DR BRIAN BRIDGES

BD.BRIDGES@ULSTER.AC.UK

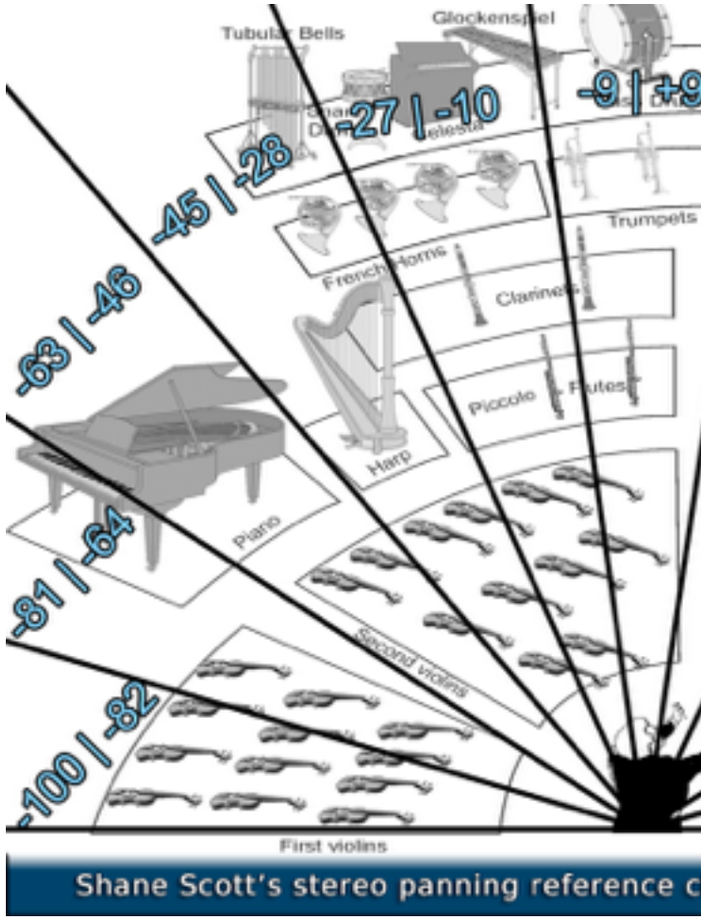
MUS302 WEEK 5
INTRO TO SPATIAL AUDIO AND SPATIAL MUSIC

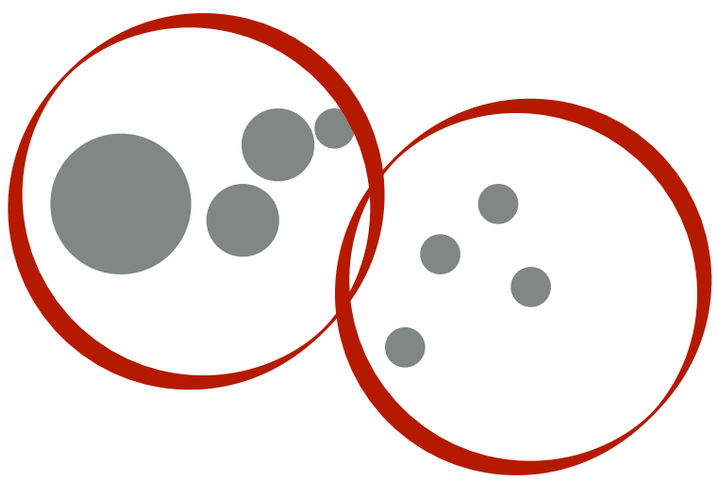
SPATIAL SOUND, SPATIAL MUSIC

- We all have some basic awareness of spatial sound
- We know we have two ears (and our main audio standard uses two speakers)
  - We know that the experience of listening over headphones is different (more intimate?) from that of listening over speakers, though we may not notice exactly how it is different…how exactly is it different?
- We know that certain animals can move their ears to track sounds
- We know that surround sound (extra speakers) can give us front/back perspective in cinemas, etc.
- If we do audio production, we understand that panning controls can 'swing' a sound between left and right, or around in a circle

BUT WHY ARRANGE MUSIC SPATIALLY?

- Why is any of this important for music? Do we really need to move sounds in space: is it just a gratuitous special effect? Should music/sound really be spatial? Why do we want sounds to surround us?
- Why do we place different musical instruments/sound sources in different apparent spatial locations? What are we trying to achieve? What can it help us with?
- How does this relate to musical structure? (Do different melodic notes mean different spatial locations? why/why not?)
- How does this relate to the structure of our wider sonic environment? (What happens to our perception when sounds occur close together in space?)





SPATIAL POSITIONS, SPATIAL STREAMS

- Does space just help us hear the direction of a sound, or does it do something else for our perception?
- What happens to the apparent tempo of this presentation as we change our spatial perspective?

HOW ELSE ARE SOUNDS SPATIAL?

- Spatial difference helps us separate sounds into different sound streams/sound objects (based on Albert Bregman's auditory stream segregation effect)=> one stream becomes two due to panning every second note left and right
- Does this provide us with one answer as to why we need space in music production activities? (Are there any other reasons why we might want to pan different sound sources...maybe ones which are close together in frequency?)
- Spatial audio techniques therefore allow us to associate or dissociate sound sources/sound events and craft a sense of musical structure through apparent causality
    - Spatial audio is therefore able to turn 'normal' musical motion in pitch and time into spatial motion, creating a sense of 'sonic virtual reality'

INTRODUCTION TO SPATIAL AUDIO

- Spatial audio approaches attempt to create the impression of spatially displaced sound sources by using speakers or headphones to supply some of the **localisation cues** which we use to decode source direction in environmental sound

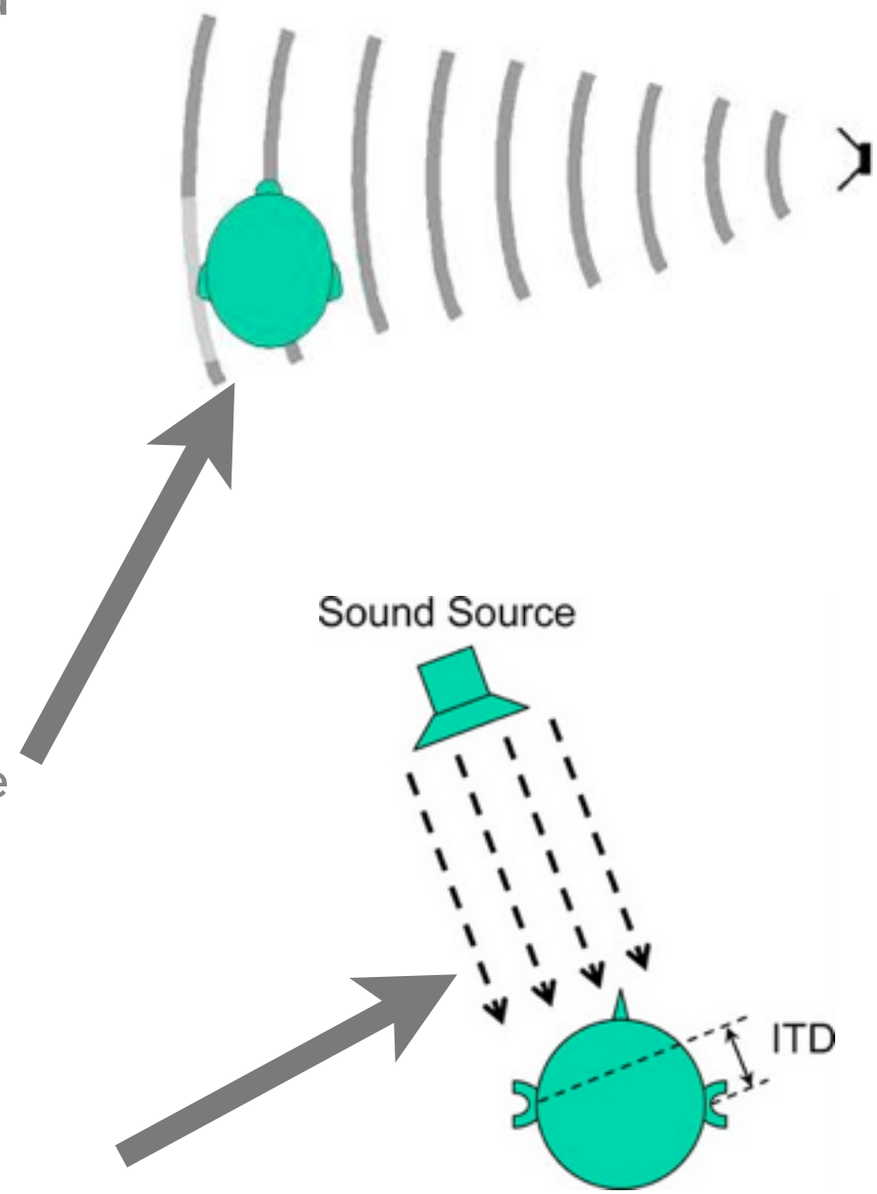
- In other words, with spatial audio, we're trying to configure our audio signal so that it 'fools' us into hearing a virtual sound source coming from a particular direction

- Stereo is a basic form of spatial audio, which uses sound level differences between two equidistant speakers to create an impression of left–to–right perspective between the speakers

- More advanced spatial audio techniques are possible when using headphones or when using multiple loudspeakers

HEARING SOUND IN SPACE: BASIC LOCALISATION CUES

- We have the ability to **hear the horizontal direction of a sound due to the fact that we have two ears**
- **We compare (A) the level and (B) the arrival time for sound sources at our two ears**
- We use these differences in the signal to decode direction
- **If the signal is near-identical at both ears**, we understand that the **sound is coming from straight ahead** (or straight behind)
- **If there is a difference in sound level, we assume that the sound is coming from the side at which it is loudest** (this is the most common stereo cue used in audio production)
- **If there is a difference in arrival time** between the different sound waves at the two ears, we assume that **the sound is coming from the side which detects the sound event first**



CONTRASTING SOUNDS, BIG DIFFERENCES

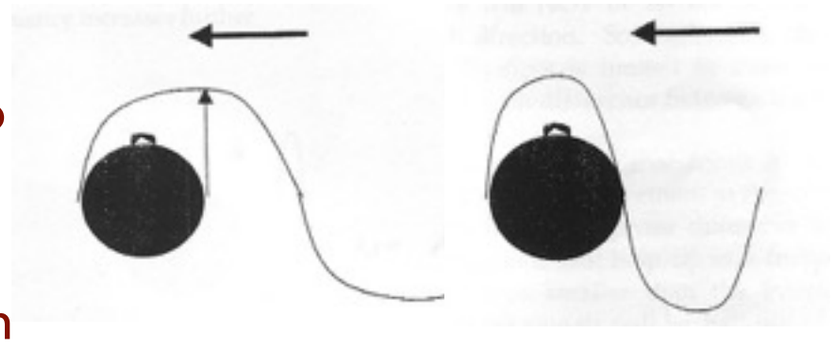
- Short and long sounds have very different spatial effects
- Short: we hear the full articulation of a note repeated…we often have noisy transient elements in sounds => easier to hear direction… (we will find out exactly why in a moment)
- Long sounds in music: tones/drones…we focus on the sustain portion, we hear the harmonic texture, we don't focus on how the sound started (attack transient) => harder to hear direction (we will find out exactly why in a moment)

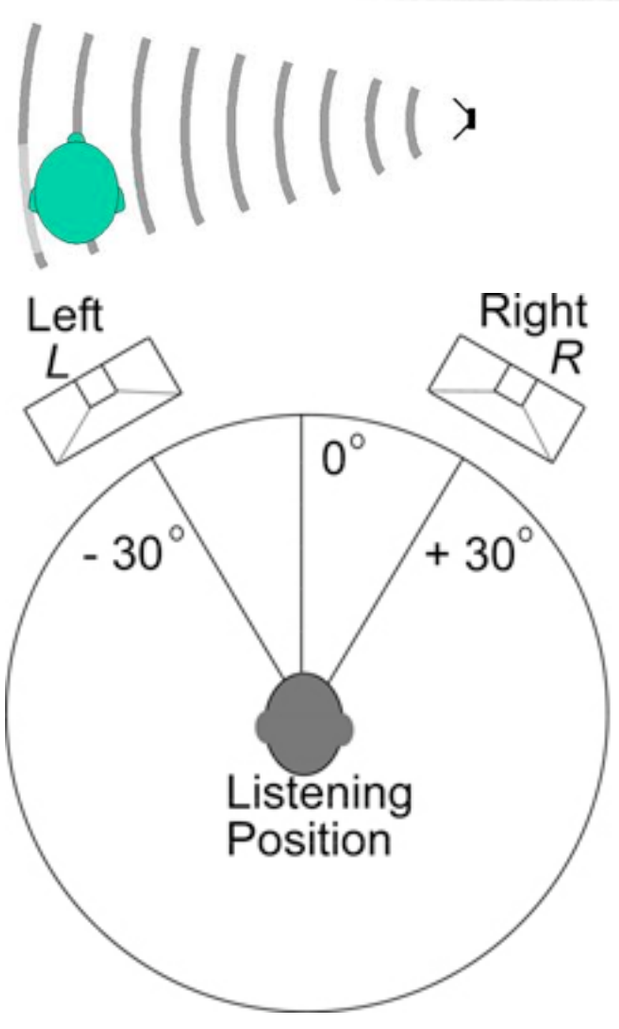
EXAMPLES 1&2: SCATTER (JONATHAN NANGLE) AND SPEAR FRAGMENT (IAN MCDONNELL)

