Surround: The Current Technological Situation

David Griesinger
Lexicon
3 Oak Park
Bedford, MA 01730
www.world.std.com/~griesngr

There are many open questions

- 1. What is “surround sound”
- 2. Who will listen to it?
- 3. How will they listen?
  - 1. How do you play surround back?
- 4. How do you record it?

- My answers will be personal- (different)

Surround (Multichannel) Basics:

- Means “more than two loudspeakers”
- In practice there are:
  - quad (deceased) 2/2
    - two in front and two behind- in the corners
  - Dolby Surround + Pro Logic Surround - 3/1
    - three in front, one channel behind (maybe more speakers behind) - still very alive!
  - 3/2 Surround
    - three in the front, two behind - (quad with a center.)
  - 5/2 - Sony SDDS - five in front for theaters
  - 3/3 - Dolby Digital EX - three in the rear
  - 3/4 - Lexicon Logic 7 - four in the rear

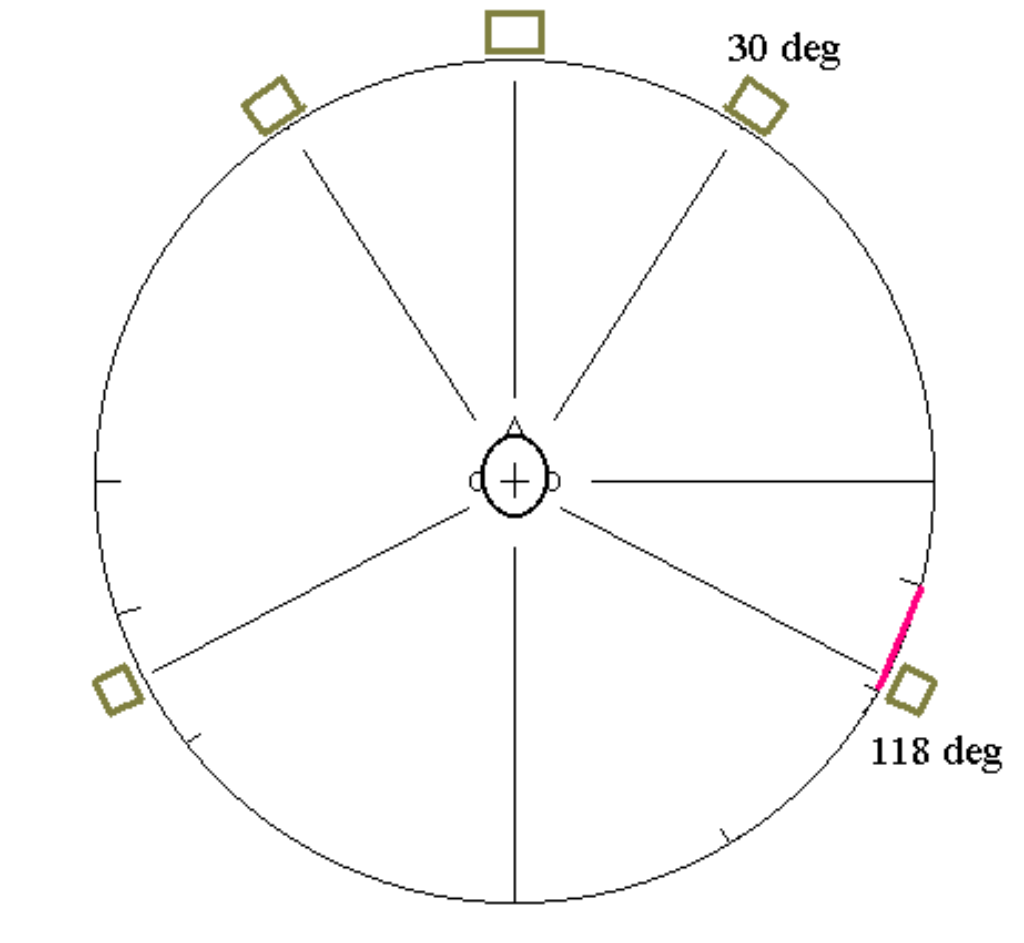
Discrete versus Matrix:

- Discrete systems record a separate channel
  - Most current systems use data compression
  - Dolby Digital AC3, DTS Digital, MPEG
- Matrix systems mix multichannels to two
  - and then attemp to unmix them on playback
  - no data compression is required
  - performance can be surprisingly good
    - but the most widespread decoders are usually OK for film and not so great for music

How do we listen to surround?

- In practice - in just about every way possible
- but, Naturally...

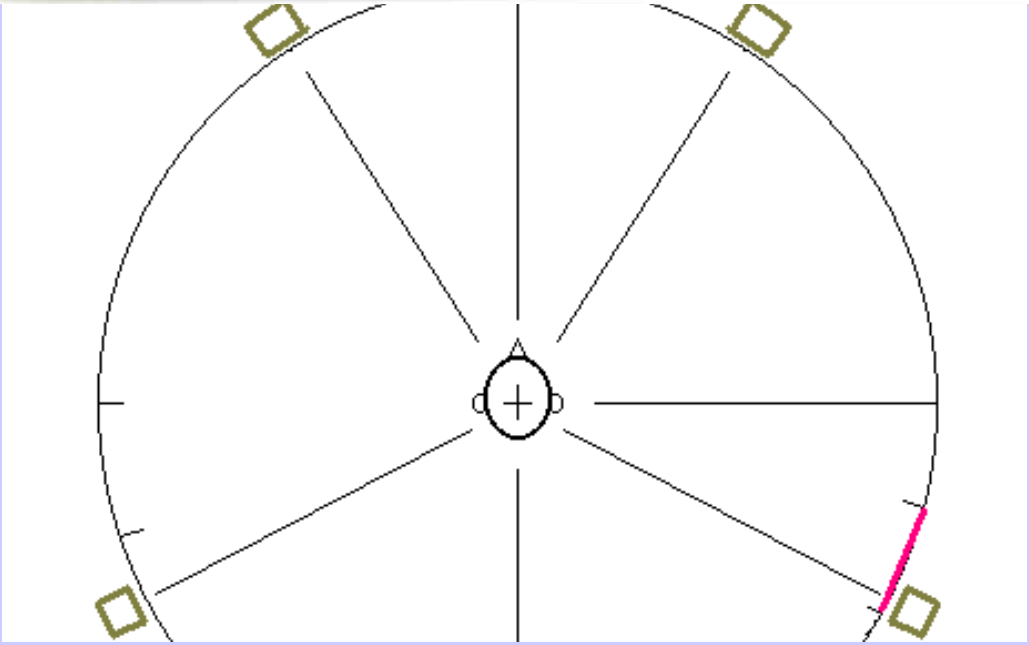
We have a Standard for 3/2!



