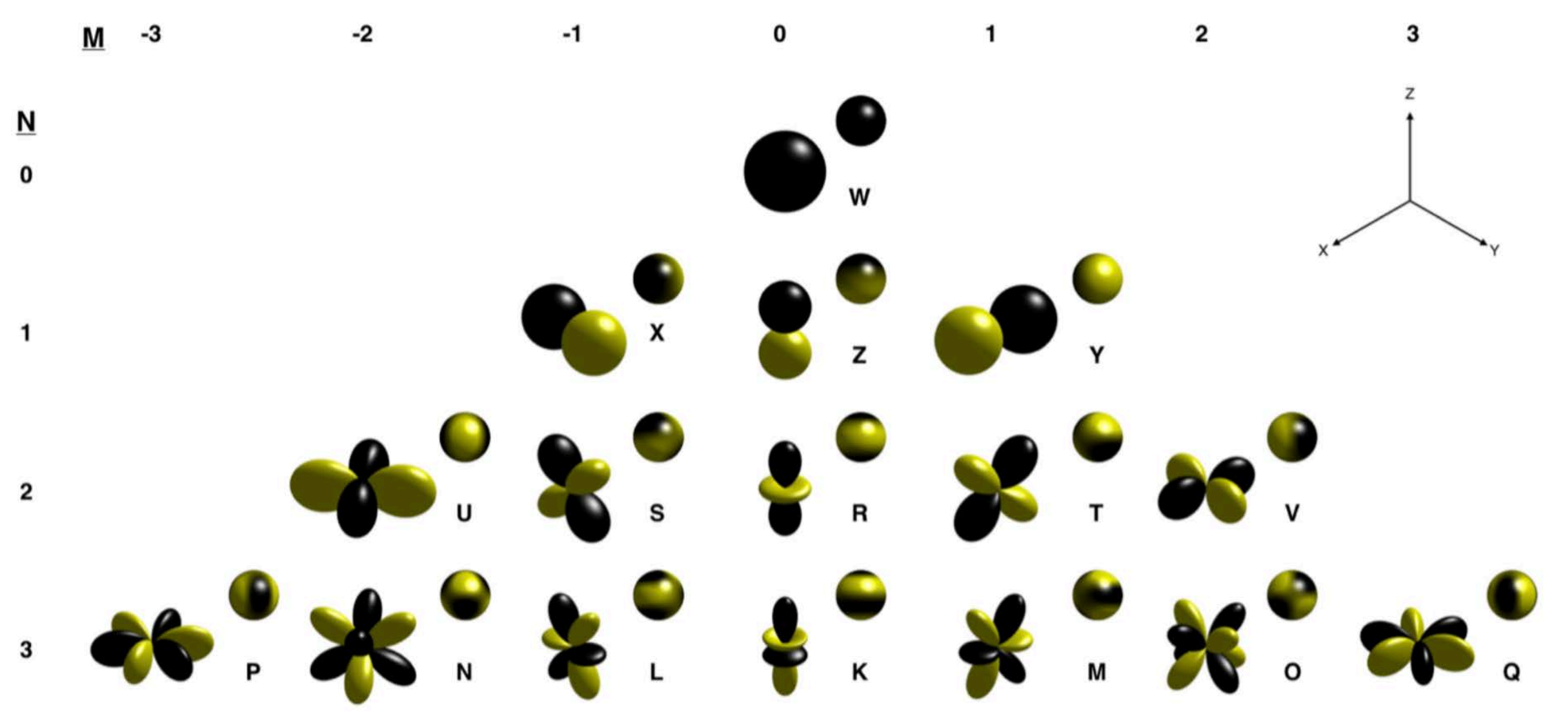
Stereophonic



- Mixed to a predefined setup
- Two channels feed directly to two speakers

- Mixed to a sound field
- Unknown number of speakers used in playback (the above is just an example)
- Number of channels is $(n + 1)^2$, where ‘n’ is the order of ambisonics used

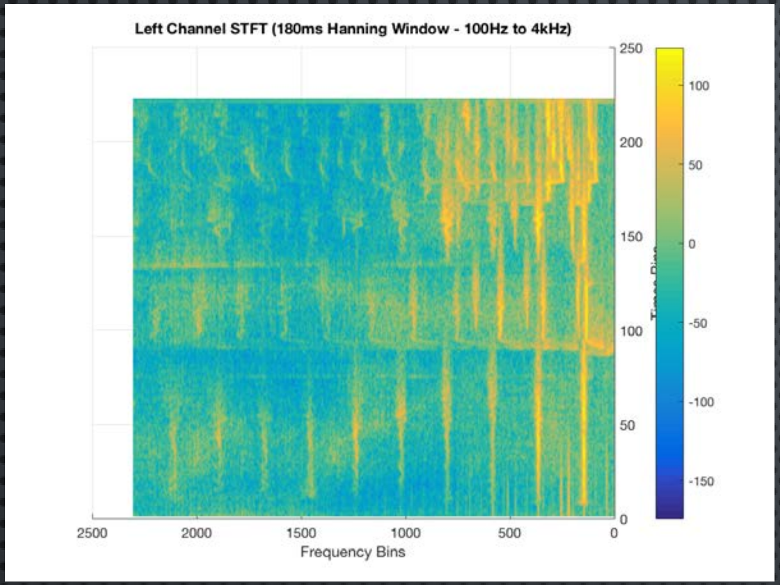
Ambisonic Channels (up to 3rd order)

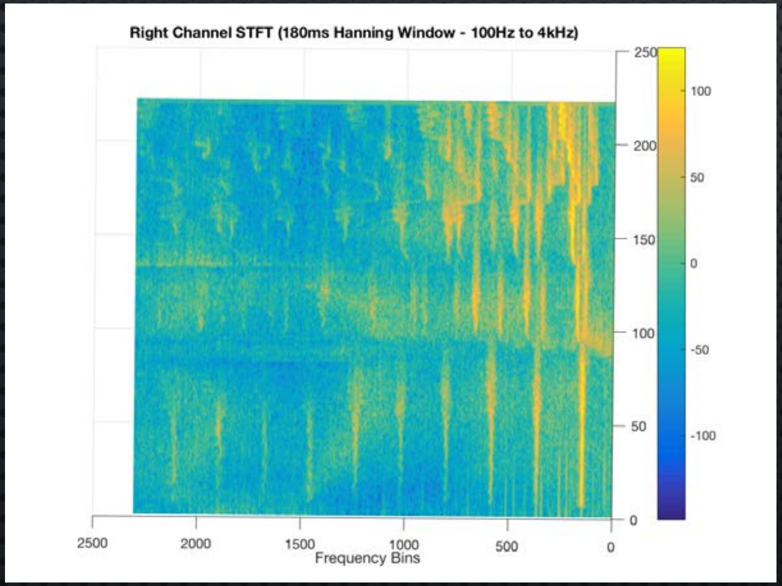


ACN	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Channel Label	W	X	Z	Y	U	S	R	T	V	P	N	L	K	M	O	Q