- What do we notice about the materials chosen (and the different effects of these materials?)
- What do we notice about the sense of spatial perspective and grouping?
- Do any other factors contribute to the sense of structure of the pieces?
- (The pieces are by Dublin–based composers who are members of the Spatial Music Collective, which specialises in spatial music in both electronic and acoustic forms)

EXPLAINING THE DIFFERENCE: SPATIAL AUDIO CUES

- Sustained tones: hard to track the time difference between signal at two different ears (many cases where cycle length will mean no clear difference in signal at two ears)
- Short (and high) tones: either a clear time difference or more high frequency content (hence more level difference due to head shadowing)
- Q:Which cue is more important for us in 'standard' stereo setups?
- As a result, which type of materials should we use in spatial music if we wish to hear direction clearly?
- Are there any conceivable circumstances in which we might want to use longer, sustained tones (either from an instrument, or from the acoustic response of a space)?



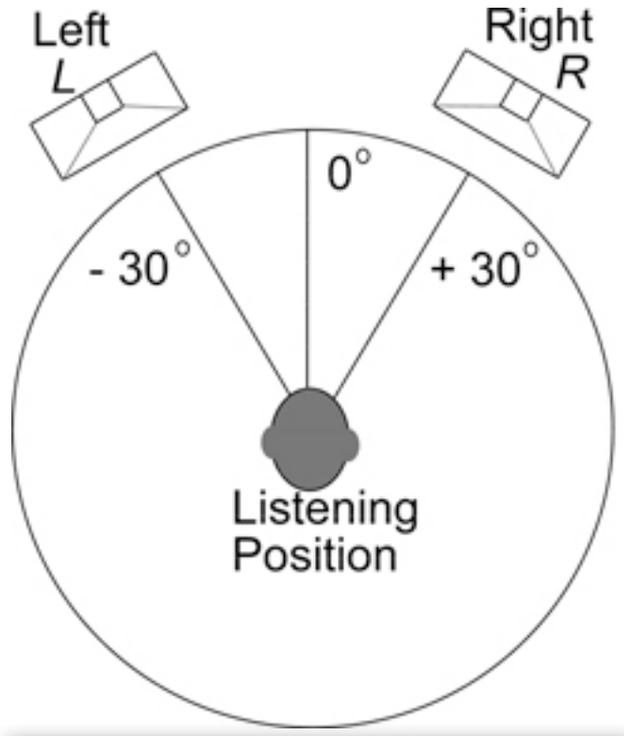


ARE SOUNDS SPATIAL? SEE (ERICKSON, 1975)

- You've begun to examine environmental sound as part of this module
- Have you identified any high frequency/short/transient sources? What sort of objects/actions produce these sources? How do they 'impose' themselves on the soundscape? Can you associate any descriptive terms with them?
- Have you identified any sustained sources? What sorts of objects/actions produce these sources? Which features do you concentrate on with these sources? Can you associate any descriptive terms with them?
- How could these different aspects relate to music/sonic structure in your pieces? How does your source material change your sense of structure?

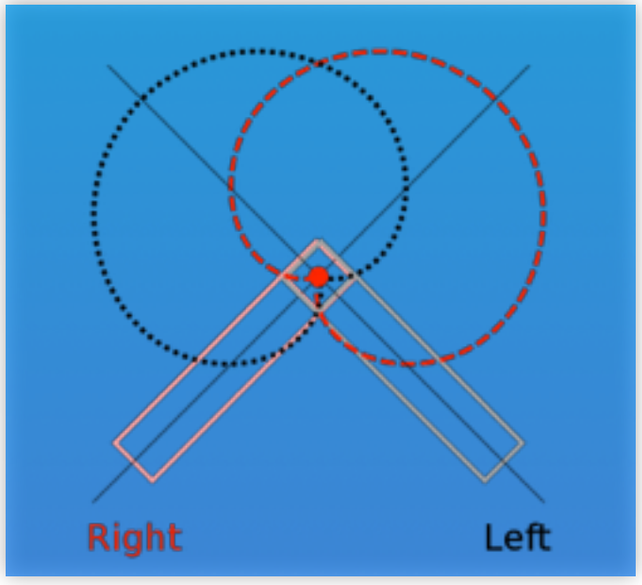
STEREO CUES AND PANNING (FOR LOUDSPEAKERS)

- Two speakers (at least) are used. Stereo speaker configuration - equilateral triangle with loudspeakers forming the base and the listener at the apex. Loudspeakers 30 degrees off centre on each side in relation to listener.



- The listener is in front of a pair of loudspeakers which are equidistant. There will be no time difference cues. We use level difference cues whereby a pan pot will increase level in one channel and decrease in another.

- We can also record this type of cue using **coincident pairs** of directional (e.g. cardiod) microphones--**xy pair**--(lower level off-axis response simulates head shadowing effect)



SPATIAL CONVENTIONS AND SPATIAL CONFUSION

- How has the use of stereo spatialization changed over the course of popular music history?
- Are there any particularly notable conventions now? How does stereo interact with the musical materials we use?
- Have you ever listened to stereo and mono versions of the same album? Do you hear different details in the music in each version?
- Have you ever had the experience of confusing the direction of a sound event? Have you noticed any common factors which contribute to this confusion? Are these related to the environment or the materials? Or both?

BEYOND STEREO

- Stereo presents us with a limited horizontal plane of 60 degrees (constrained lateral response) and no direct impression of verticality (though we get some from high/low pitches and bright/dark timbres)

- This is in marked contrast to our environmental experience of a spherical sound world which may present a source anywhere on a 360 degree circle in the horizontal (azimuth) or vertical (elevation) planes

- Using more speakers and a few careful combinations of audio techniques may provide us with a more convincing and immersive soundscape...this is the means by which much electroacoustic music is presented today

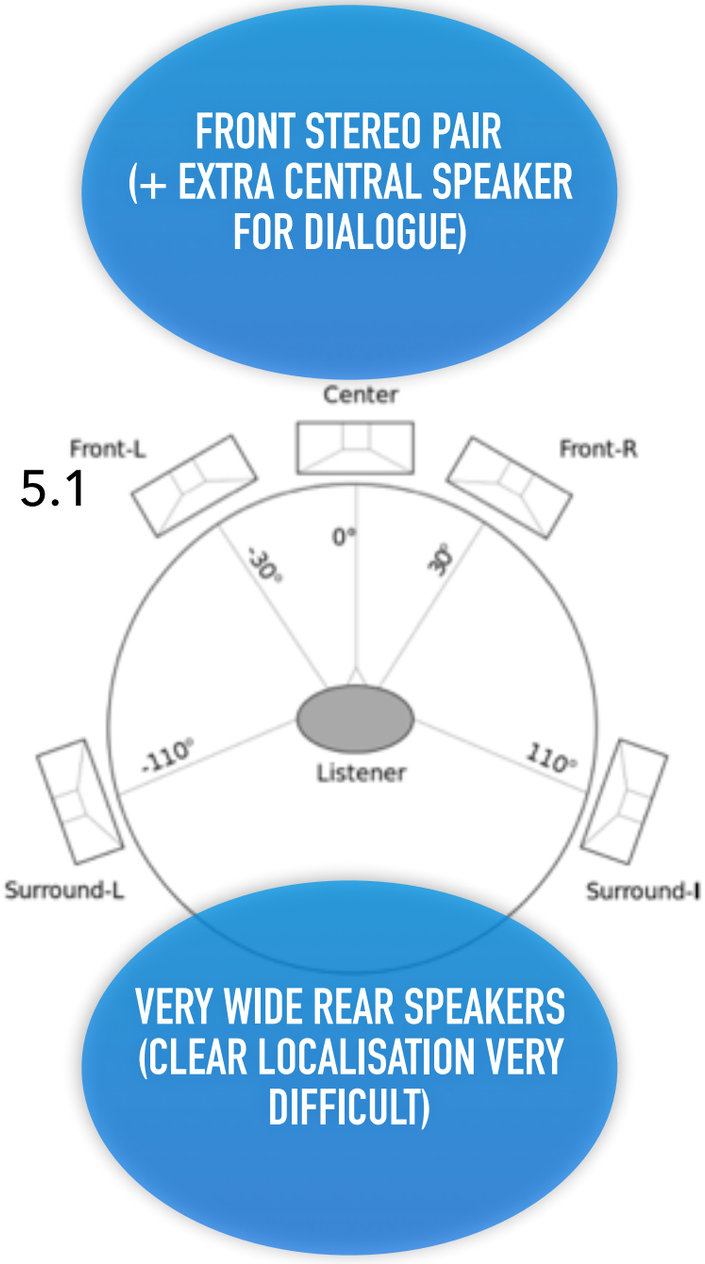
- For the present purposes, our practical focus will be on stereo spatial audio techniques: (1) 'standard' stereo for loudspeakers and (2) a surround technique known as binaural encoding/recording

‘SURROUND SOUND’ & SPATIAL AUDIO

- What we commonly term **surround sound** techniques (e.g. 5.1) are an inheritance from sound effects channels in the cinema, where the priority is to immerse the audience in multidirectional sound effects based on a front/rear foreground/background paradigm

- **For example, 5.1** (5 main speakers, 1 sub-woofer) is **basically a front/rear stereo variant** with a central channel reserved in cinematic convention for the dialogue track; the aim is immersion and incremental development upon stereo-based film sound conventions

- However, other (more developed) spatial audio approaches may be more concerned with accurate representations of directional movements between speakers



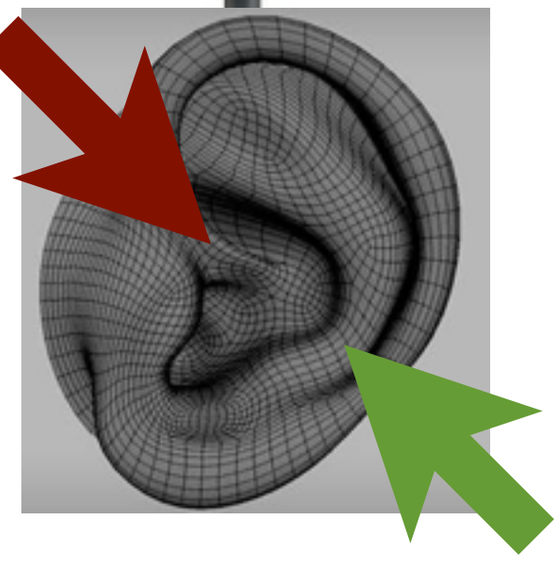
INTRODUCTION TO BINAURAL ('HEADPHONE SURROUND) ' RECORDING/PROCESSING

- 'Ear's-eye view' for 'headphone surround'
- Recording: Microphones placed in 'dummy head' simulate the frequency-dependent filtering effect of a human head/outer ears as well as general level and time differences
- Outer ears, head and upper body will produce different filtering effects for different frequencies (physical filter based on size of body 'object') which will be direction-dependent

...as if different 'colour'