- It is for a single listener at a fixed point

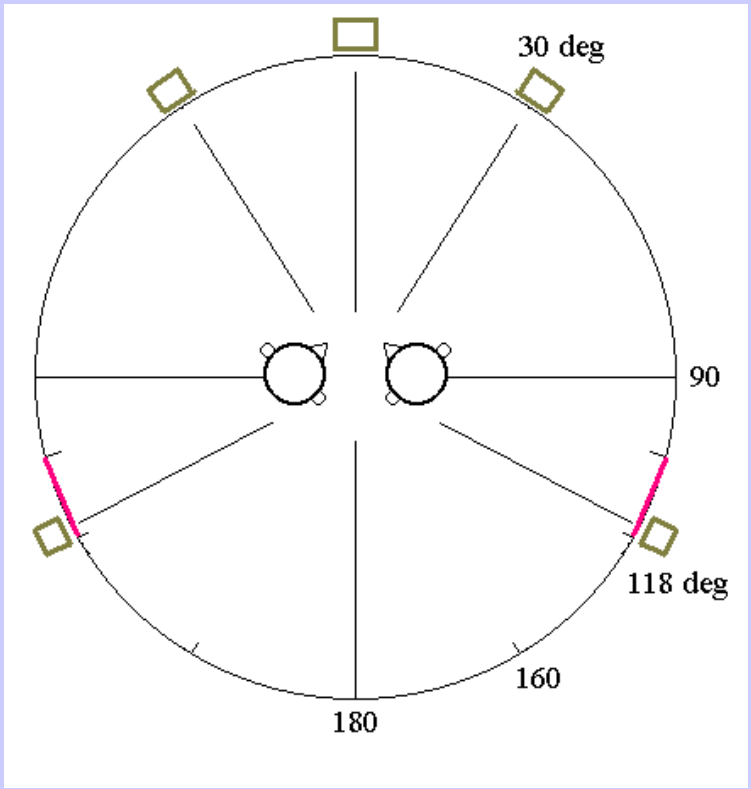
Music versus Cinema

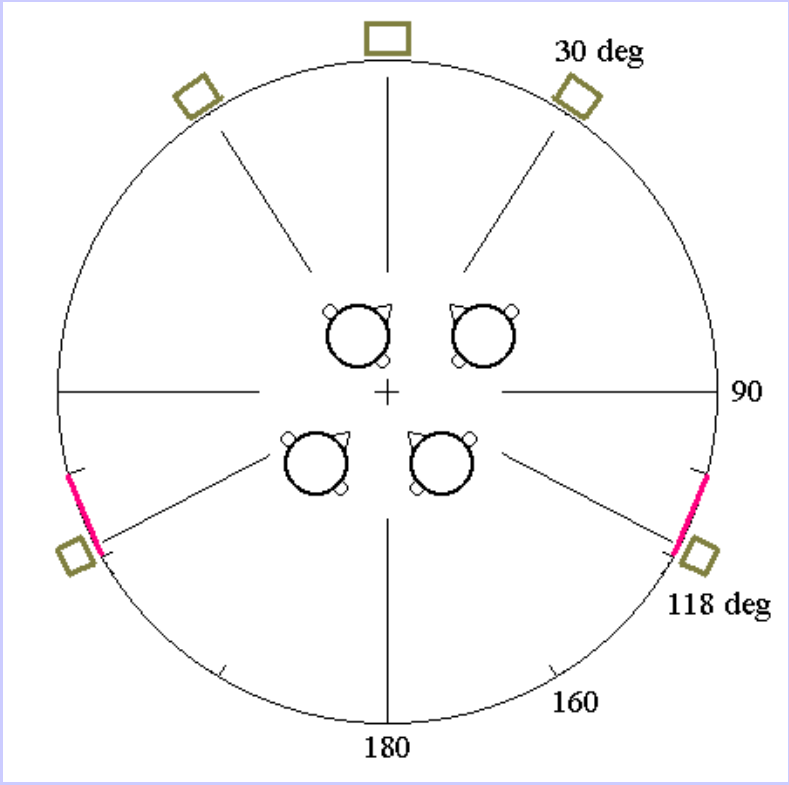
- Cinema is a group experience
  - we watch films as couples, families, or groups
- Recorded music listening is often solo
  - but how often are we in the sweet spot?
- The market has spoken!
  - Cinema wins
    - The great majority of surround systems purchased are for home cinema
    - The great majority of recorded material in surround is cinema
    - And the majority of surround music is a reproduction of a live concert.



- The standard does not allow friends to attend the concert.

Is the standard sensible?





- If we add other people,
- Will everyone be disappointed?

We need a large listening area, not a sweet spot

- Mono was democratic - it sounded equally bad over a large area.
- Two channel stereo sounded better than mono over a wide area.
  - but it localized well only on a line.
- 3/2 surround can sound better than stereo over a large area
  - but most people assume it localizes well at only one position.
  - this is NOT progress.

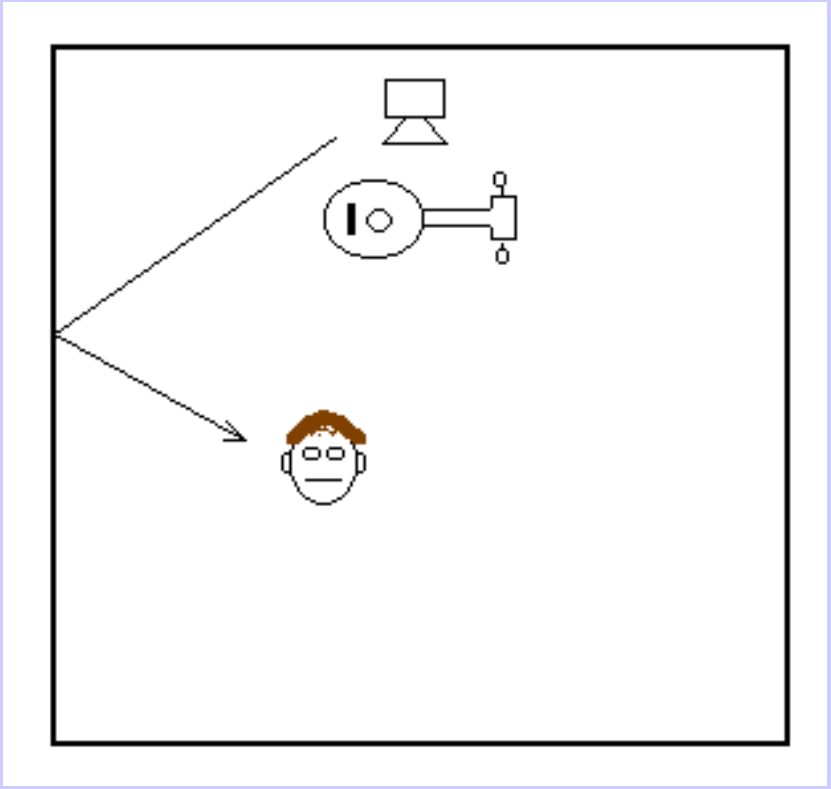
Our Goal:

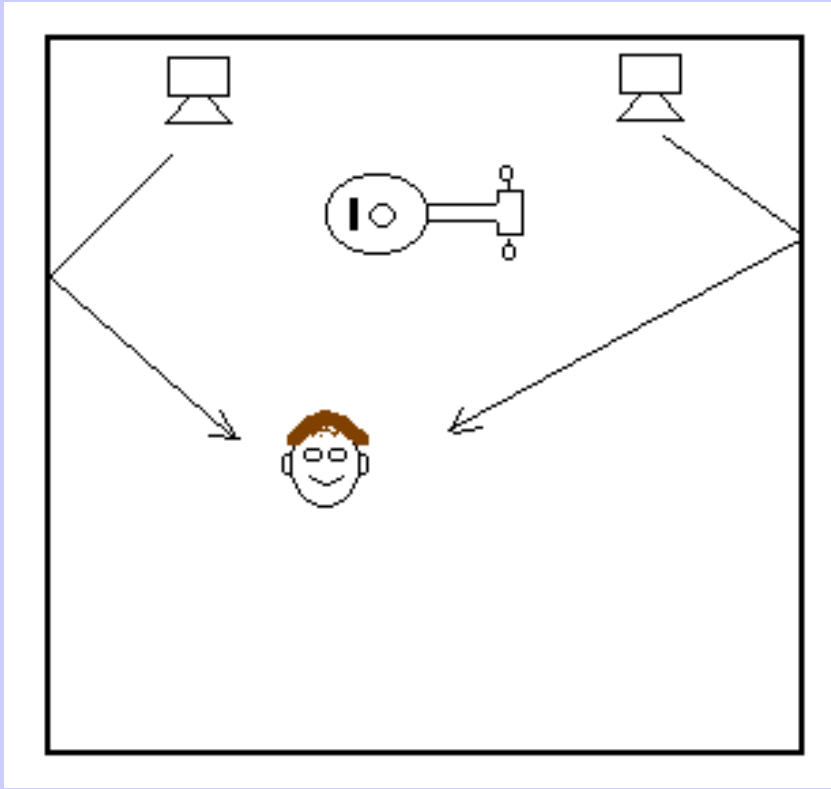
- Making a recording with good frontal localization over a large listening area
- Maximizing the sense of space throughout the entire room
  - If the recording sounds a little better in the sweet spot - that is OK.

Isn’t stereo enough?

- Mono was pretty good, and is still inadvertently common
- Two reasons for using two or more channels
  - 1. Horizontal localization (sound image)
  - 2. Enlarging the apparent size of the playback room
  - (Horizontal localization is probably over-rated.)

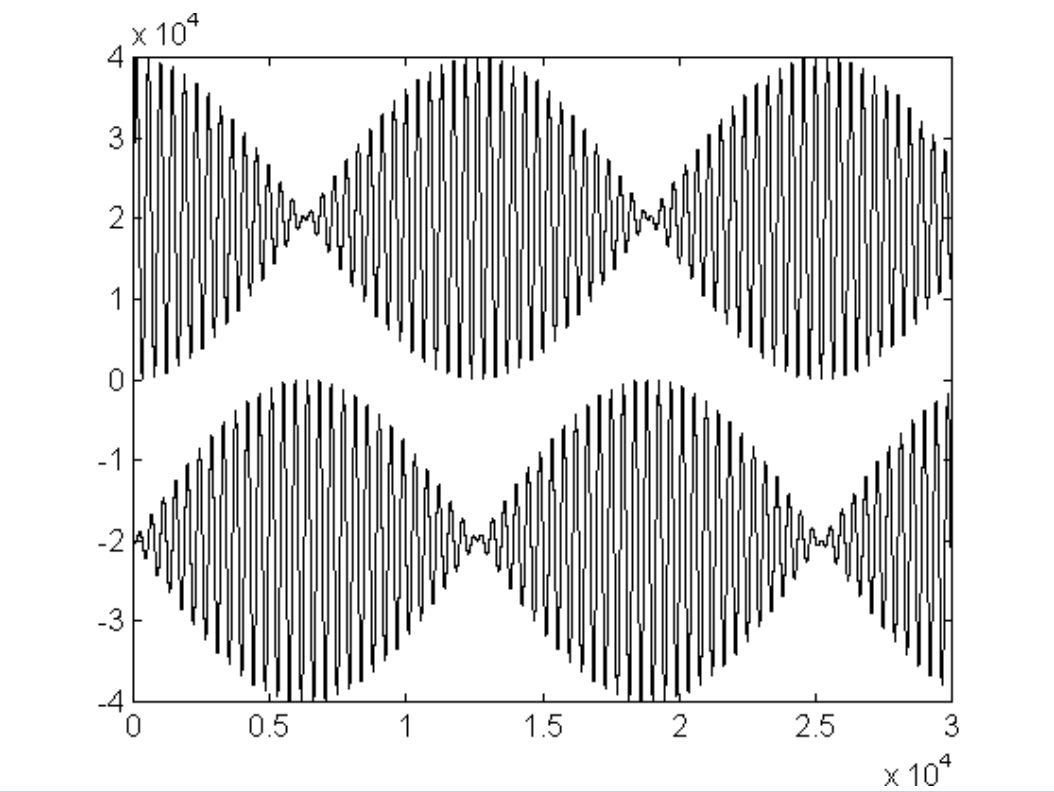
Mono vs Stereo





With a single loudspeaker there is no possibility of reproducing the spatial properties of the original space. With two, plus a recording that includes decorrelated reverberation, a few of these properties will come through. The playback room will sound larger everywhere.