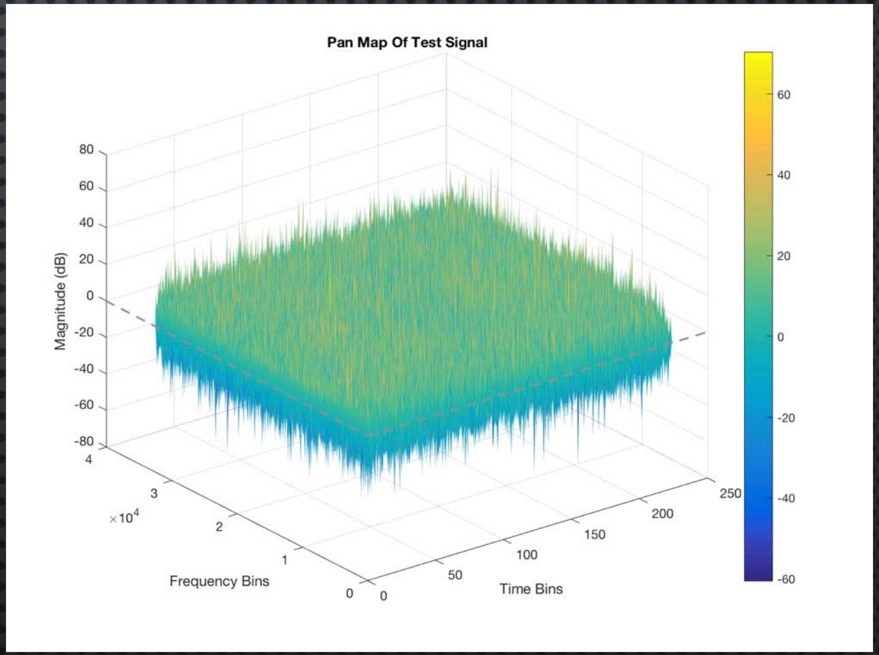
How The Algorithm Works

Algorithm: STFT and Pan Map Creation





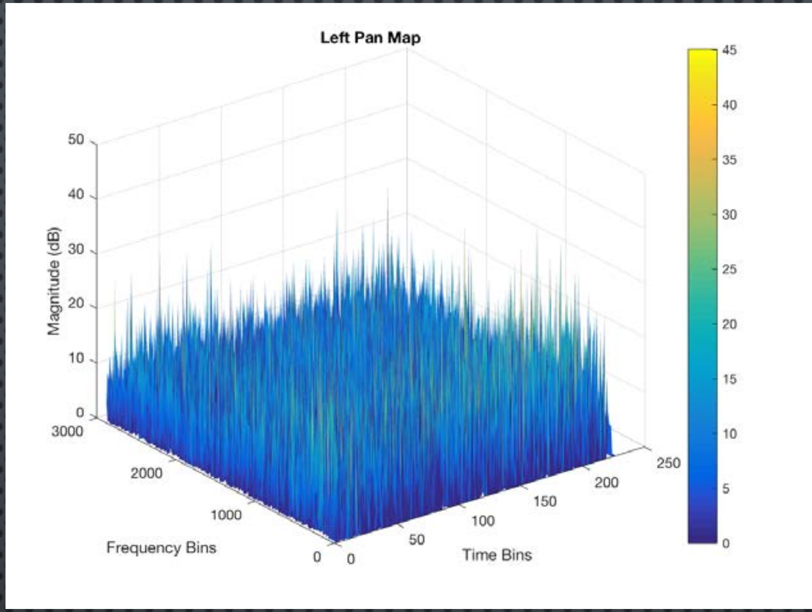
$$PanMap(k,m) = 20\ log\left(\frac{|Left(k,m)|}{|Right(k,m)|}\right)$$

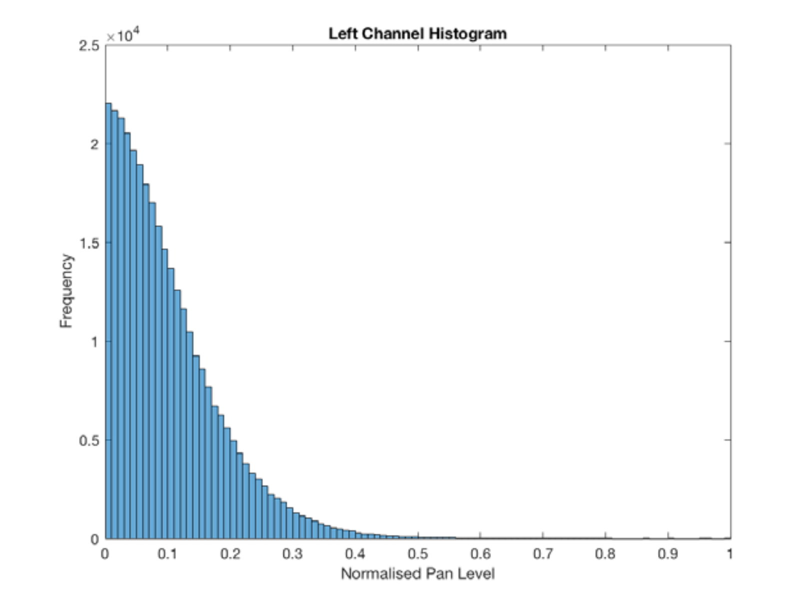


(Cobos and Lopez, 2008)