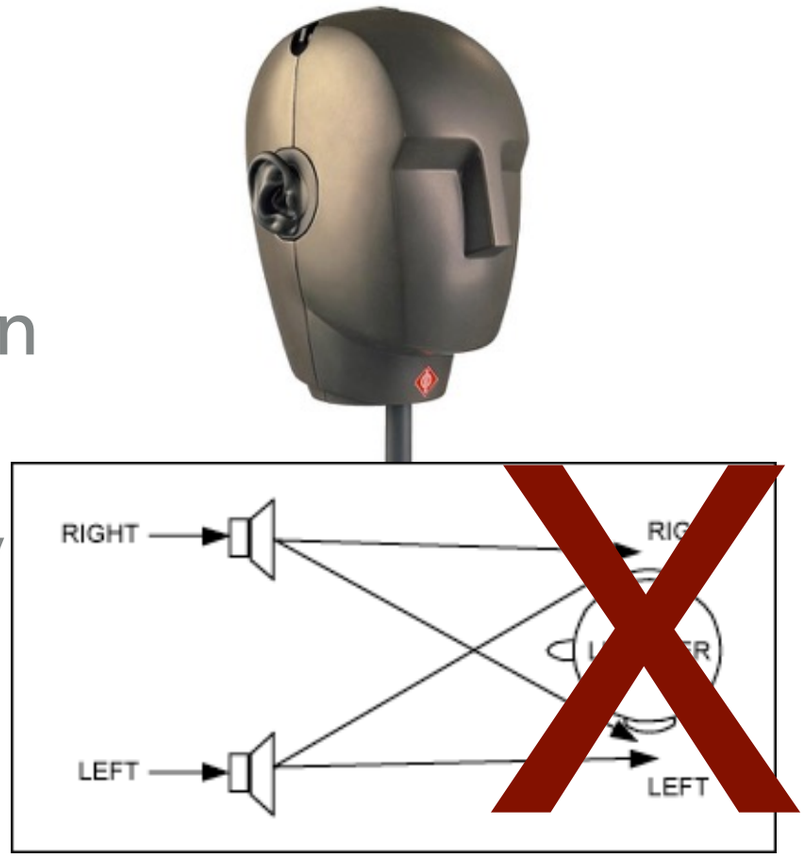
to sounds

from different directions

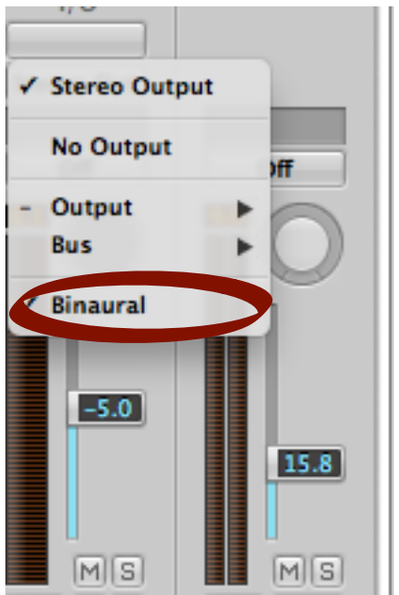


INTRODUCTION TO BINAURAL ('HEADPHONE SURROUND) ' RECORDING/PROCESSING

- 'Ear's-eye view' for 'headphone surround'
- Only effective over headphones (or in certain carefully controlled conditions)--need to keep stereo channel information completely separate to maintain clarity of cues (cannot have crosstalk between channels)

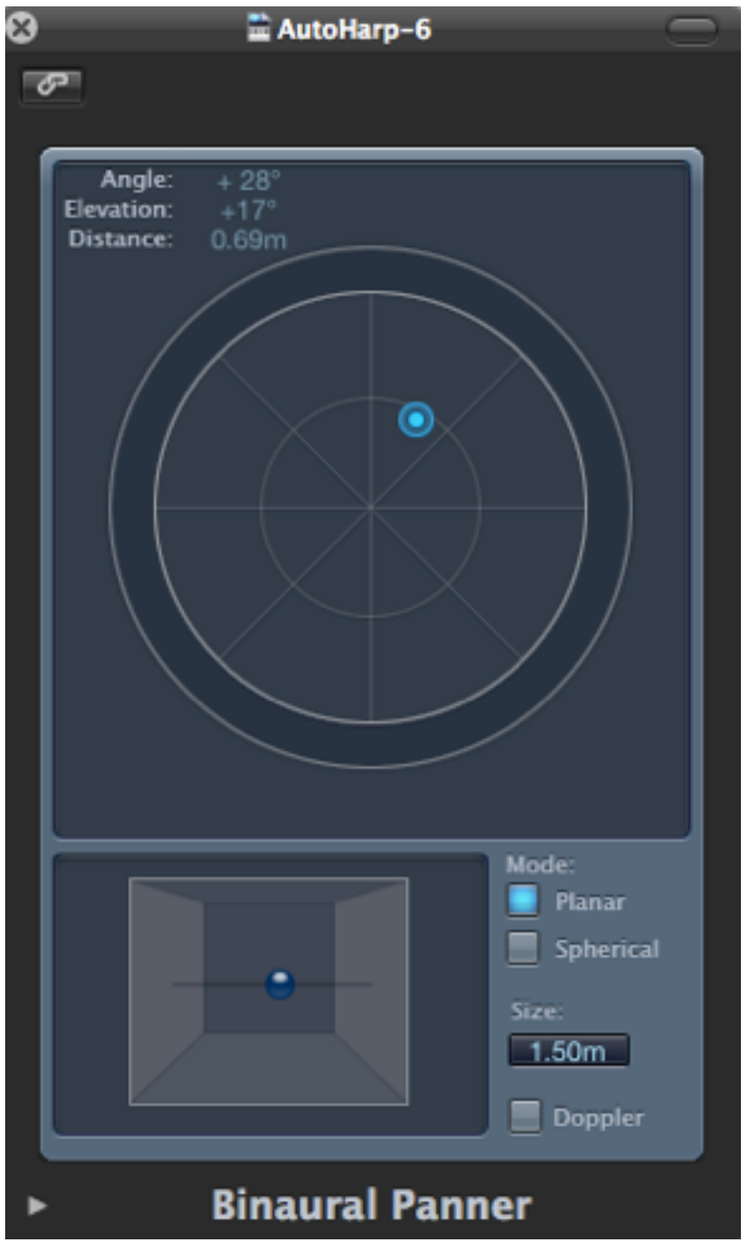


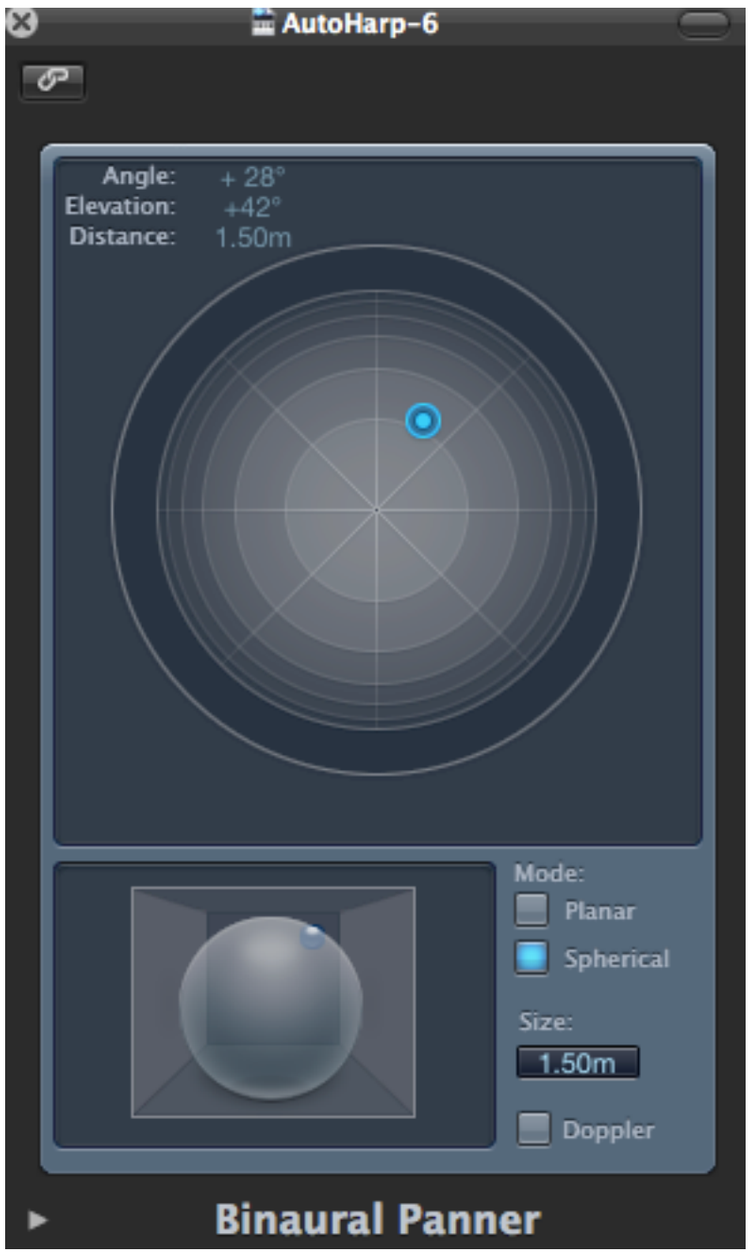
- Encoding/mixing: Logic Pro has a 'binaural panner' which applies this principle (experiment for yourself with headphones)



BINAURAL PANNER: WITH/WITHOUT VERTICAL CUES

REMINDER: MARK CLEARLY AS ONLY FOR HEADPHONES

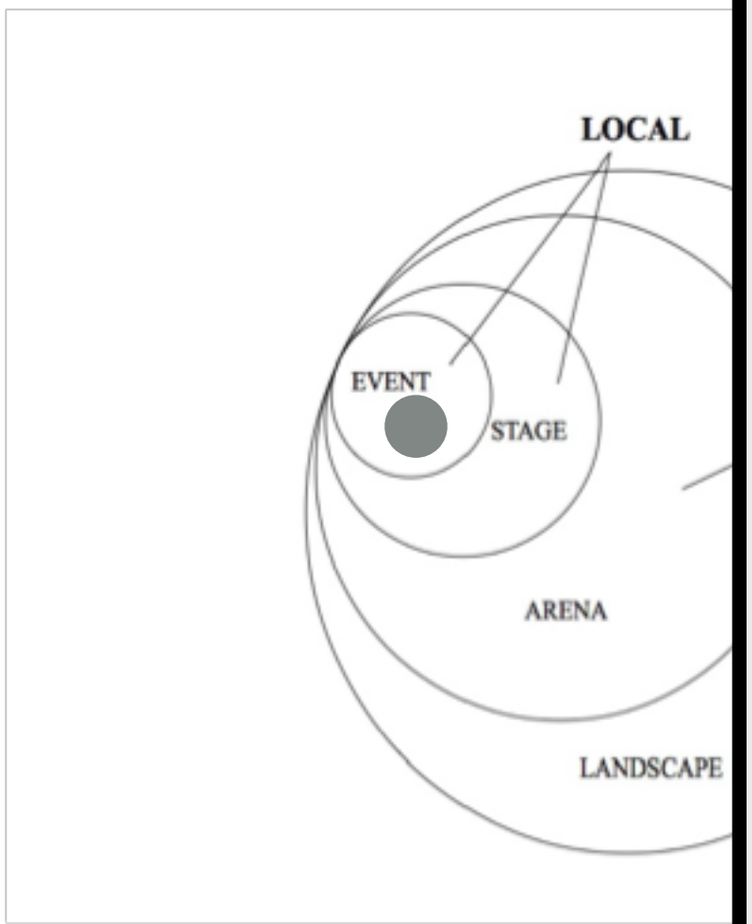




SPACES AND FRAMES

Space in music isn't just about co-ordinates

we don't have *that clear* a spatial sense...=> it's more about relationships and zones/frames