Demonstration of Low Frequency Envelopment



– we can design a beat frequency signal

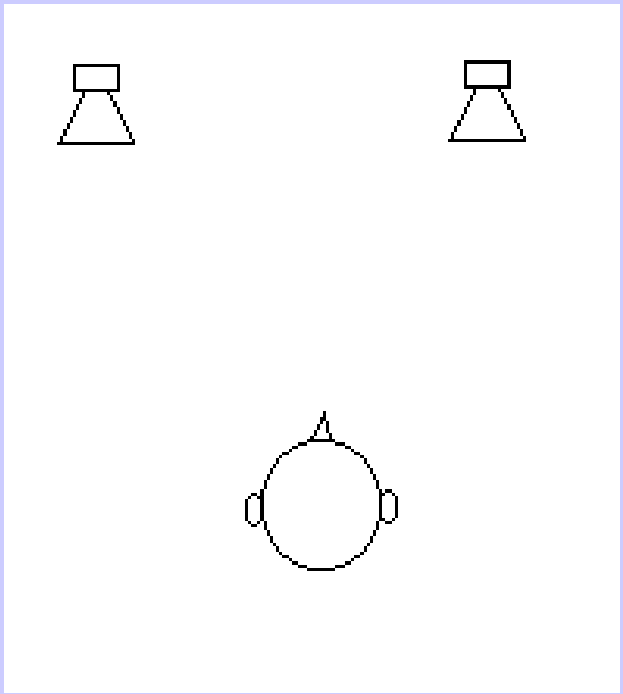
Demo Signal

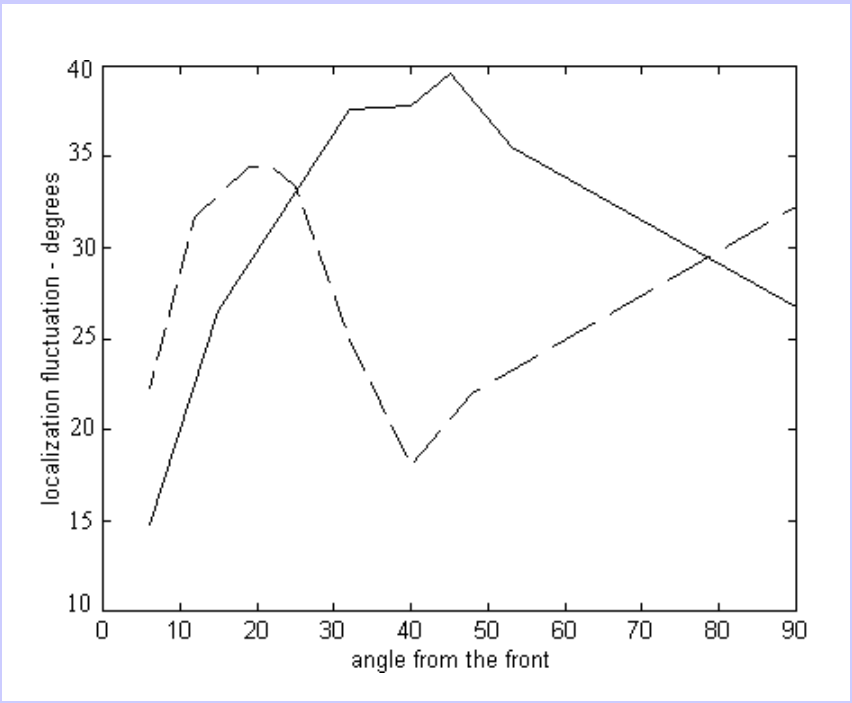
- Demo signal is made by taking the sum and the difference of two tones. One tone is about 3Hz higher than the other in frequency, so there is an obvious beat, 6 times a second.
- The tone pans from left to right and back at 6Hz.
- The phase alternates from in phase to out of phase at a 6Hz rate.

Comments on the demo

- When you play just one channel the sound gets louder and softer but does not move around the head.
- Play both channels and the sound moves around.
- In some places and for some frequencies only L+R is heard, and the sound is pure tone.
- In this lecture room the sound does not move around, but sounds outside the head and spacious. This is envelopment.

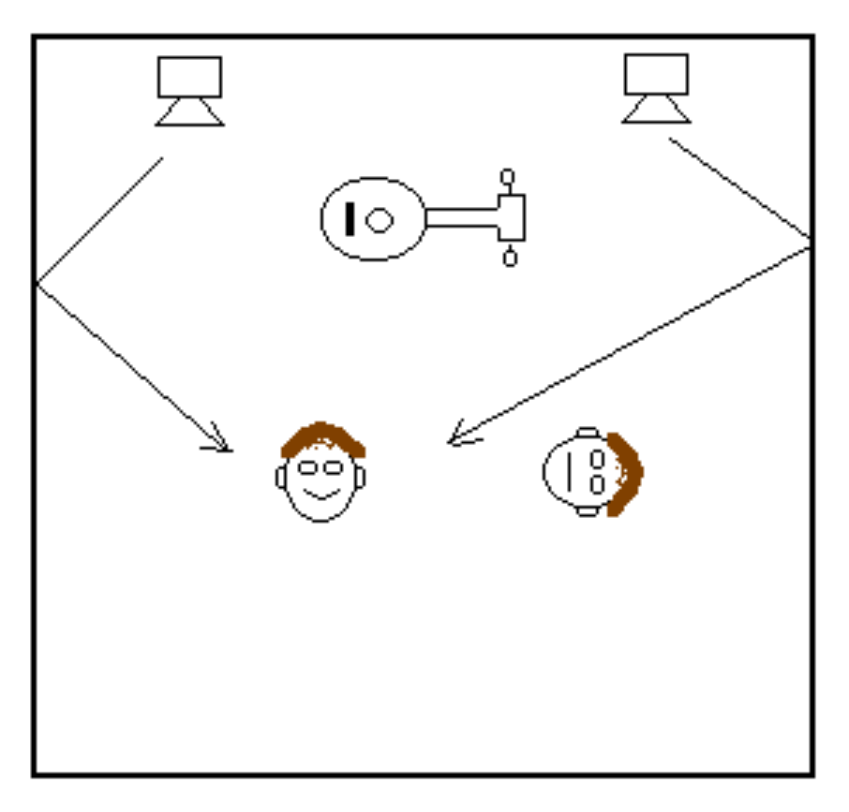
Envelopment at High Frequencies





Hearing envelopment depends on differences in the signals between the ears. The difference depends on frequency. At about 1500Hz the front speakers will sound as if they are at the sides if the signals are decorrelated. Thus standard stereo can be enveloping even without the room - but only above 1000Hz!

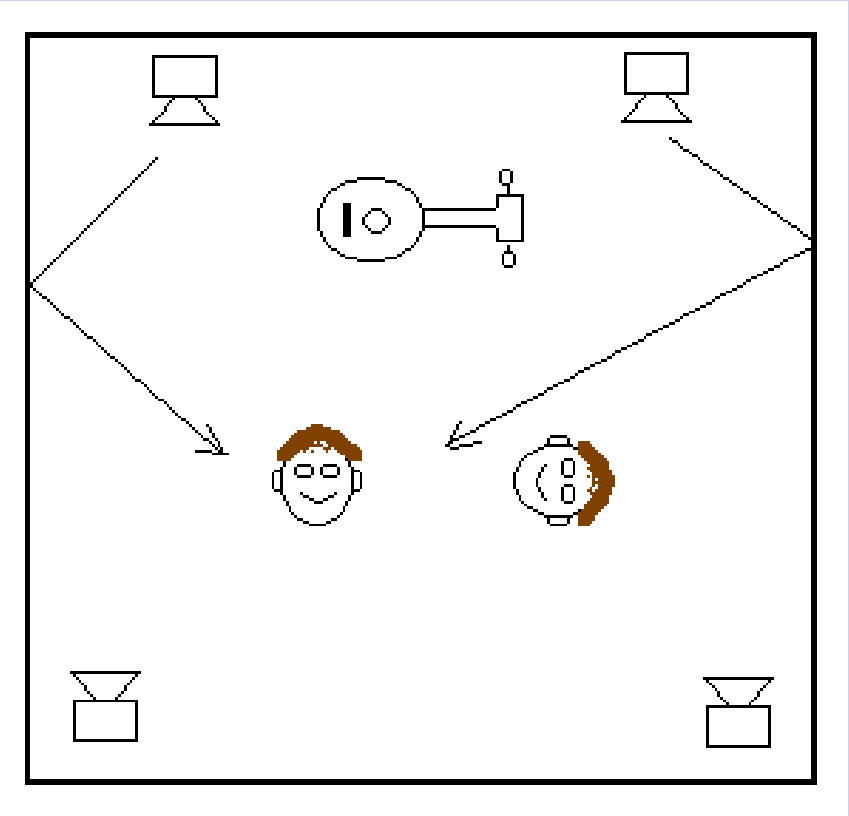
Why more than two channels?



Two independent drivers can recreate the spatial fluctuations of the original space - but only in one direction. A listener facing the side hears only the acoustics of the playback space.

Like it or not - we all turn our heads occasionally.

We need at least four drivers to recreate a natural space.



But these drivers must reproduce reverberation that is DECORRELATED!

(Decorrelation can be difficult to achieve.)

Experiments at Delft show at least 8 independent plane waves are needed to create a natural reverberant field.