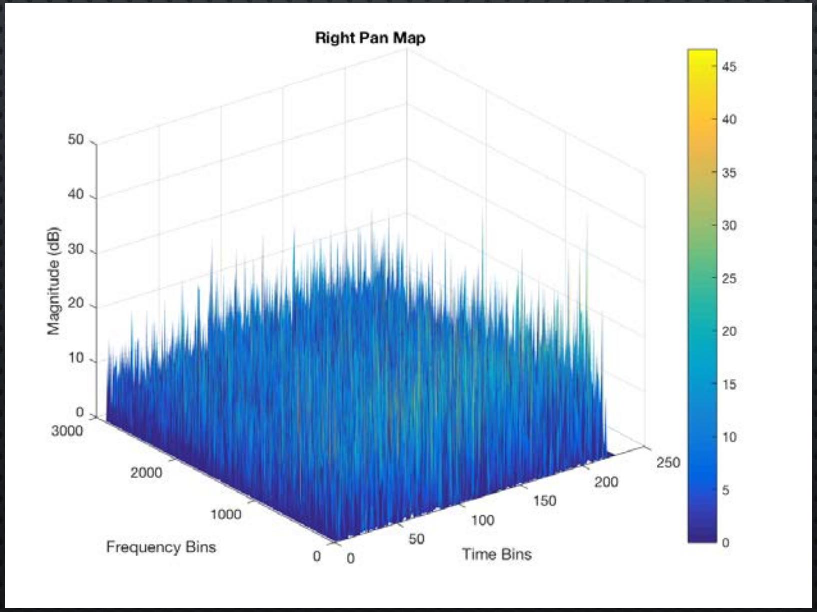
Algorithm: Histogram Creation

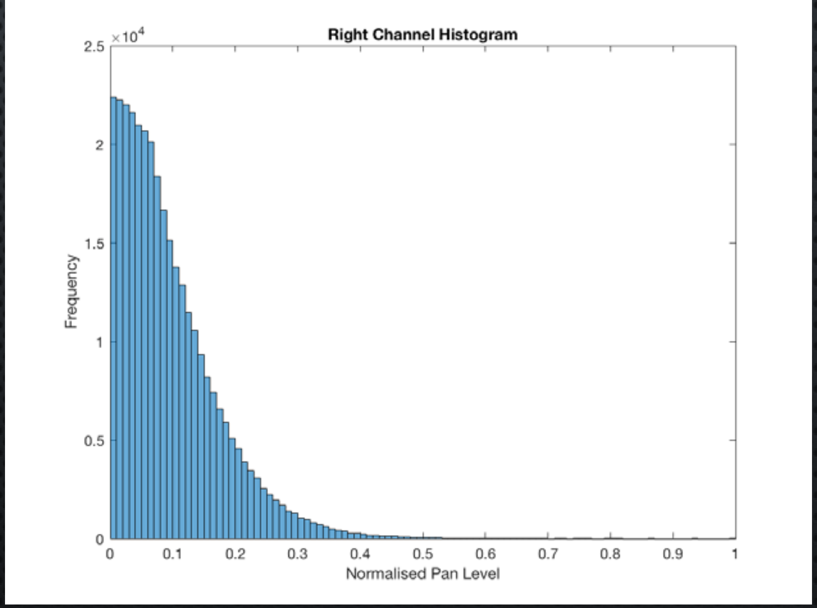
$$LeftPanMap(k,m) = \begin{cases} PanMap(k,m)\ if\ PanMap(k,m) \geq 0 \\ 0\ if\ PanMap(k,m) < 0 \end{cases}$$





$$RightPanMap(k,m) = \begin{cases} PanMap(k,m)\ if\ PanMap(k,m) < 0 \\ 0\ if\ PanMap(k,m) > 0 \end{cases}$$





Otsu’s Method

Otsu’s method finds an optimal threshold to divide a grayscale image into a foreground and a background. Using an extended version of Otsu’s method, the previously found histograms can be split into several sources by finding multiple optimum thresholds.

1 optimum threshold →

Algorithm: Finding Optimum Thresholds

Inter-class variance:

$$\sigma_B^2 = \sum_{k=1}^{M} \omega_k(\mu_k - \mu_T)^2$$

Total mean:

$$\mu_T = \omega_1\mu_1 + \omega_2\mu_2$$

Probability that a chosen data point falls within class *k* :

$$\omega_k = \sum_{n\in C_k} p_n$$

Class mean:

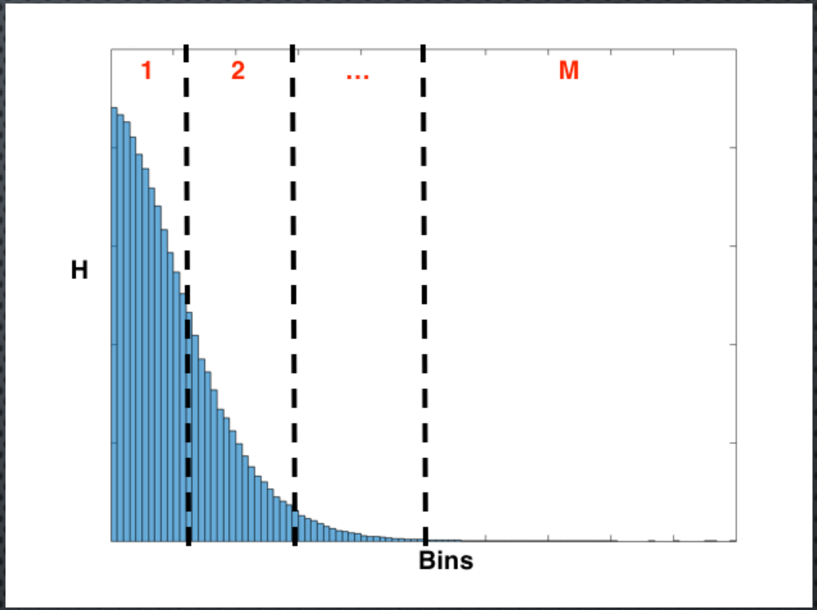
$$\mu_k = \sum_{n\in C_k} n\frac{p_n}{\omega(k)}$$

Probability that a chosen data point falls within bin *n* :

$$p_n = \frac{H(n)}{N}$$

Sum of all H values:

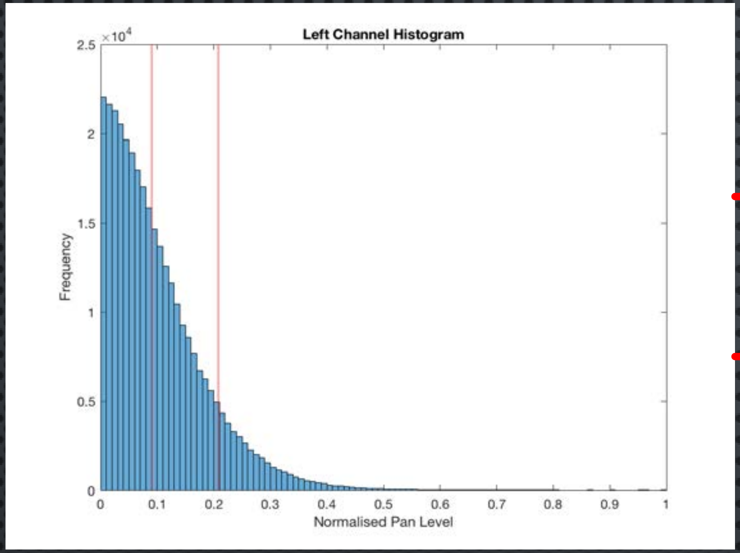
$$N = \sum_{n=1}^{L} H(n)$$

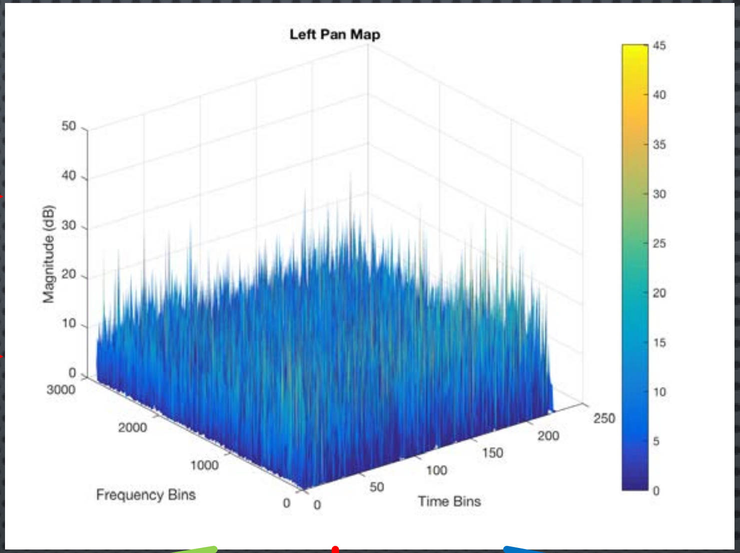


By varying the locations of the thresholds, optimum thresholds are found when σ_B^2 is a maximum