Simon **Emmerson** has

written about **performance 'space frames'**

in his book *Living Electronic Music...*

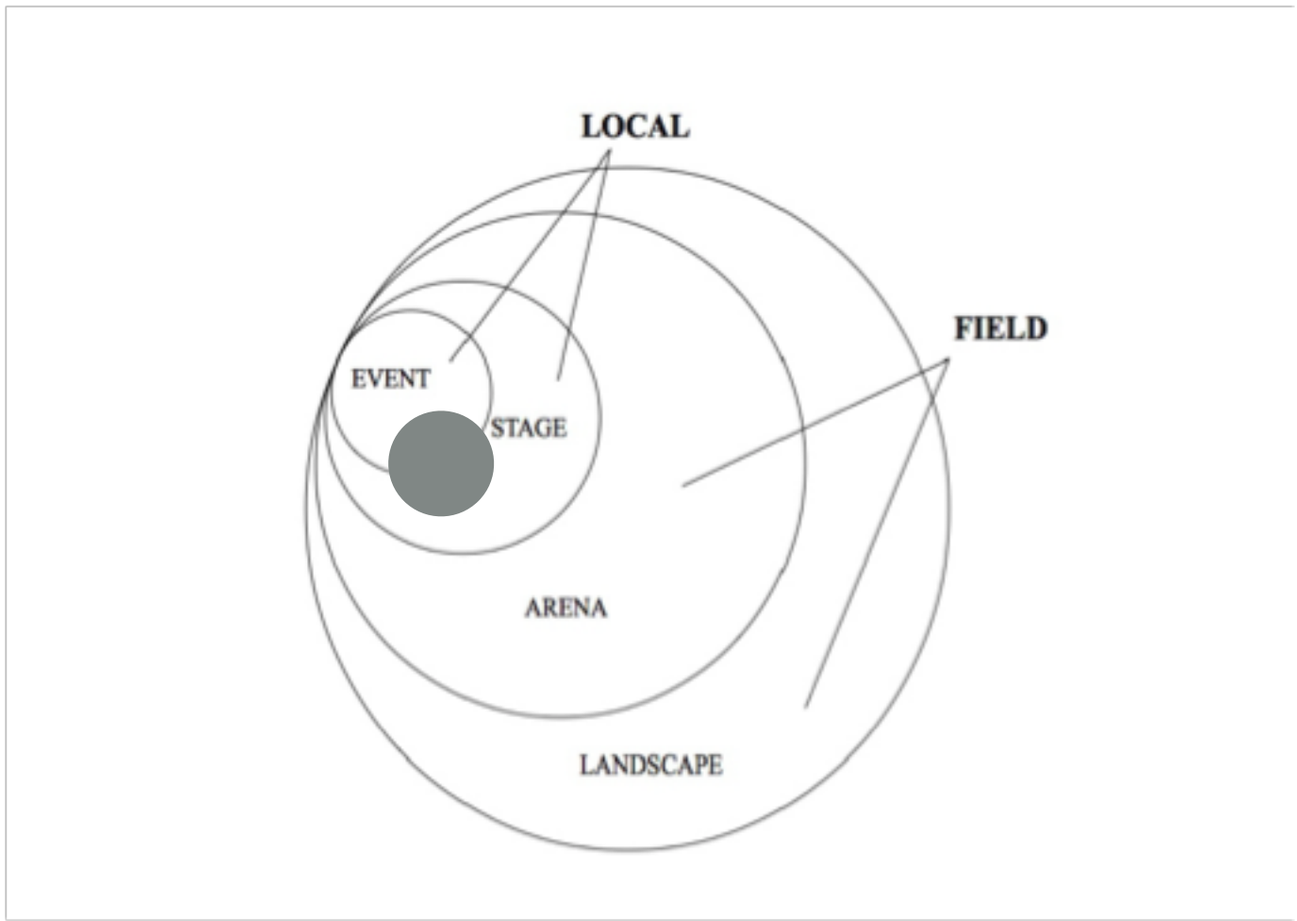
but **space frames are also important for composition**

(composition could be viewed as 'non-realtime performance')

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?

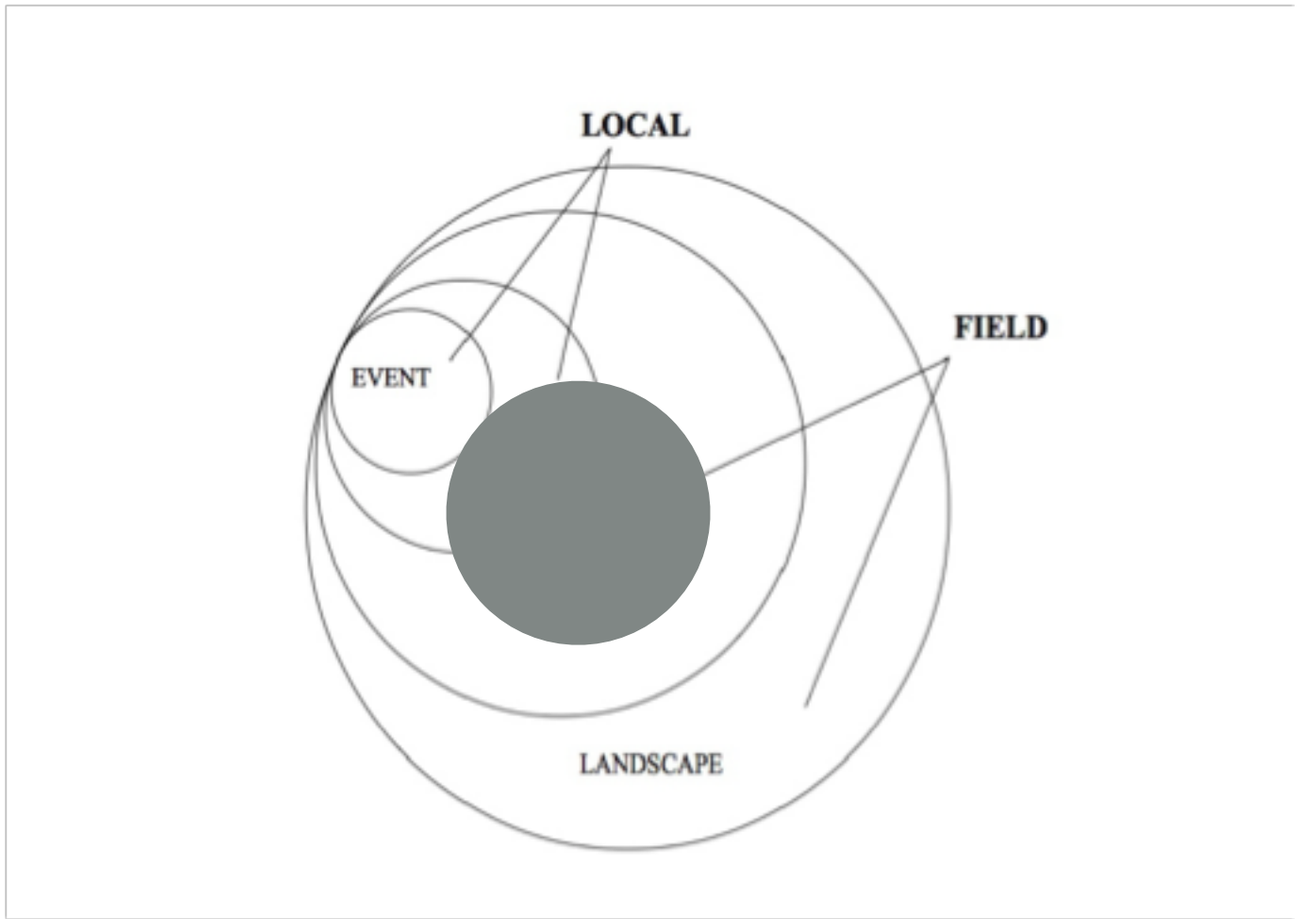


Space frames (Emerson, 2007)

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?

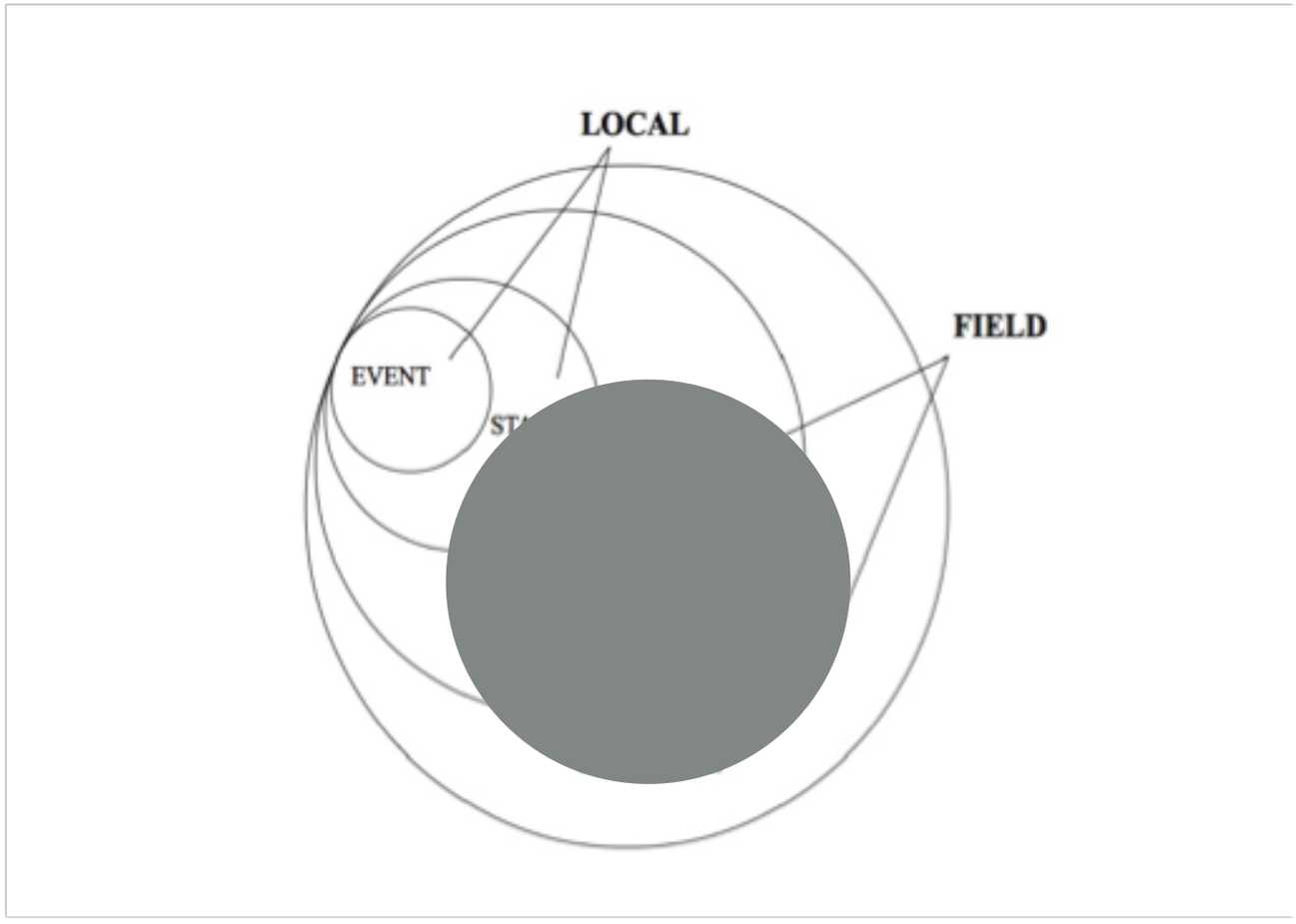


Space frames (Emerson, 2007)

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?