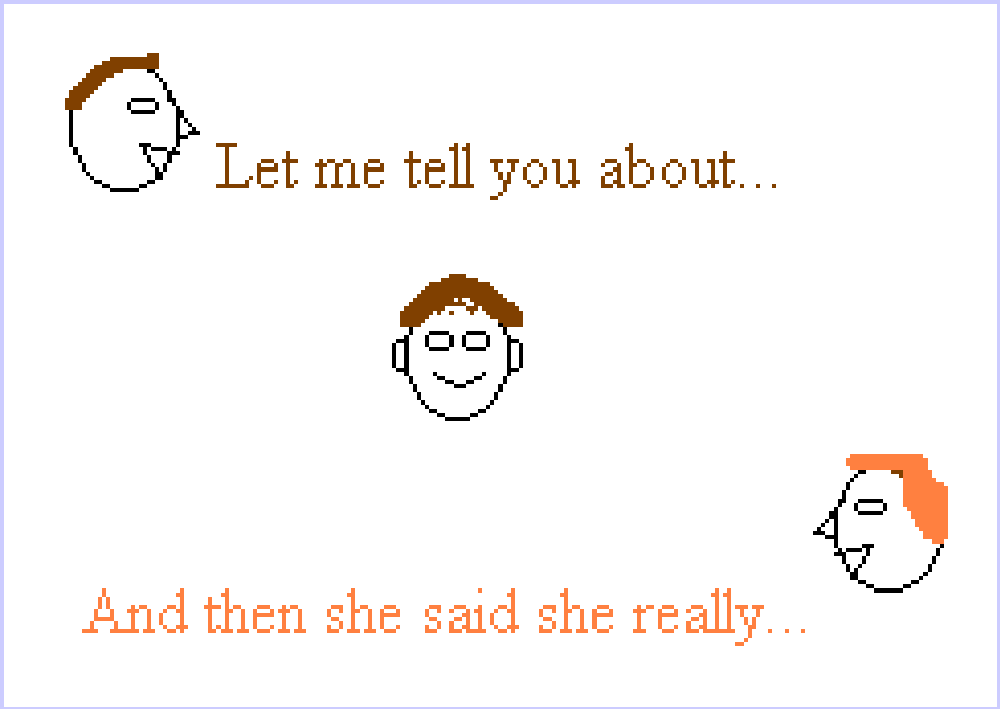
A driver in the center

- improves the listening area
  - we can pan sound sources between left and center
  - and between right and center
- But sources panned between front and rear are highly unstable.
- We can use the rear speakers for localization
  - but only if we avoid panning sounds to the side.

How do we hear sound images?

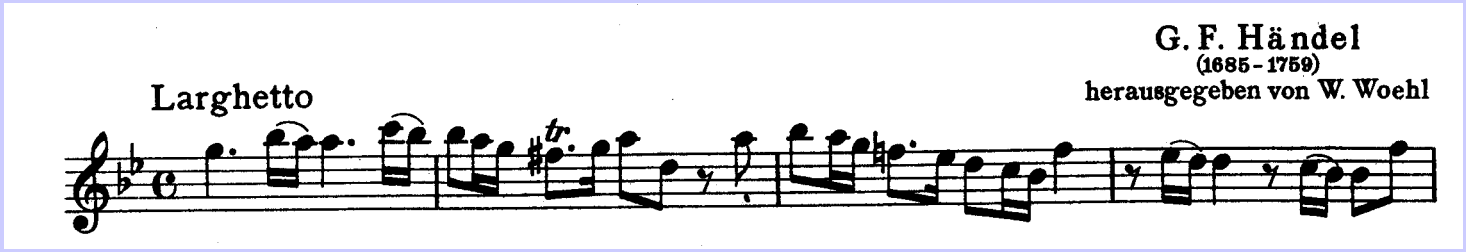
- Sound images are highly visual
- We hear sounds where we expect them to be.
- Blind Horizontal imaging depends on the difference of arrival time of signals with a rapid rise in amplitude.
  - We localize the beginnings of sounds
  - when sounds are continuous, localization disappears.

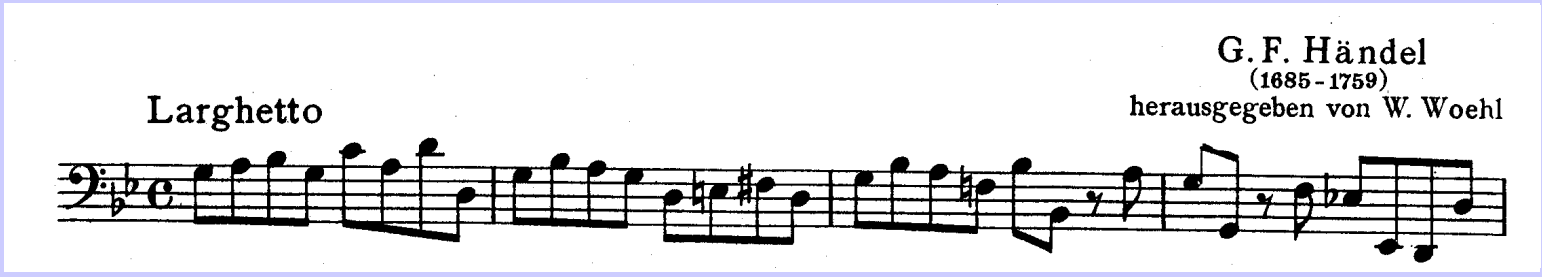
The cocktail party effect



We can separate the speech content from two different people, and choose which we will try to understand.

Sonic images and streaming:





We are capable of separating sound from two separate sources - even when they come from the same horizontal direction. These are examples of foreground sound streams.

Foreground separation requires sound events of finite duration

- We separate sounds by finding the start and ends of individual sound events,
  - and then we sort them by timbre and direction.
- Where two events overlap in frequency and time:
  - separation becomes impossible
- we determine sound direction during the start of events.