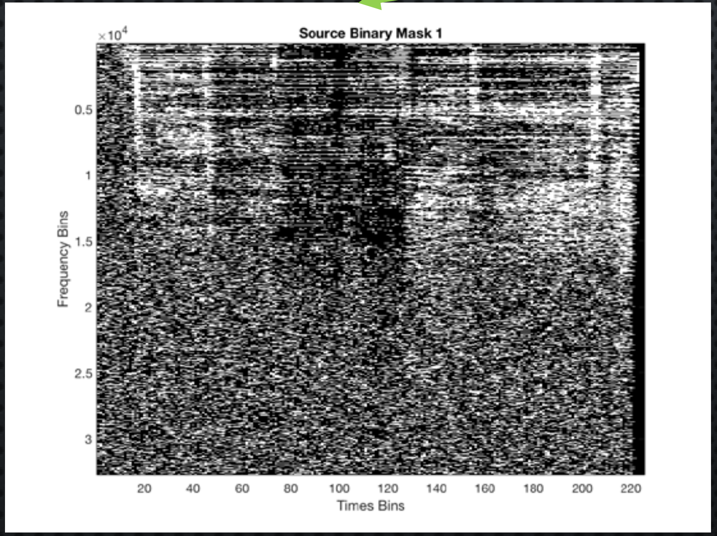
(Otsu,1979)