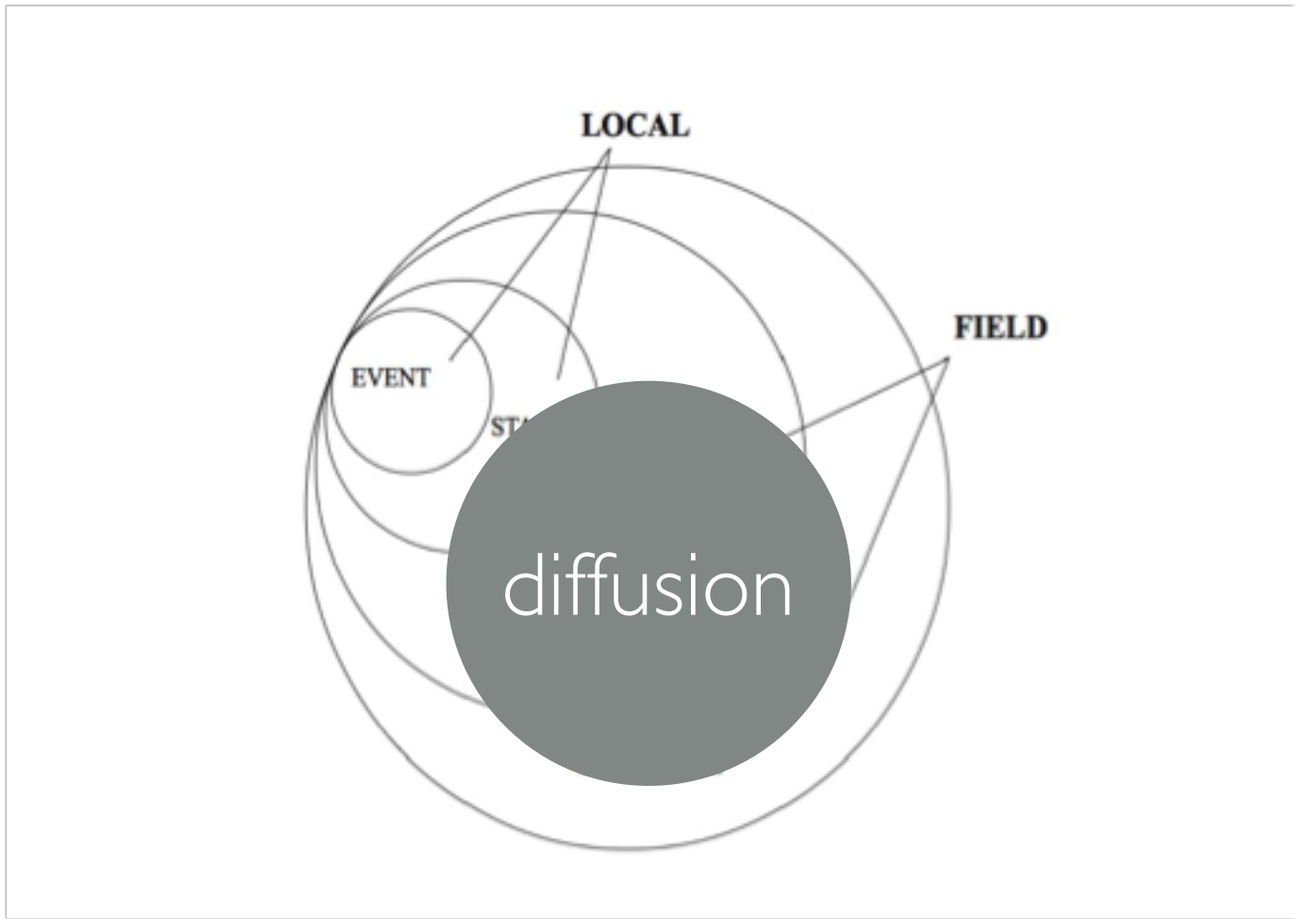


Space frames (Emerson, 2007)

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?

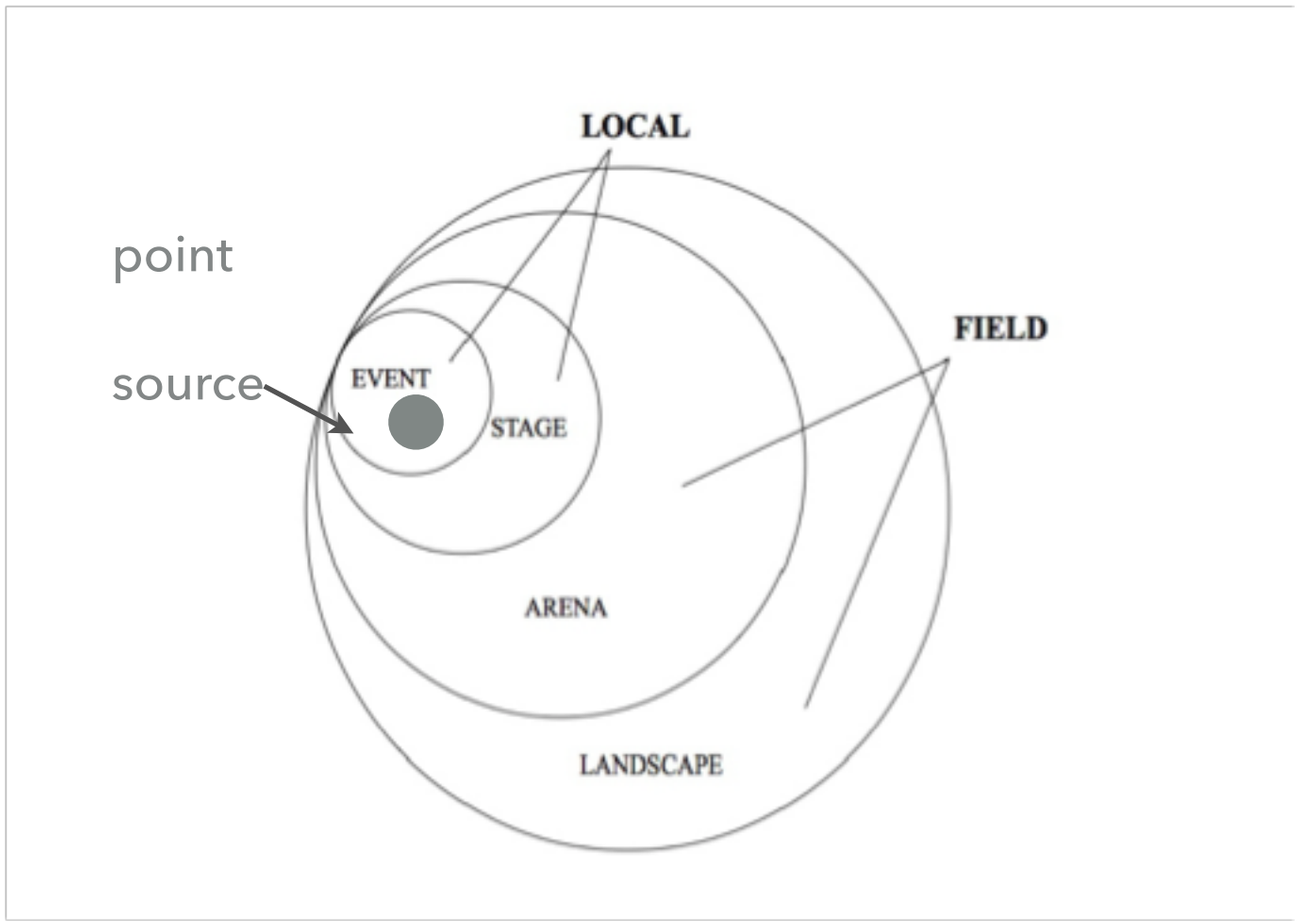


Space frames (Emerson, 2007)

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?

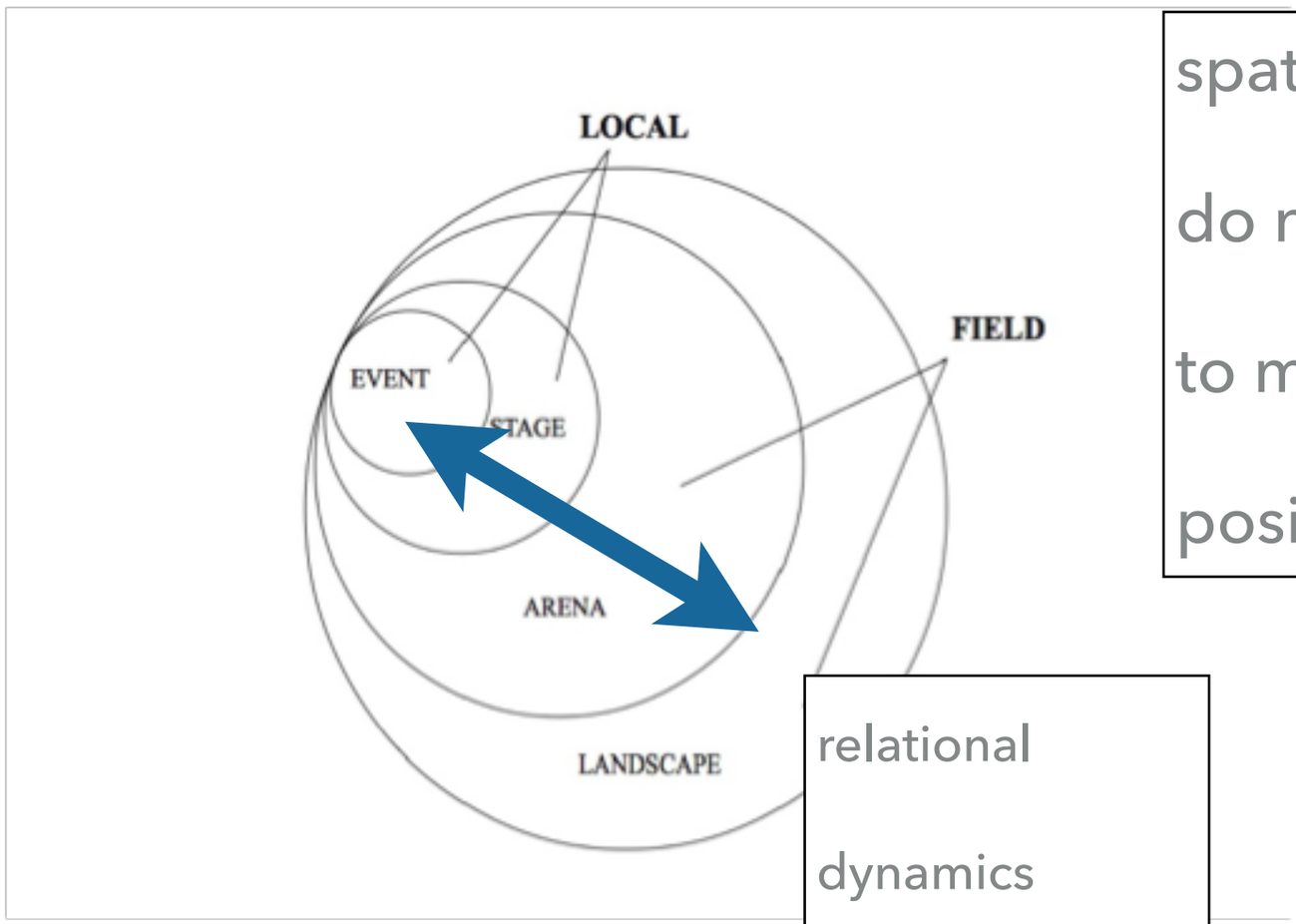


Space frames (Emerson, 2007)

SPACES AND FRAMES

how do these frames relate to our perceptual experience?

how do different types of musical sources relate to the different frames?



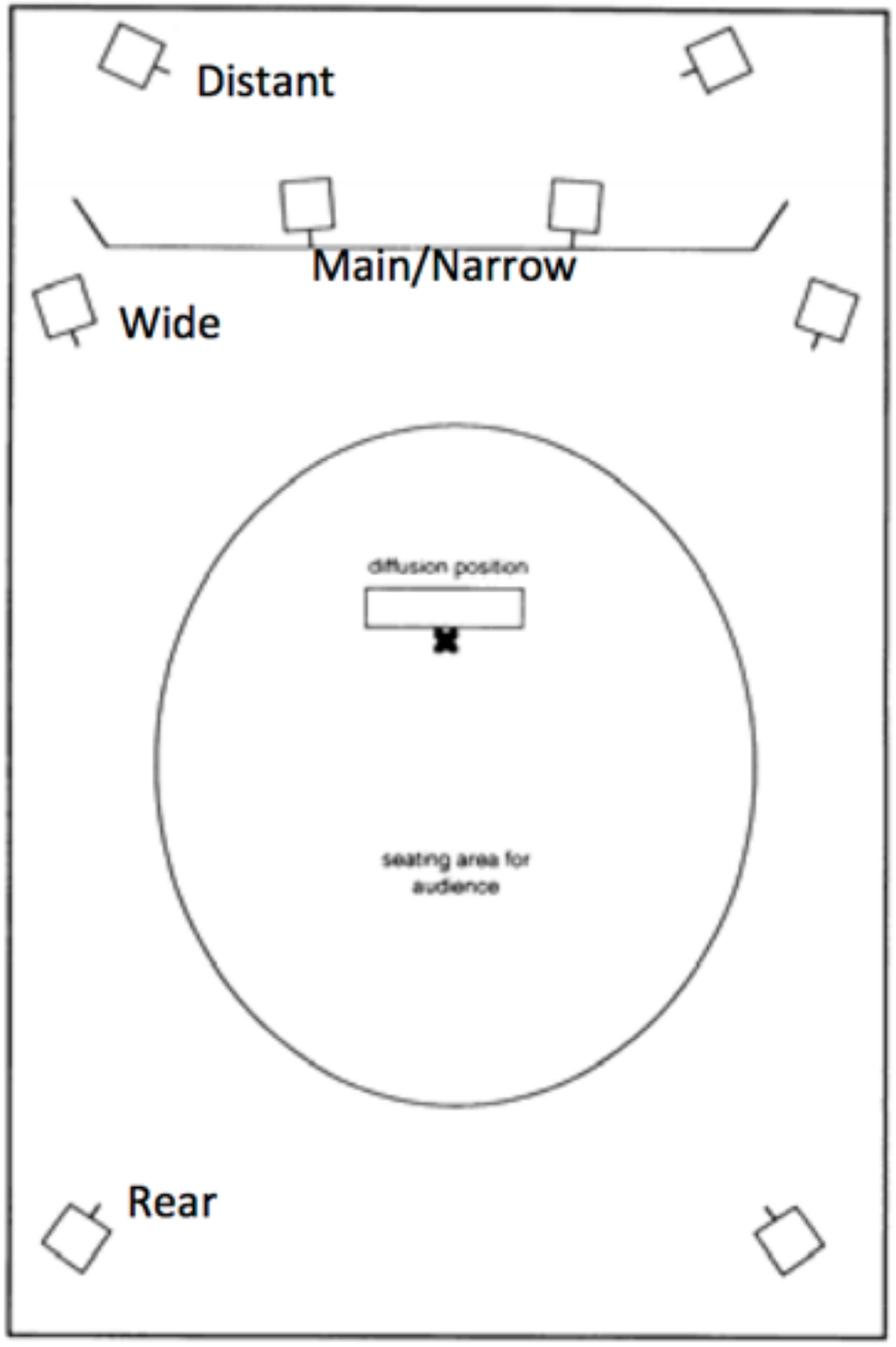
Space frames (Emerson, 2007)