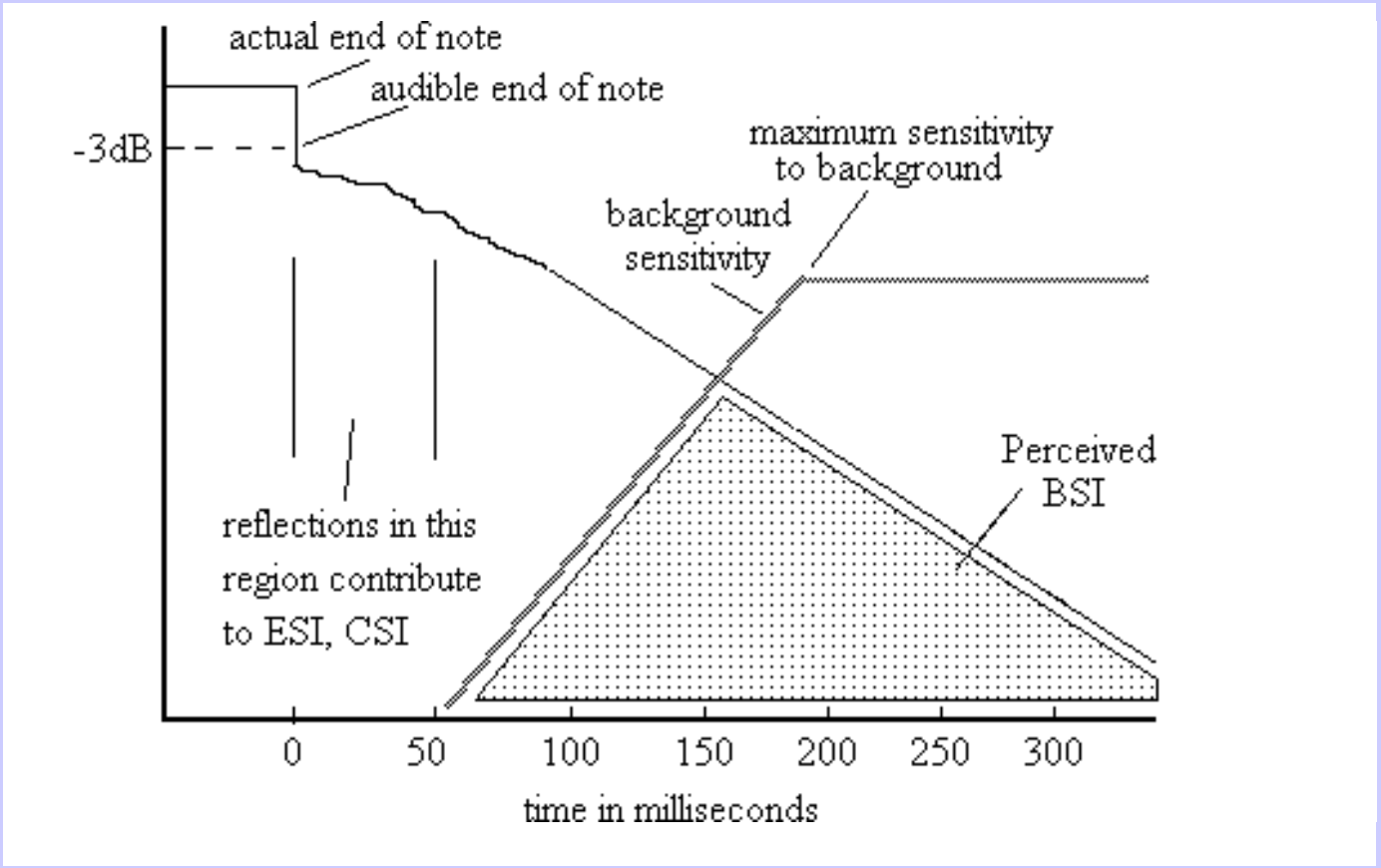
Background Streams

- There is neural circuitry for separating another type of stream.
  - Sound that does not belong to one of the foreground streams is assigned to the background stream
  - There is only ONE background stream
  - The background stream is perceived as continuous
    - even when it is composed of quickly decaying reverberation

The background requires a foreground

- The background stream can only exist where there is one or more foreground streams.
- The foreground stream(s) must be composed of sound events with clear starts.
- Ideally the foreground events also have a clear end
  - but this is not essential.

Separating the background takes time.



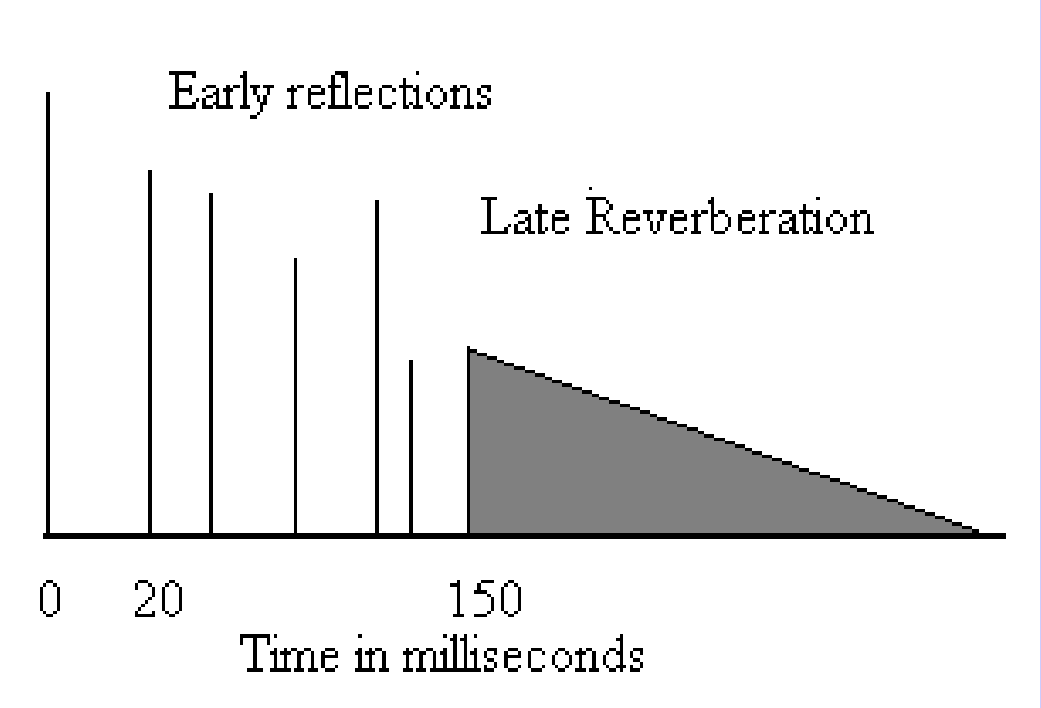
Haas Effect

- The Haas effect arises because the neural circuits wait 50ms after the apparent end of a sound event to be sure it is really over.
- Sound which arrives during this waiting period is assigned to the previous event
  - it is not heard as a separate event
  - although the spatial properties of this sound can be heard to some degree.