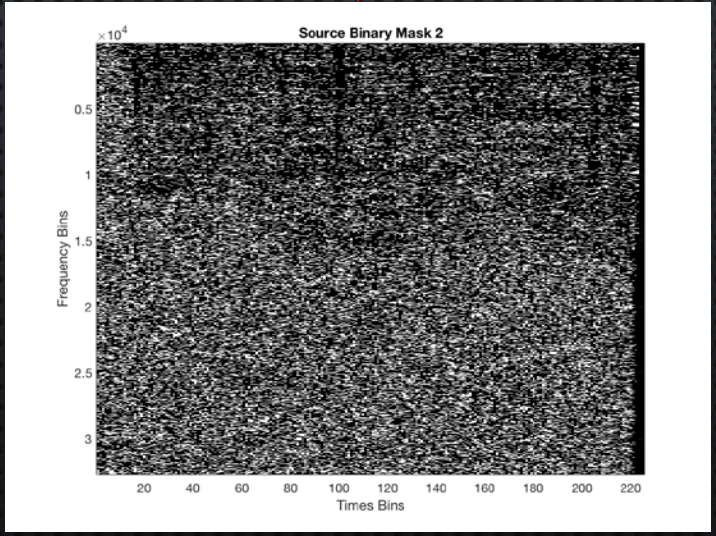
Algorithm: Finding Binary Source Masks

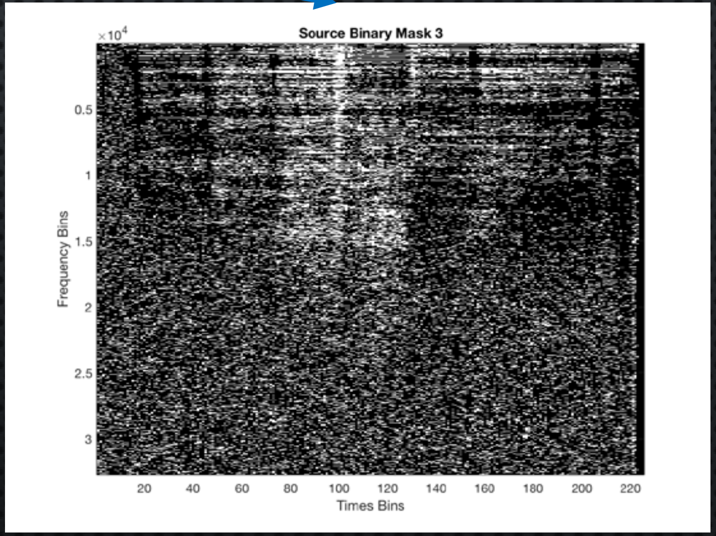




Apply thresholds to divide the pan map into binary source masks

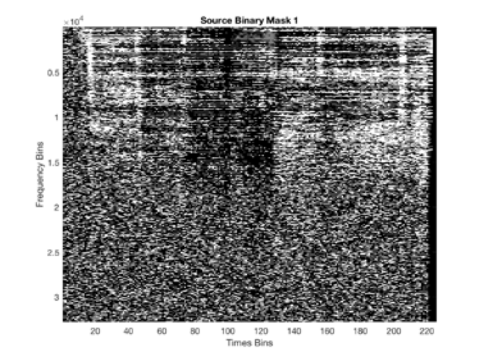




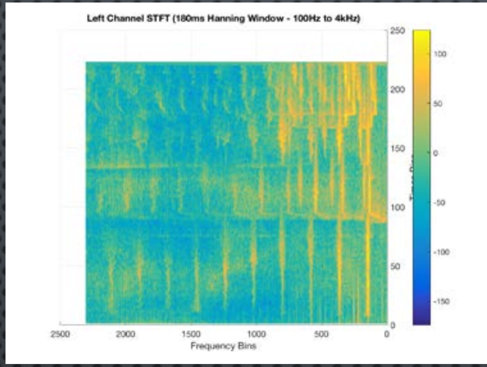


Algorithm: Extract Audio Source

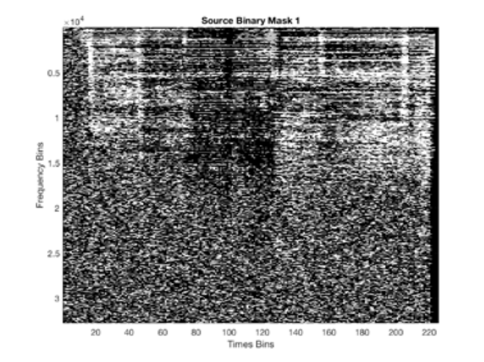
Left Channel



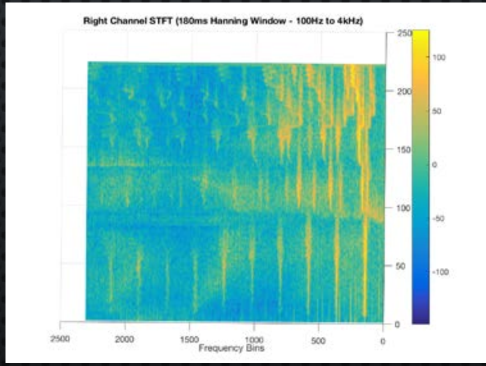
×



Right Channel



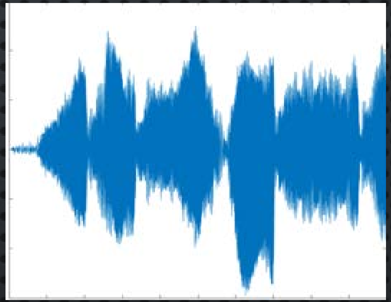
×



+

ISTFT

Extracted Audio Source



Algorithm: Azimuth Estimation

Need to find

Left Channel Gain = G_L Right Channel Gain = G_R

Already know

$'x'$ In Left Channel = xG_L ‘x ‘ In Right Channel = xG_R

By assuming (constant power panning)

$$G_L^2 + G_R^2 = 1$$

Can be found

$$G_L = \frac{xG_L}{\sqrt{(xG_L)^2 + (xG_R)^2}} \qquad G_R = \frac{xG_R}{\sqrt{(xG_L)^2 + (xG_R)^2}}$$

Algorithm: Effect of Width Angle

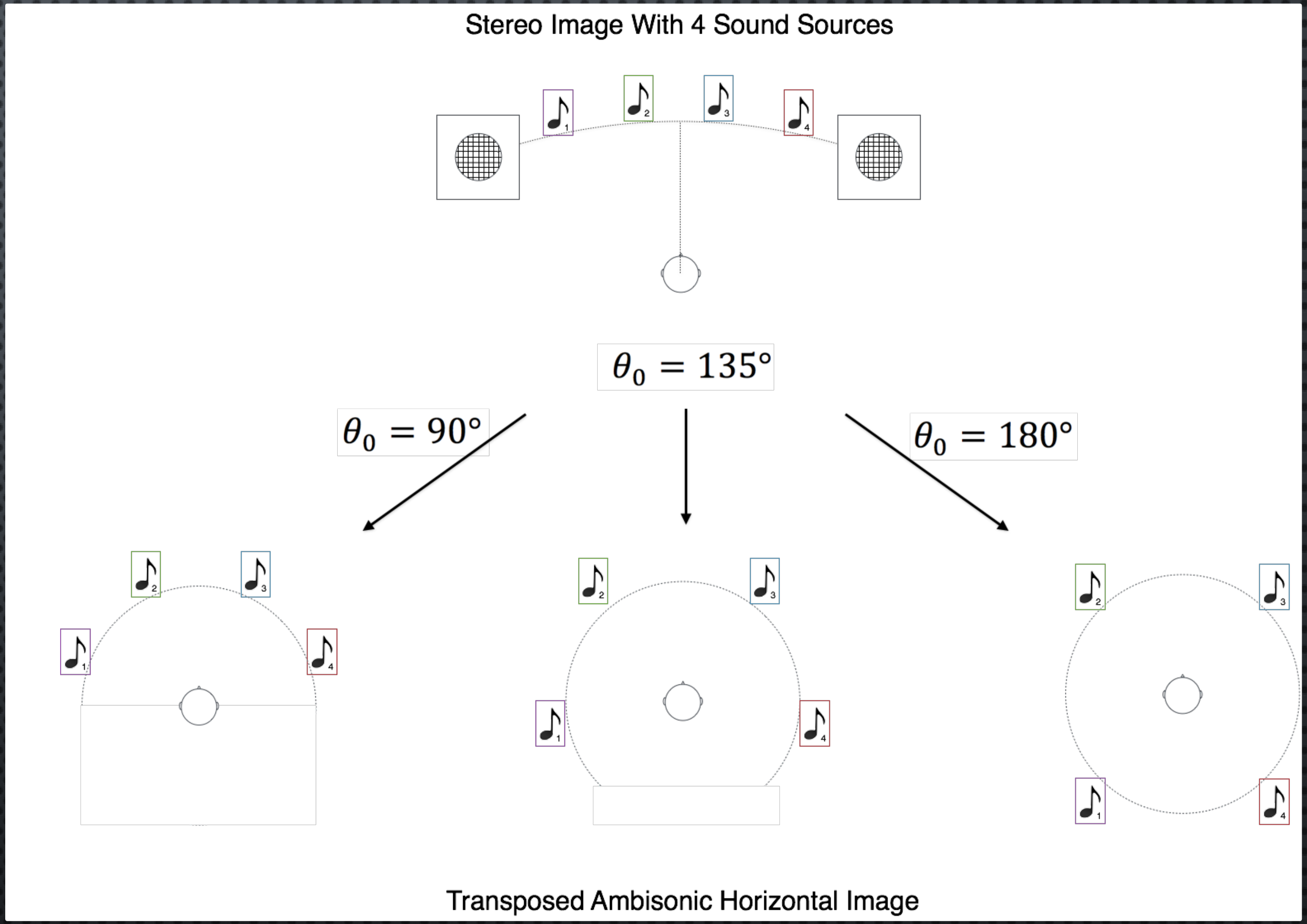
Tangent Panning Law
(Constant Power Panning)

$$\frac{\tan(\theta)}{\tan(\theta_0)} = \frac{G_L - G_R}{G_L + G_R}$$

$$\theta = \tan^{-1}\left(\tan(\theta_0) \cdot \frac{G_L - G_R}{G_L + G_R}\right)$$

θ = Source Azimuth

θ_0 = Desired Width



(Griesinger, 2002)