SPATIAL AUDIO TECHNIQUES AND PHILOSOPHIES

- Two broad categories
- (a) Diffusion/'loudspeaker orchestra', focus on space-frame delineation and basic in/out, diffuse/point-source movements and dynamics (this approach is generally more ad-hoc, practice-led, based on experimentation)
- (b) Attempt to accurately recreate spatial audio cues and soundfield, generally via (relatively large number) of symetrically-placed speakers

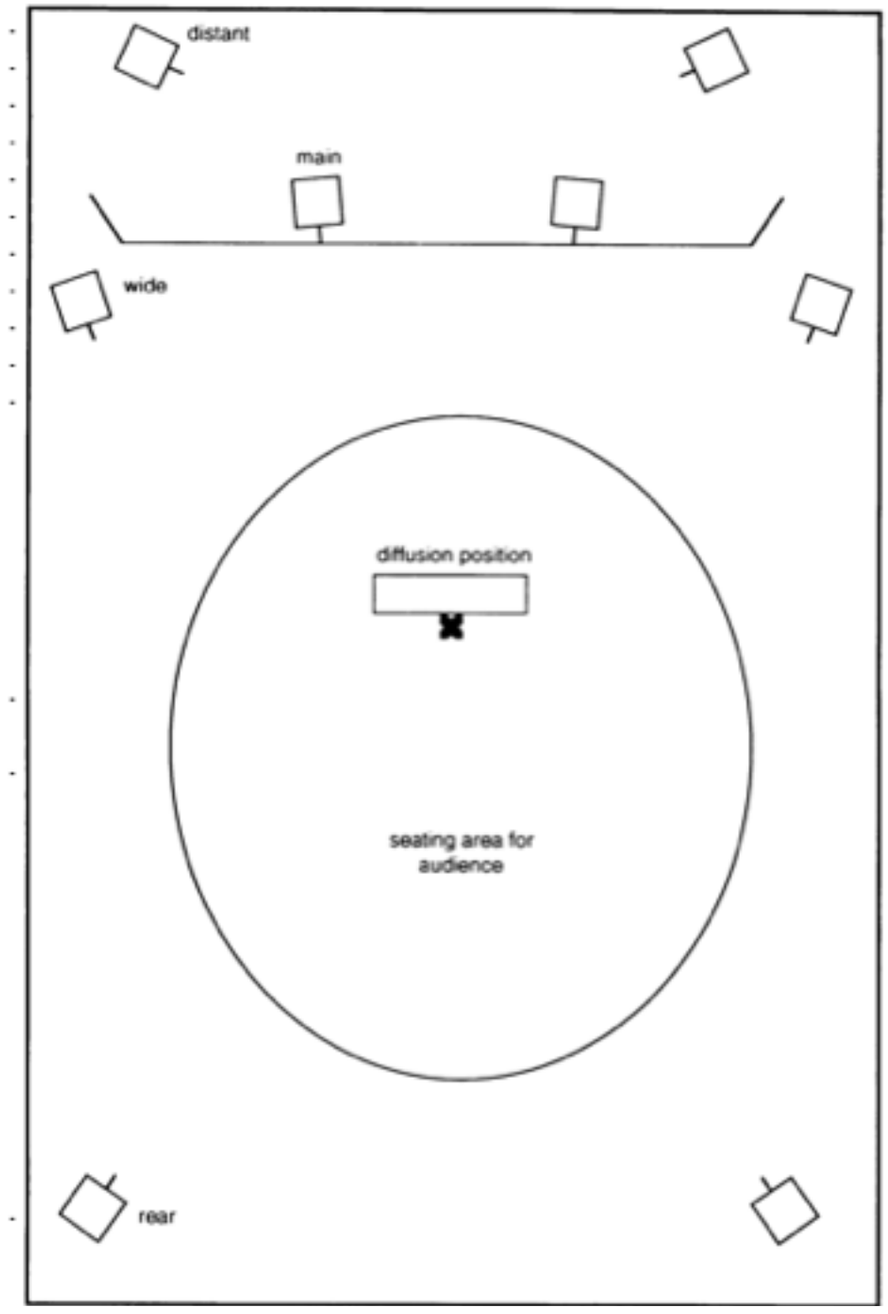
(A) DIFFUSION/LOUDSPEAKER ORCHESTRA

- Diffusion setup (8-channel): the BEAST setup (after Birmingham Electro–Acoustic Sound Theatre); see Harrison (1998)
- Different stereo pairs serving different listener positions (and allowing for local/field, front/back or diffusion articulations, in addition to 'standard' stereo perspective)
- Often fed from stereo track to mixing console with 8 output busses (live performance!)
- Can create dynamic and performative impact
- Works well for material with significant transient detail (perhaps combined with more slowly-articulated envelopes)

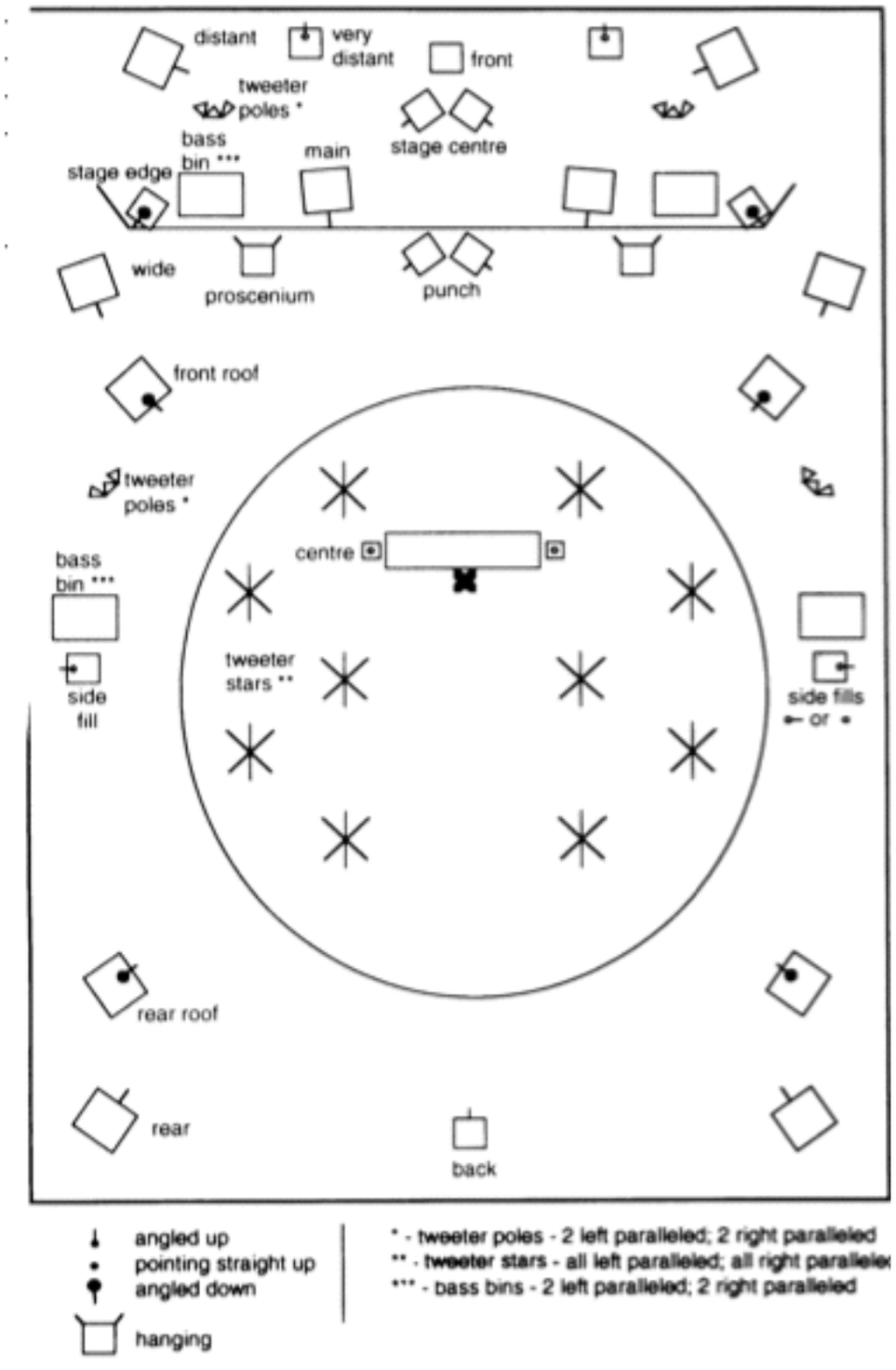


HARRISON'S DIFFUSION AND THE BEAST

- Jonty Harrison is a composer who has worked extensively with diffusion
- Background: **limited 'sweet spot' for stereo** necessitates multiple speakers in concert hall to **fill in potential 'gaps' in stereo field at different locations** for a distributed audience
- Can feed different amounts to different speakers to increase dynamic range and draw attention to different parts of the performance/diffusion space
- Creates a performance out of a fixed media piece: Harrison (1998) 'the composer proceeds by drawing out implicit larger structures from the explicit morphologies of individual sound objects'
- Harrison/Birmingham initial BEAST setup: multiple stereo pairs; (1) standard, (2) wide, (3) distant, (4) rear)––latter apparently provides for 'anchoring' of stereo image



EXTENDED BEAST



Multiple diffusing speakers tailored to concert space

Open to criticism - lots of potential for **out-of-phase sound** (same sound materials from multiple locations may create confused spatial imaging)

However, it does attempt to account for needs of distributed audience and tailoring to performance space (pragmatic approach) and is also quite dynamic/dramatic

BEAST SETUP IN ACTION

https://www.youtube.com/watch?v=jM65sUdaS_4

(B) RECREATING SPATIAL AUDIO CUES OR SOUND FIELDS VIA SYMMETRICAL SPEAKER ARRAYS

- We are familiar with one attempt to recreate spatial audio cues via the use of stereo panning based on equidistant stereo pairs using amplitude panning for simulating level difference cues

- A similar approach can be extended to spatial standards and approaches such as 5.1, 7.1, etc.