The perception of distance

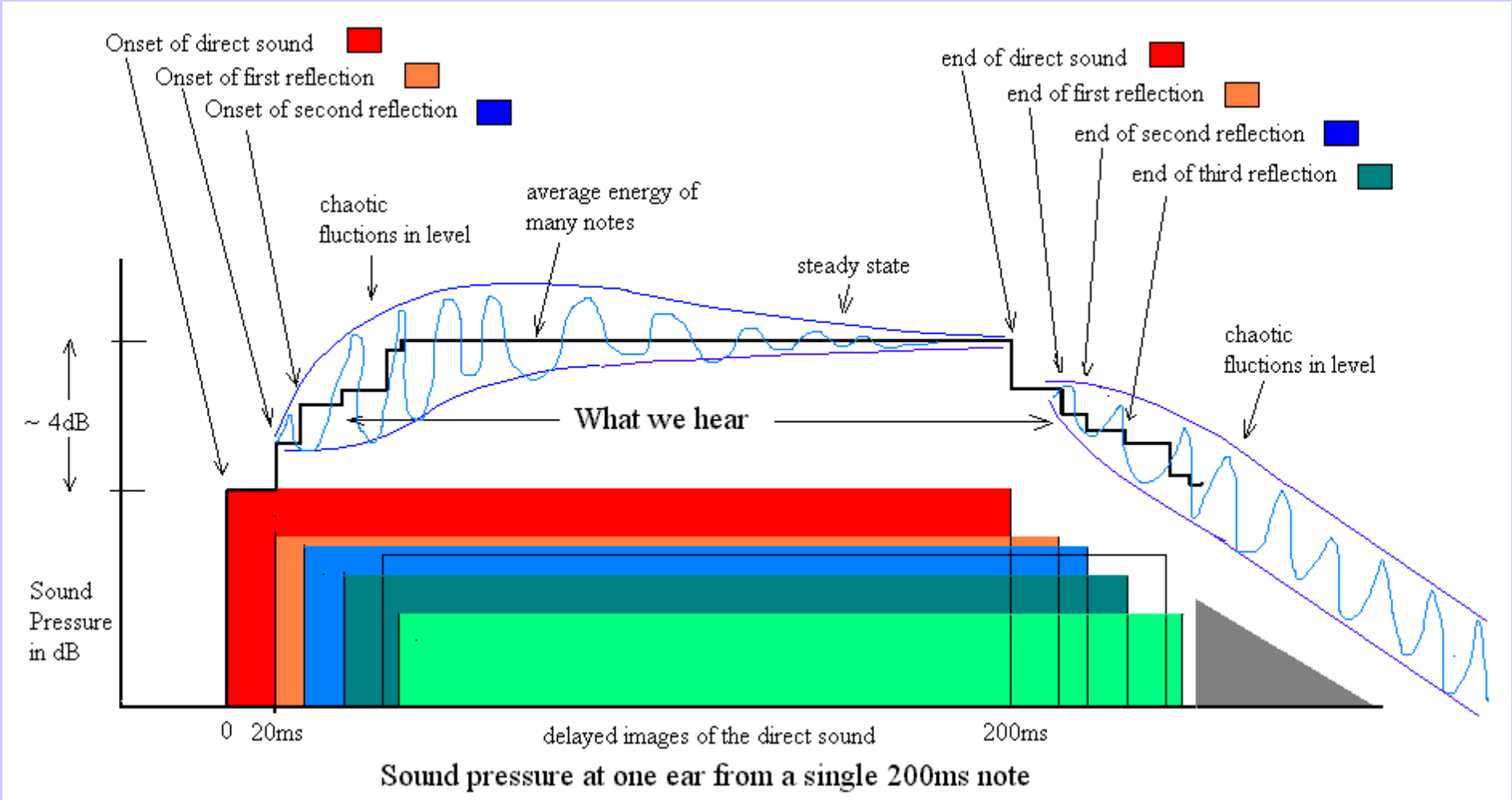
- Reflections (particularly lateral reflections) that arrive during the Haas period increase the apparent distance of the sound source
  - without changing the intelligibility
  - and without being audible as reflections.
  - You cannot tell the direction of these reflections,
  - but they sound most natural when they come from all around.

The “typical” impulse response



- is not what we hear
  - unless we like listening to gunshots
- Music is convolved with the IP

What we hear:



Sound events are convolved with the impulse response to produce the sound pressure we hear. It is the fluctuations in level caused by reflections that is audible, not the reflections themselves.

Spatial Fluctuations

- It is the difference in the fluctuations between the two ears that gives rise to the spatial properties of the sound.
- Fluctuation differences during the event - and up to 50ms after the event - give rise to a distance perception
  - and a perception that the sound is in an acoustic space of unidentifiable size and shape.
- Fluctuations occurring >150ms after the end of the event are heard as envelopment.
  - And as a clue to the size of the space.

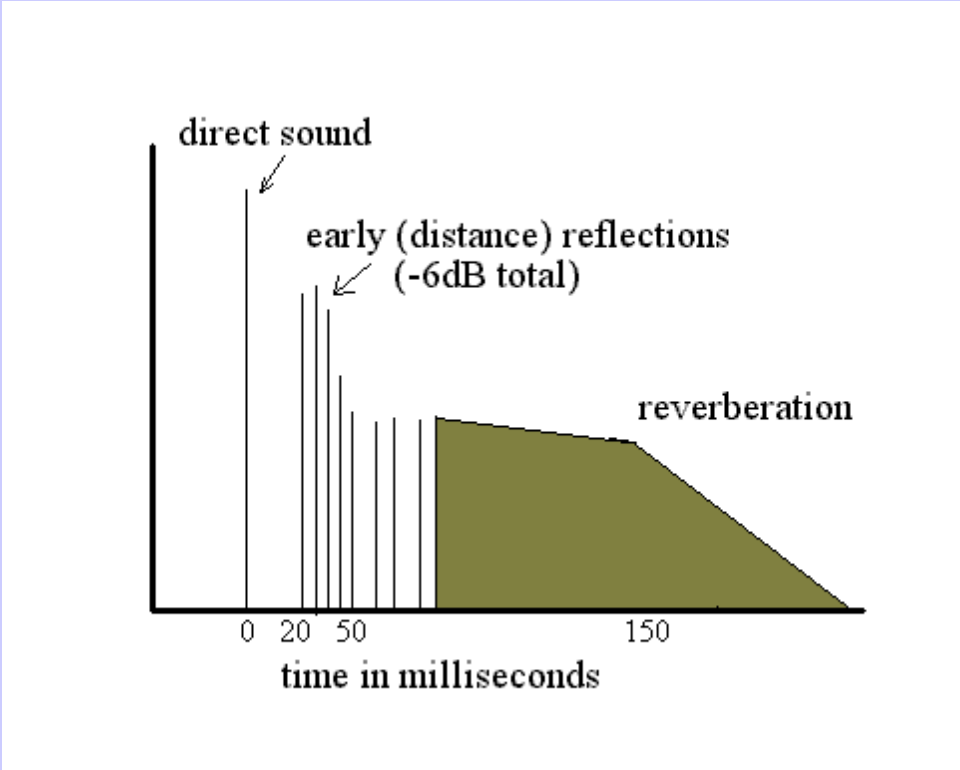
Reflections between 50ms and 150ms

- Reflections between 50ms and 150ms add a sense of distance, but at the cost of reduced intelligibility
- Reflections in this range sound “muddy”.
- The do not create envelopment.

Sound engineers need to control both perceptions separately!

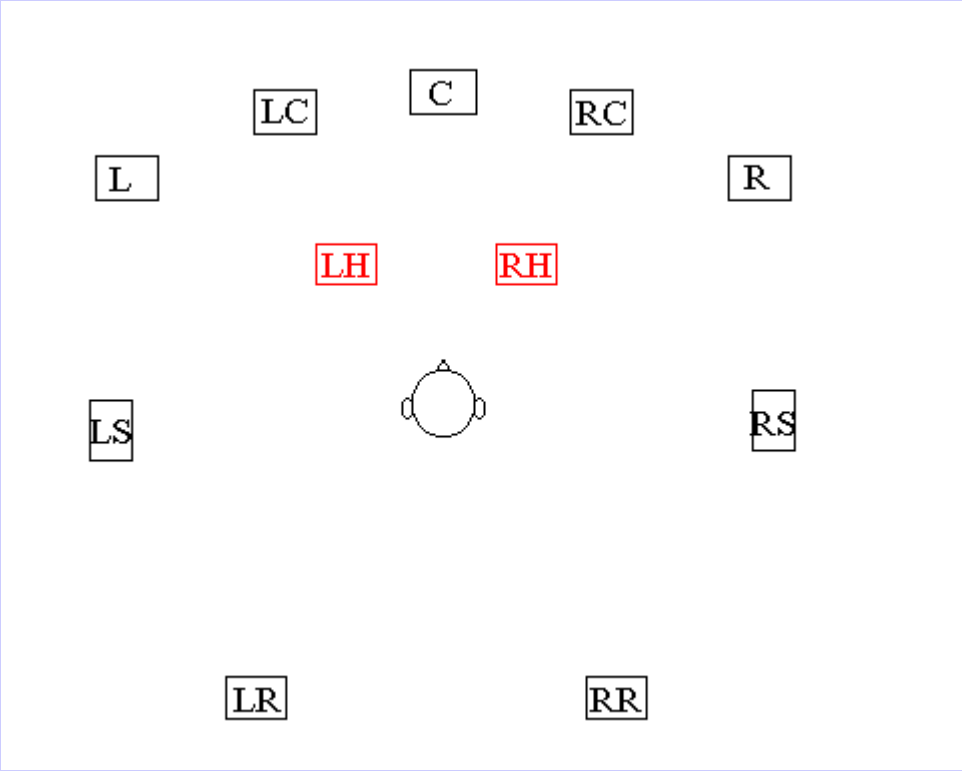
- A recording with to little early lateral reflections sounds too close and artificial
  - There is an optimum level for early reflections
    - -4 to -6dB total energy relative to the direct sound
- The level of energy >150ms is critical
  - There is a ~3dB change in audibility for a 1dB change in reverberant level
  - Audibility depends strongly on reverberation time.