Algorithm: B-format Encoding

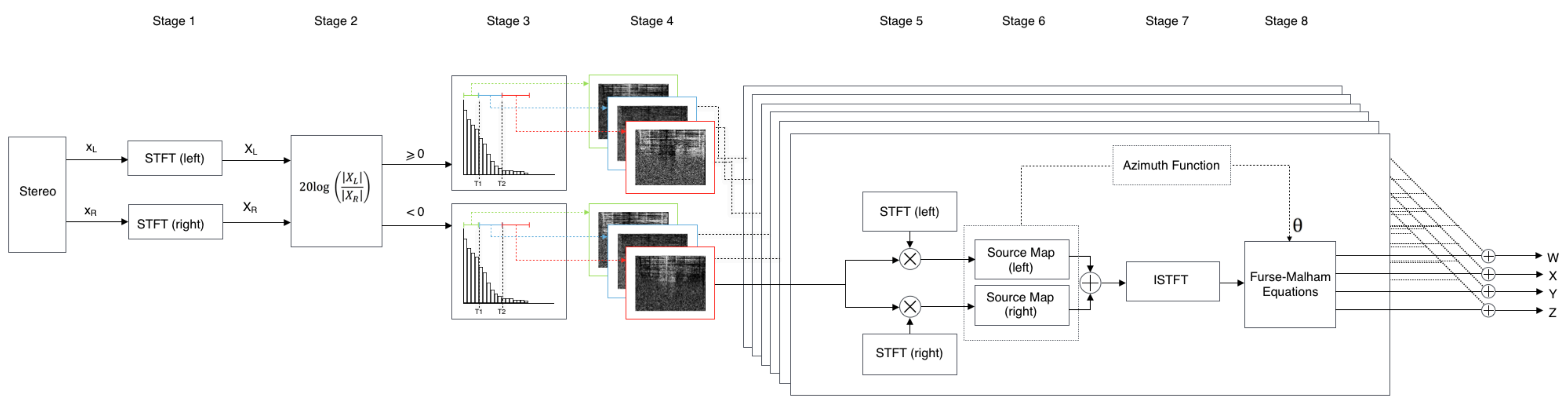
Label	Order	Angle/Elevation Representation
W	0	sqrt(1/2)
X	1	cos(A)cos(E)
Y	1	sin(A)cos(E)
Z	1	sin(E)
R	2	(1/2)((3sin(E)sin(E)-1)
S	2	cos(A)sin(2E)
T	2	sin(A)sin(2E)
U	2	cos(2A)cos(E)cos(E)
V	2	sin(2A)cos(E)cos(E)
K	3	(1/2)sin(E)(5sin(E)sin(E)-3)
L	3	sqrt(135/256)cos(A)cos(E)(5sin(E)sin(E)-1)
M	3	sqrt(135/256)sin(A)cos(E)(5sin(E)sin(E)-1)
N	3	sqrt(27/4)cos(2A)sin(E)cos(E)cos(E)
O	3	sqrt(27/4)sin(2A)sin(E)cos(E)cos(E)
P	3	cos(3A)cos(E)cos(E)cos(E)
Q	3	sin(3A)cos(E)cos(E)cos(E)

Elevation = 0°

Label	Order	Angle/Elevation Representation
W	0	sqrt(1/2)
X	1	cos(A)
Y	1	sin(A)
Z	1	0
R	2	(1/2)(-1)
S	2	0
T	2	0
U	2	cos(2A)
V	2	sin(2A)
K	3	0
L	3	sqrt(135/256)cos(A)(-1)
M	3	sqrt(135/256)sin(A)(-1)
N	3	0
O	3	0
P	3	cos(3A)
Q	3	sin(3A)

(Blue Ripple Sound, 2015)

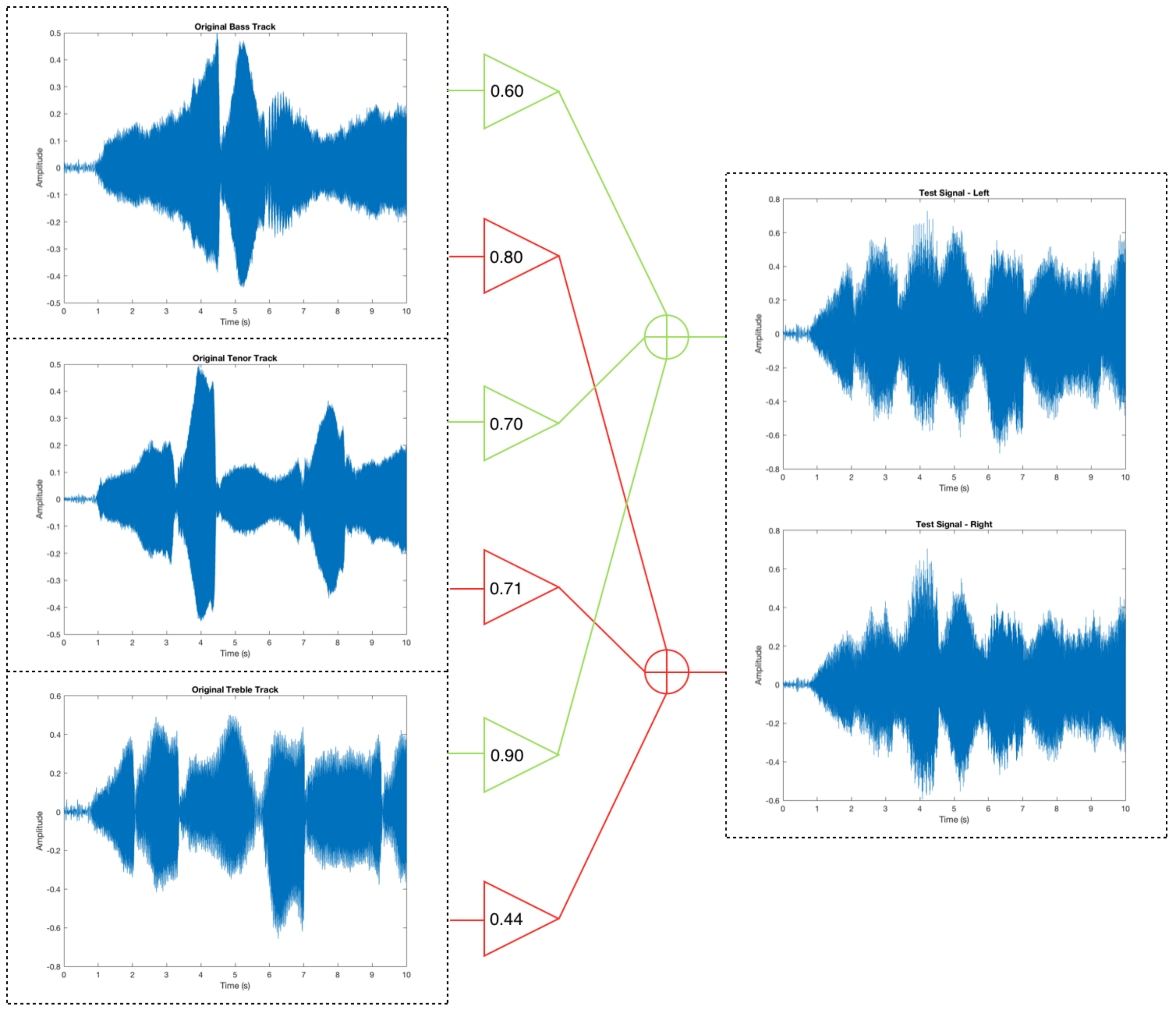
Algorithm: Overview



(converting stereo to 1st order B-format using 2 optimum thresholds)