- 5.1 and 7.1 aren't symmetrical (and, as we've seen with 5.1, the rear speakers are badly placed in terms of providing clear left/right imaging and central image at the rear)

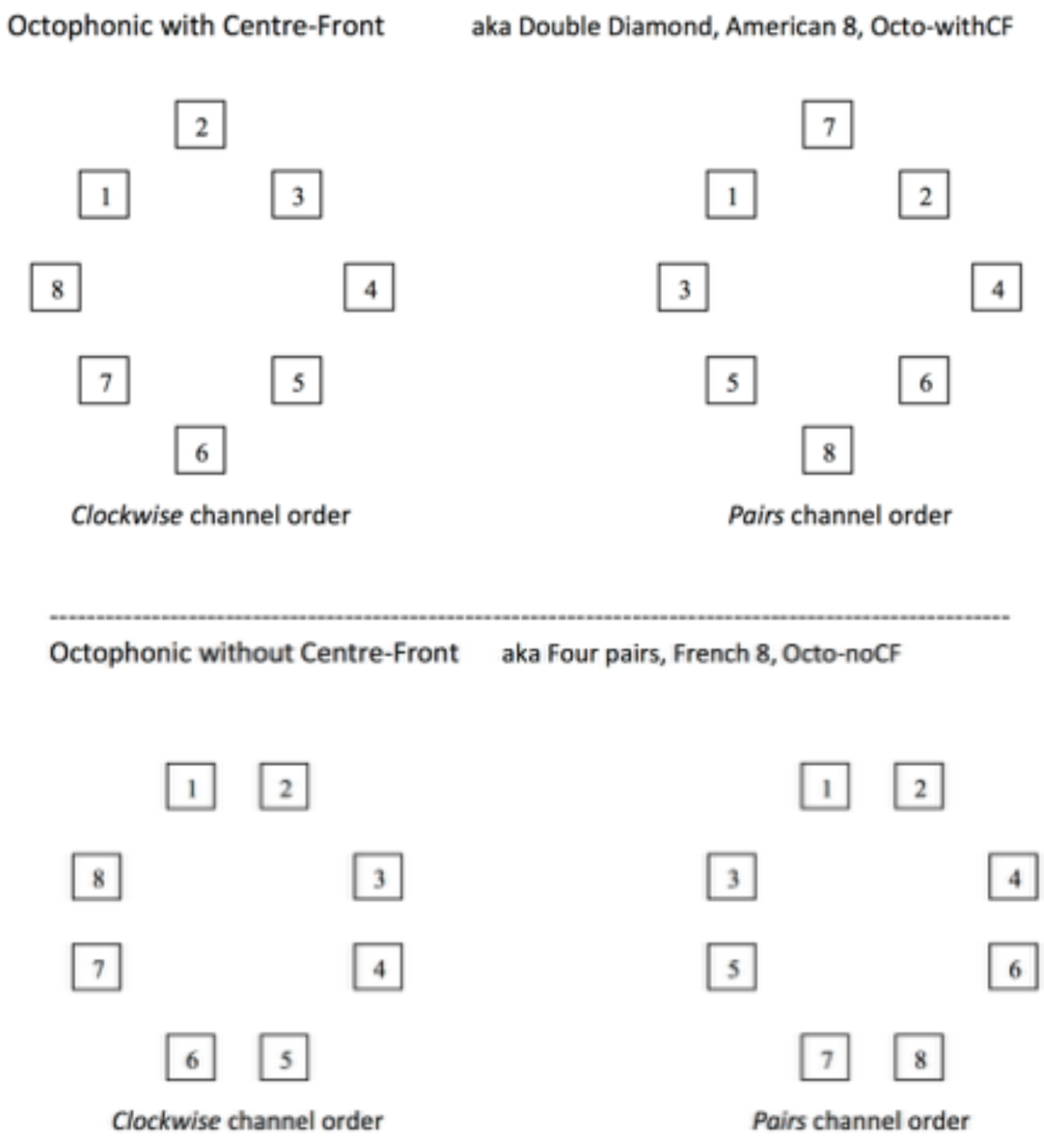
(B) RECREATING SOUND FIELDS VIA SYMMETRICAL SPEAKER ARRAYS

- We can try to use a fully symmetrical array (e.g. 8-channel/octophonic ring setups); simple approaches would have you place an audio channel at each speaker or use basic panning to move sounds

- However, there is still the problem that simple amplitude panning (i.e. simple crossfading between speakers) doesn't properly recreate a more complex spatial soundfield

- A moving sound source will seem to 'jump' between different speaker positions rather than fading smoothly, and the image changes dramatically with different listener positions (nearest speaker heard affects apparent direction of sound)

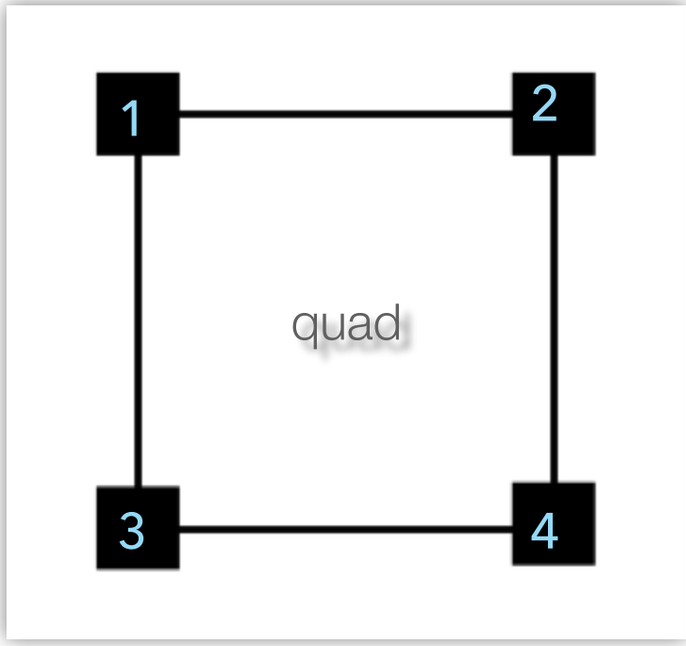
- Solution: make a number of speakers contribute to the creation of the sonic image (more realistic recreation of soundfield)….find ways to 'pan' which involve a number of speakers, not just one or two



8-channel variants; from Bates (n.d.)

AMBISONICS: AWAY FROM DISCRETE-CHANNEL SPATIAL AUDIO

- So far, we have viewed spatial audio as being based on **discrete channels**: i.e. you assign a track to a particular speaker output, or you use a surround panner to crossfade a track between two different speakers...this is an intuitively clear approach, but it is **not necessarily the most efficient or accurate approach to multichannel audio**



- A particularly **problematic case for discrete channel approaches is that of the early quadraphonic standard (1970S)**: 4-channel audio with symmetrical speaker placement of 2 stereo pairs (front and rear)

- This produces a **90 degree separation between each speaker: this causes the virtual sound images found in stereo pairs of speakers to 'break down'** (60 degrees is considered to be the maximum advisable separation), resulting in the sound being heard 'at the speaker/ channel' as opposed to creating a spatial image with apparent realism between speaker positions

*In general, the **exclusive use of level-difference cues in discrete channel approaches is particularly prone to localisation errors**: moving closer to one speaker will significantly distort the spatial 'image'; c.f. Haas effect/law of first arriving wavefront*

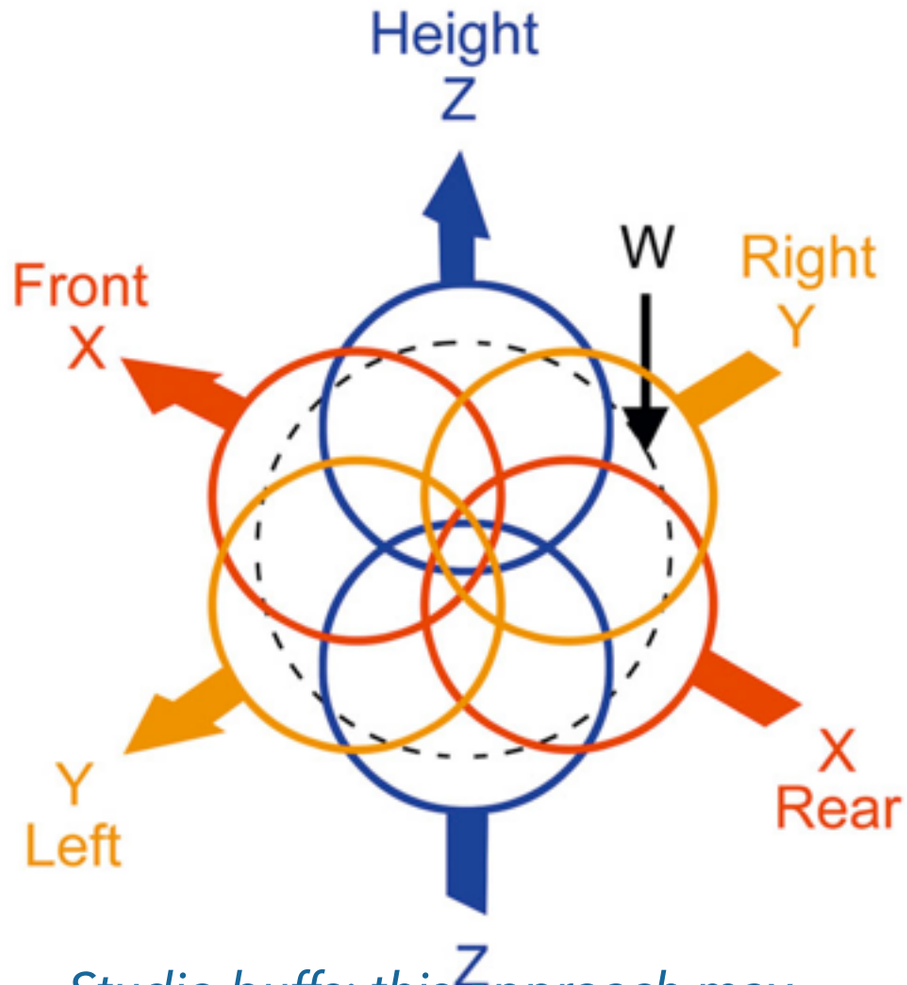
AMBISONICS: AWAY FROM DISCRETE-CHANNEL SPATIAL AUDIO

- Therefore, for more accurate rendering/reconstruction of spatial sound, discrete channel approaches have significant limitations (particularly if a limited number of speakers are used)

- Michael Gerzon discovered an **alternative approach-- ambisonics-- based on trying to create a more accurate reconstruction/rendering of a spatial sound field** which is not subject to distortion based on speaker/listener position

- This approach is based on **using each speaker to contribute to the rendering of the sound field, rather than using a speaker only for a source at (or very near) its position**

- It does this by **splitting the sound into different components for overall level (marked W) and directional elements for horizontal and vertical directions (X, Y, Z)**. A given speaker will contribute a greater or lesser degree of the level info and directional components, depending on the location which is to be reproduced. [This 4-component signal is known as B-format.]



Studio buffs: this approach may remind you of mid-side stereo recording: mid (omnidirectional mic) provides level, figure of 8 mic provides directional info)

AMBISONICS IN PRACTICE

- **In practice, ambisonics provides us with a means of encoding spatial information with a high degree of accuracy** (including vertical location), which can then be decoded using a wide range of speaker setups
- **Mono signals/channels can be accurately and smoothly panned to a variety of apparently 'between-speaker' positions** by an application which calculates W, X, Y (and, if vertical aspect, Z) components for its intended location (and then decodes it to the speaker setup, with varying W, X,Y, Z components for each speaker)
- It is less susceptible to distortion of auditory perspective because a number of channels are contributing to the resulting sound, providing a more realistic approximation of an actual spatial soundfield
- **Very convincing spatial movements can be generated using this technique**: if you want a virtual source to smoothly circle around a listener's head, this is the approach for you
- **The good news:** an ambisonics application will do the encoding and decoding for you

ALTERNATIVE APPROACH: VBAP: VECTOR BASE AMPLITUDE PANNING

- (Discussion from introduction in *Ardour* manual)

- VBAP was developed by Ville Pulkki at Aalto University, Helsinki, from 1997 onwards. It works by distributing the signal to the speakers nearest to the desired direction with appropriate weightings, aiming to create a maximally sharp phantom source by using as few speakers as possible:

- A one speaker, if the desired direction coincides with a speaker location,

- B two speakers, if the desired direction is on the line between two speakers,

- C … and three speakers in the general 3D case (the 3D case allows for height

- Its strenght is in adaptibility: it easily accounts for irregular speaker placement via distance factors

https://www.youtube.com/watch?v=nCl0md3gSMI

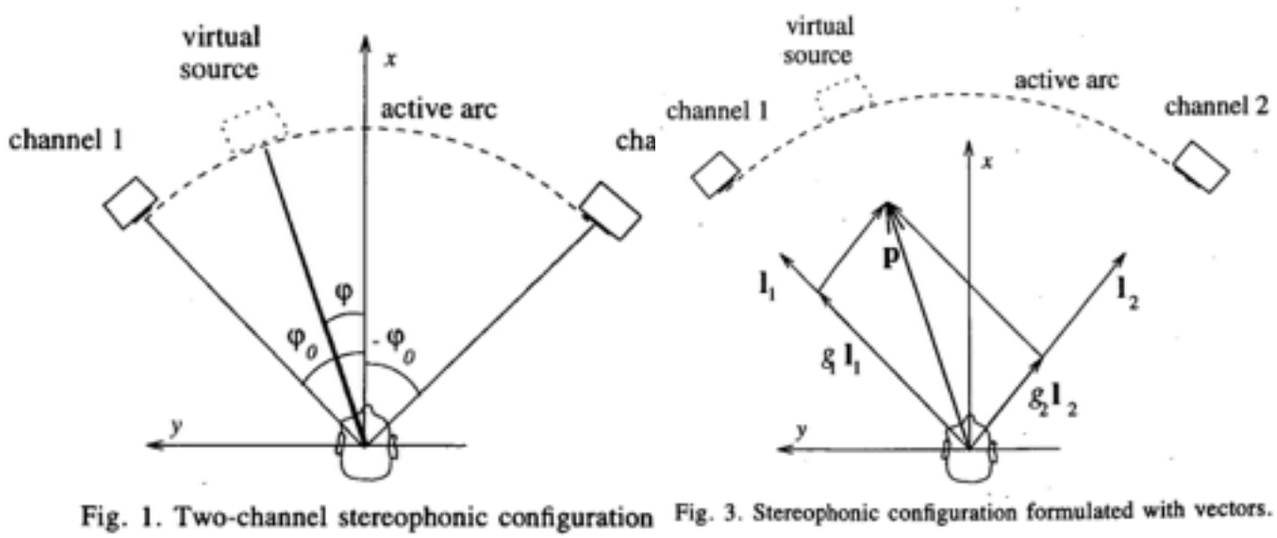
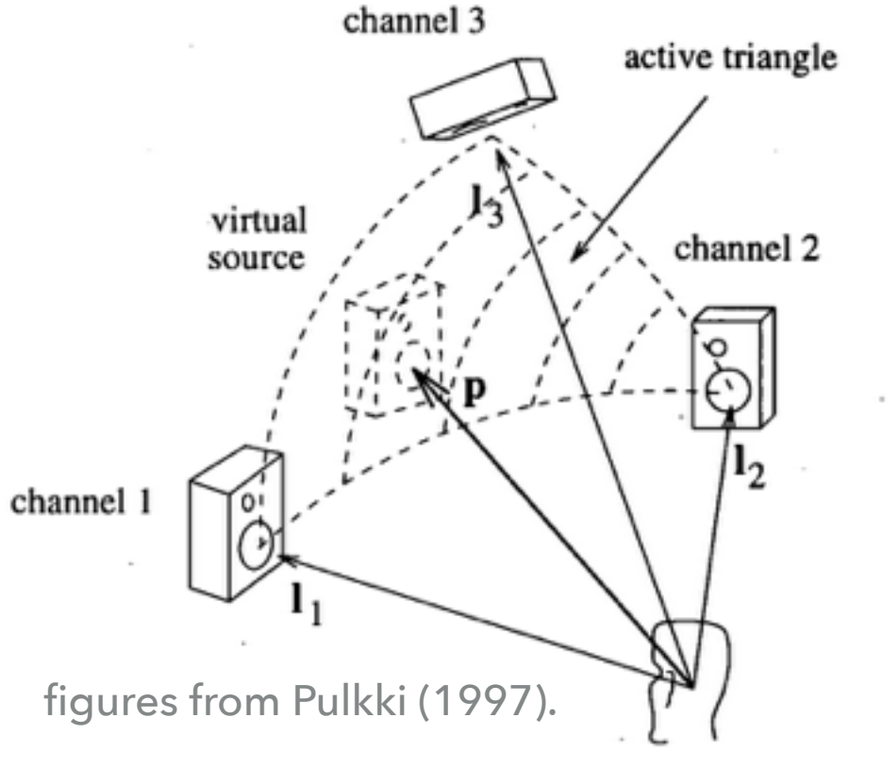


Fig. 1. Two-channel stereophonic configuration Fig. 3. Stereophonic configuration formulated with vectors.



figures from Pulkki (1997).

Fig. 5. Sample configuration for three-dimensional amplitude panning. Loudspeakers form a triangle into which the virtual source can be placed.

ALTERNATIVE APPROACH: VBAP: VECTOR BASE AMPLITUDE PANNING

- (Discussion from introduction in *Ardour* manual)
- Thus, if you move the panner onto a speaker, you can be sure that only this speaker will get any signal. This is handy when you need precise 1:1 routing.
- The drawback of VBAP is that a moving source will constantly change its apparent sharpness, as it transitions between the (two or) three states mentioned above.
- More elaborate implementations of VBAP also include a spread parameter, which will distribute the signal over a greater number of speakers in order to maintain constant (but no longer maximal) sharpness, regardless of position (advantage over some more limited amplitude surround panning approaches)

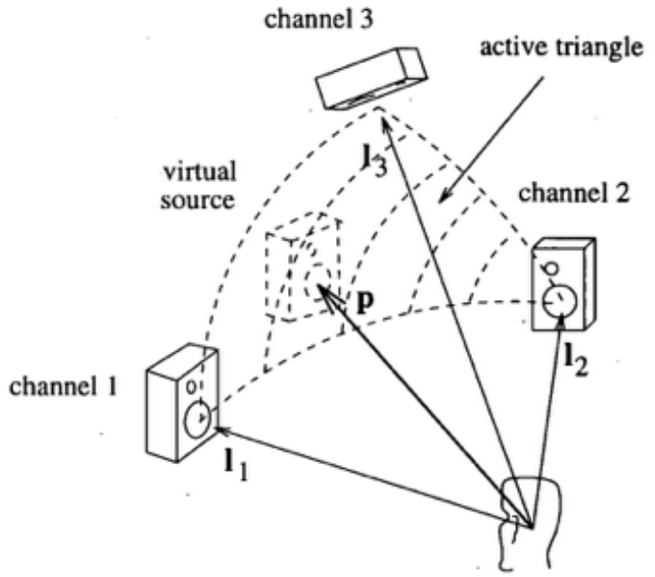


Fig. 5. Sample configuration for three-dimensional amplitude panning. Loudspeakers form a triangle into which the virtual source can be placed.

460

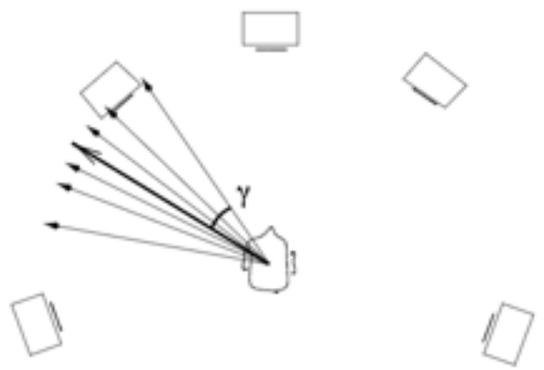


Figure 6: Panning directions used in 2-D spreading.

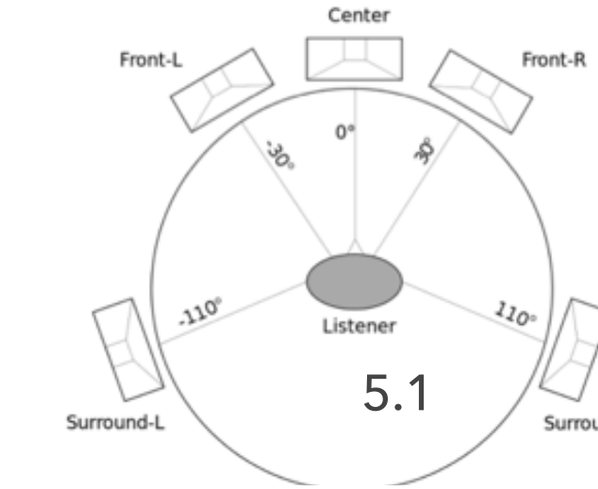
latter figure from Pulkki (1999)

Introduction to Studio 2

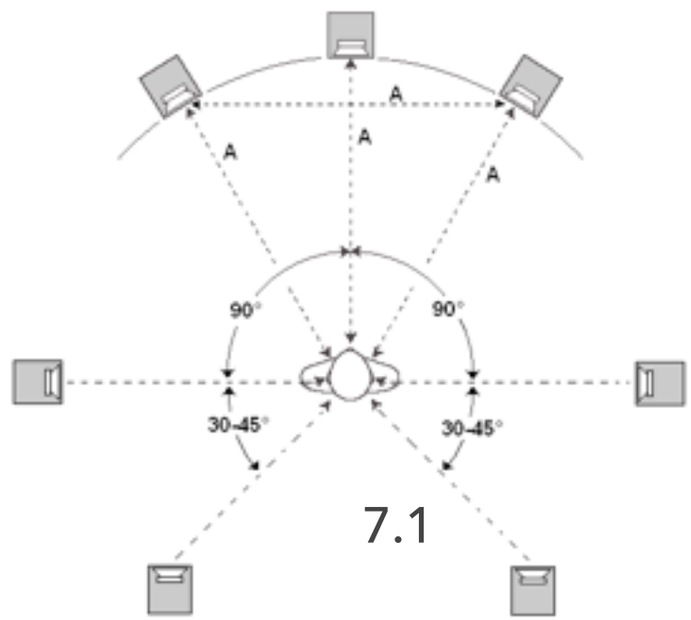
INTRO TO STUDIO 2

- Studio 2 is designed with the aim of making more accurate spatial audio presentations possible, whilst still offering possibilities of broad compatibility with commercial standards
- It prioritises symmetrical 8-speaker (octophonic) sound based on 45 degree separation between each speaker
- However, it can also broadly accomodate 5.1/7.1 setups due to the presence of a front-centre speaker
- The angles won't be quite right between this and the speaker layout in 5.1/7.1 (it's closer to 7.1, as you can see from the diagram to the right), but since 5.1 and 7.1 aren't particularly accurate with regard to speaker imaging, this is less significant than you might expect

Octophonic



5.1

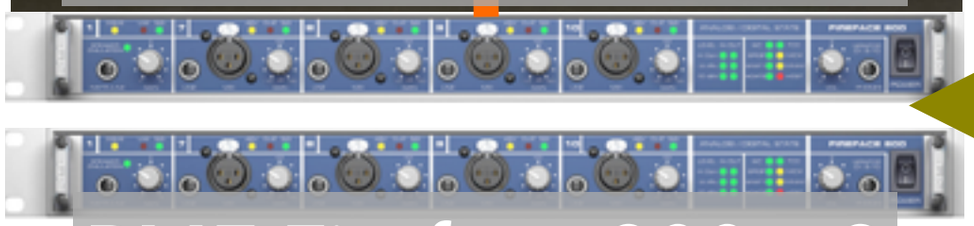


7.1

OUTLINE OF STUDIO 2 SETUP

KRK spatial speaker array

+ Genelec stereo pair



RME Fireface 800 x 2

FW: digital audio signal

MC Mix

Main Channel controls (8 faders)

control for DAW parameters (over ethernet)

SELECT FURTHER READING

Bates, E. n.d. Octophonic Array Configurations. [online]. Available at: http://www.endabates.net/Octophonic.pdf [last accessed 3/2016]

Bates, E. 2010. *The Composition and Performance of Spatial Music*. (PhD dissertation, Trinity College Dublin). Available at: http://www.endabates.net

Bregman, A.S. 1993. Auditory Scene Analysis: Hearing in Complex Environments. In: McAdams, S. and Bigand, E. *Thinking in* Sound. Oxford: Oxford University Press, pp.10–36. Available at: http://webpages.mcgill.ca/staff/Group2/abregm1/web/pdf/1993_Bregman_Hearing-complex-environments.pdf [last accessed 4/2014]

Emmerson, S. 2007. *Living Electronic Music*. Aldershot: Ashgate.

Harrison, J. 1998. Sound, space, sculpture: some thoughts on the 'what', 'how' and 'why' of sound diffusion. *Organised Sound*, 3(2), pp 117-127. [online journal available through library]

Roads, C. 2015. Composing Electronic Music; a new aesthetic. Oxford.

Smalley, D. 2007. Space-form and the acousmatic image. *Organised Sound*, 12(1), pp 35-58. [online journal available through library]

Also: Divergence Press, issue 3 had a special issue on spatial music (edited by Prof. Eric Lyon from Virginia Tech): http://divergencepress.com/Journal/JournalIssue/tabid/85/articleType/CategoryView/categoryId/3/Issue-3-December-2014.aspx

Also: http://www.soundonsound.com/sos/jan01/articles/surround.htm

Also: http://film-mixing.com/2015/08/22/understanding-stereo-and-surround-pan-laws-in-pro-tools-and-dolby-atmos/