Testing

Testing: Source Extraction



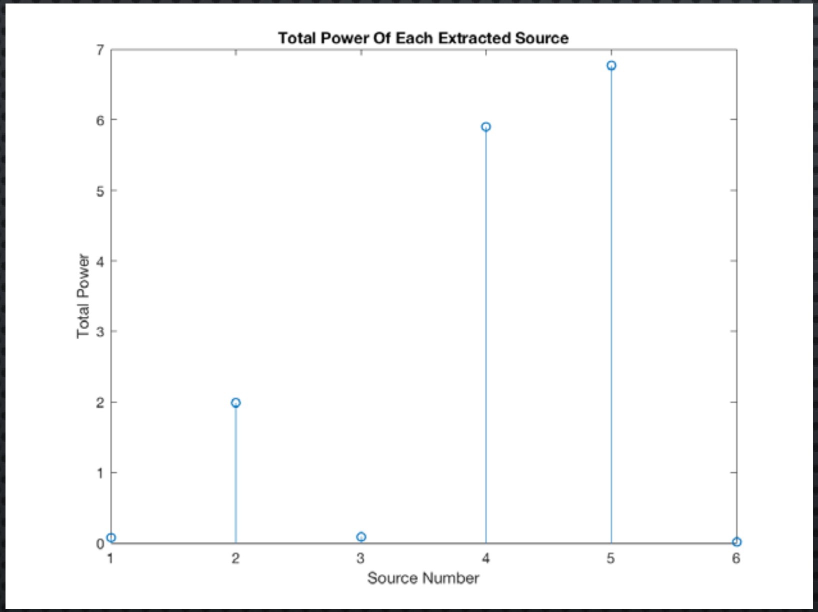
A test signal consisting of 3 audio sources was created. Using the blind source separation technique described previously the sources and their respective panning coefficients were extracted using 2, 3 and 4 optimum thresholds.

Testing: Source Extraction

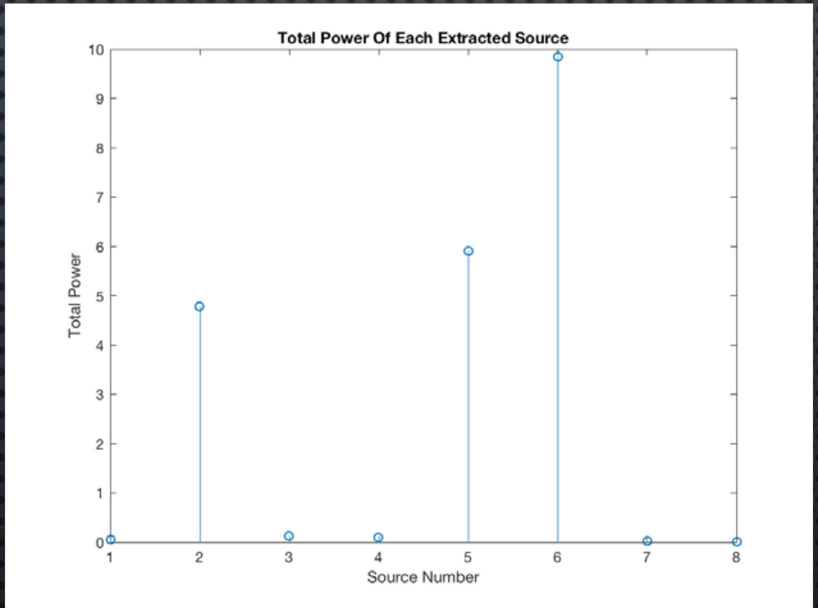
			Number Of Thresholds Used			
			2	**3**	**4**	**Original Source**
	Time Taken For Calculation (s)		7.55	9.98	11.03	
Extracted Source	**Treble**	**Gain (L)**	0.8945	0.9014	0.8977	0.9000
		Gain (R)	0.4471	0.4329	0.4406	0.4359
		Correlation	0.9780	0.9621	0.9613	
		SNR (dB)	20.76	21.07	20.97	
	Tenor	**Gain (L)**	0.6956	0.6956	0.6975	0.7000
		Gain (R)	0.7185	0.7185	0.7166	0.7141
		Correlation	0.9848	0.9848	0.9855	
		SNR (dB)	20.36	20.36	20.37	
	Bass	**Gain (L)**	0.5982	0.5995	0.6023	0.6000
		Gain (R)	0.8014	0.8004	0.7982	0.8000
		Correlation	0.9884	0.9882	0.9876	
		SNR (dB)	20.36	20.38	20.39	

Testing: Over Extraction

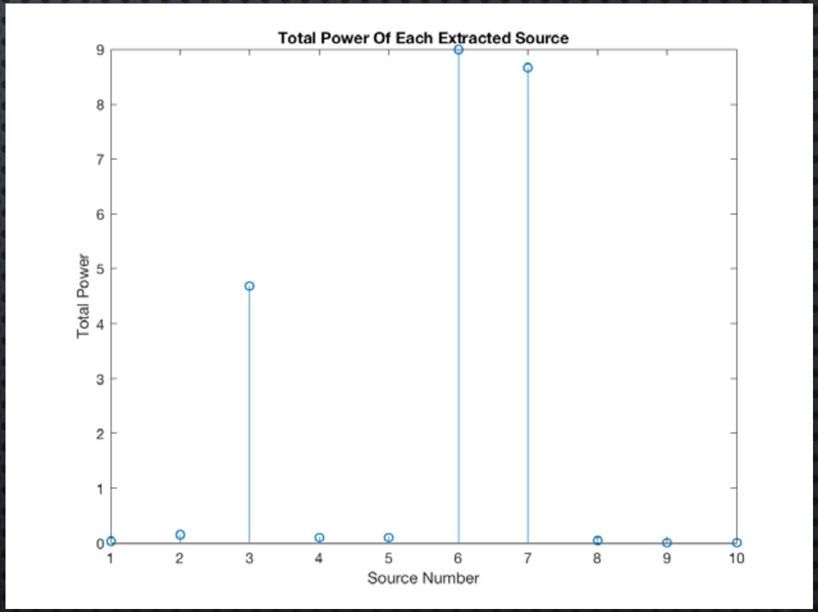
What if there are less sources in the audio than you are trying to extract?



(using 2 thresholds)



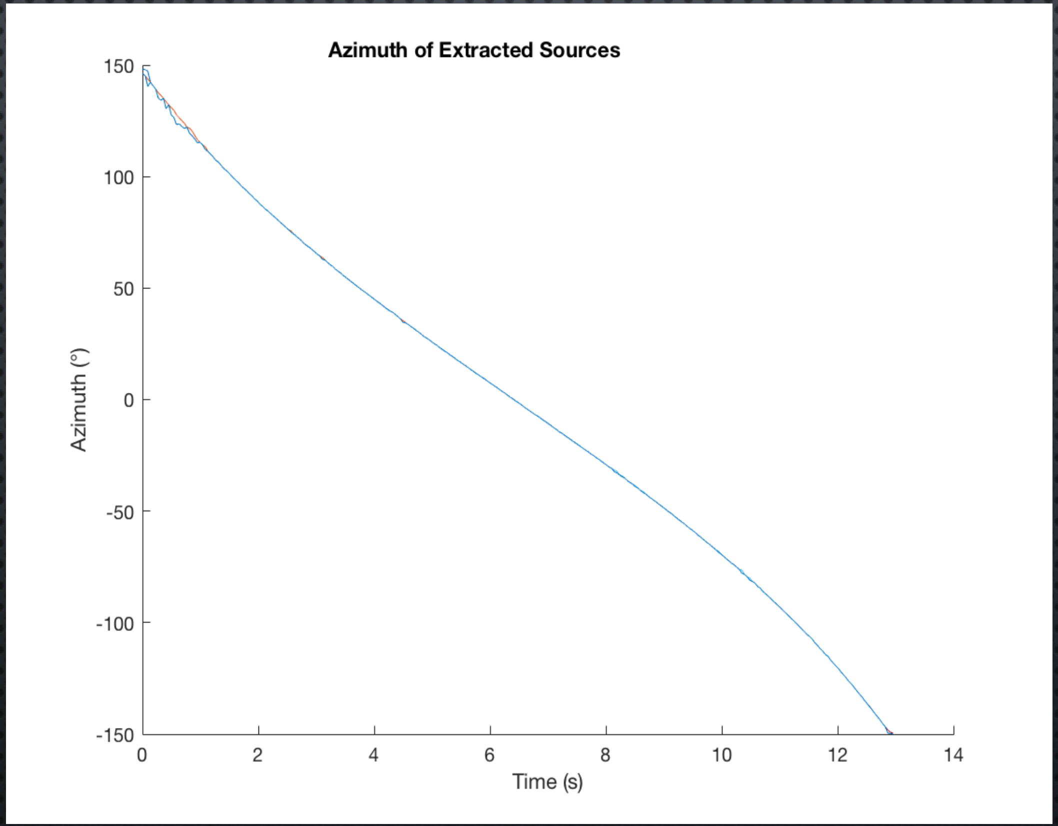
(using 3 thresholds)



(using 4 thresholds)

Testing: Source Movement

A test signal was created panning an audio source from hard-right to hard-left.



Conclusion

Conclusion: Advantages and Disadvantages

Advantages:

- The algorithm is capable of converting any standard commercial stereophonic signal to up to 3rd order B-format.
- It is shown that sound sources within the audio are extracted with a high degree of accuracy and, for constant power panned sources, their azimuth will be also be extracted accurately.
- The user is able to define the horizontal width of the outputted signal.

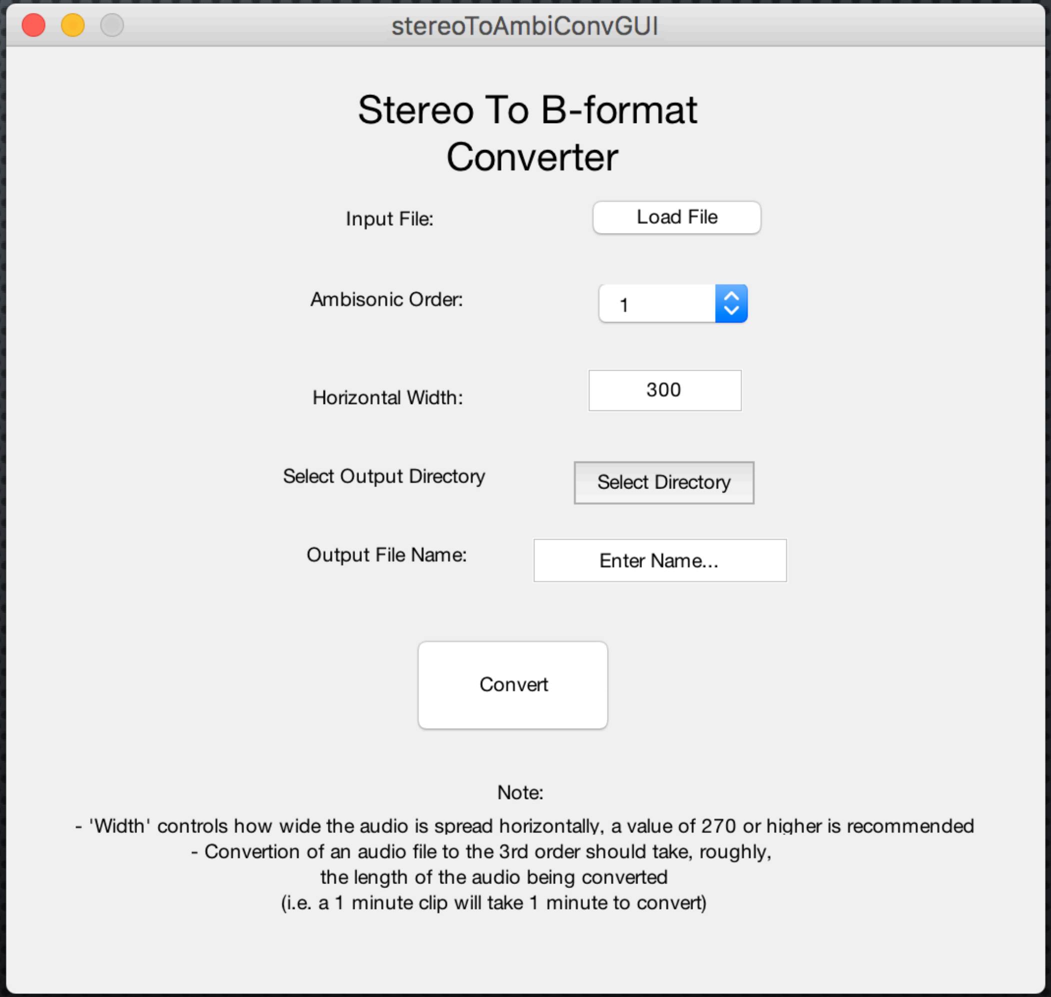
Disadvantages:

- When used with a signal with few sources noise can be picked up as a source and panned irregularly.
- The panning extraction is designed specifically for constant power panning; other panning laws will still work but not as accurately.
- No height information is outputted.

Conclusion: Future Work

Next Steps:

- Real-time implementation
- Listening tests
- Add in height?
- Combine other blind source separation methods (allow for diffuse and direct separation)



References

Blue Ripple Sound (2015) **HOA Technical Notes – B-Format** [online] Available at: http://www.blueripplesound.com/b-format (accessed: 2 Sept. 2016)

Cobos, M. and Lopez, J. (2008) **Stereo Audio Source Separation Based on Time-Frequency Masking and Multilevel Thresholding**. Digital Signal Processing 18 (960-976)

Griesinger, D (2002) **Stereo and Surround Panning in Practice.** Audio Engineering Society 112th Convention.

Liao, P. Chen, T. and Chung, P. (2001) **A Fast Algorithm For Multilevel Thresholding**. J. Inform. Sci. Eng. 17. P. 713–717.

Otsu, N (1979) **A Threshold Selection Method from Grey-Level Histogram.** IEEE Trans. System Man Cybernet. SMC-9 (1) 62–66.

Demonstration Time!

Q & A

Contact: haydon.cardew@mqa.co.uk