The optimum reverberation



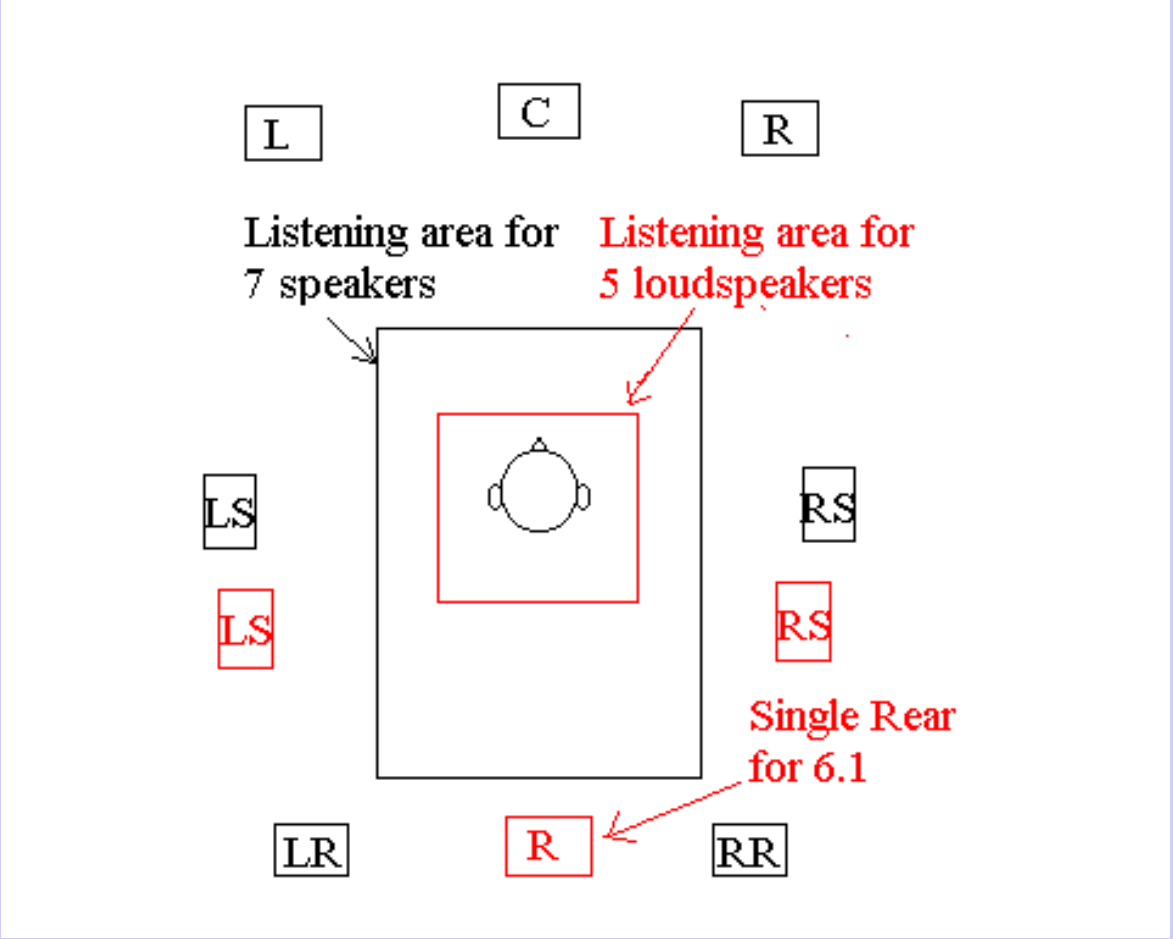
A strong early lateral field gives the sound distance and an acoustic integration. Reflections in the range of 50 to 150ms are minimized, while keeping reflections after 150ms strong.

How many speakers - and where should they be placed?



Ideally ~11 loudspeakers could be used - 5/4 with two overhead speakers

A Practical System: 3 front, 4 rear



A 3/4 surround layout has a significantly larger listening area than a 3/2 or a 3/3 layout. A 3/3 layout may reduce the listening area over a 3/2 system

How many channels to broadcast?

- Two channels is standard for broadcast
  - playback matrices exist to convert 2 channels to 4, 5, or 7 channels
    - performance is variable, but can be very good.
- 5.1 channels digital broadcasts are coming
  - matrices exist for converting the two rear channels to 3 or 4 rear channels
    - performance is variable - 4 channels sound better than 3.

Two channel matrices

- will be here here for a long time.
  - two channel surround films are (and music recordings) are being broadcast every day.
  - Dolby Pro-Logic (a 2 channel to 4 channel matrix) has a very large installed base.
  - Advanced 2-7 channel matrices will soon be common in automobiles.
  - The market for two channel matrix encoded material is still growing.

Bass Management : do we need another subwoofer?

- Current 5.1 receivers use a single subwoofer for all frequencies below 80Hz.
- It is easy to show that stereo drivers for the frequency range 40-80Hz sound better than a single driver or two drivers driven in mono.
- Below 40Hz a single driver seems to work well.
- So a 5.2 channel makes technical sense.

Matrix Comparisons

- not all two channel to seven channel matrices are created equal!
- They differ fundamentally in design
  - 1. They differ in the spatial width of the front image
  - 2. and they differ in the decorrelation of the rear channels
    - and in the ability to reproduce stereo in the rear
- Both differences strongly affect the perceived quality of the sound.

Adding a center channel reduces the width of the image!

- The center channel must be derived from a sum of the left and right channels.
- This always must reduce the separation to some degree
- The reduction in width can be controlled
  - by reducing the level in the center speaker when it is not required
  - but only one of the available matrices does this

The reverberation must be decorrelated

- reverberation in the four lateral channels should ideally be independent.
- Deriving four independent signals from two inputs is challenging.
  - At the very least, the independence should be maximized in the left/right direction
  - All the available matrices (except one) do not maintain high decorrelation in the rear channels, and only one can reproduce stereo in the rear.
- Caveat Emptor!!

Microphone Technique

- To maximize the listening area we must use the center channel.
  - This means for a centered sound source the center channel should be ~6dB louder than the left or the right.
  - This implies that NONE of the standard stereo microphone techinques will work!
- We must also avoid techniques that rely on time delay for horizontal panning
  - these techniques only work if the listener is centered.

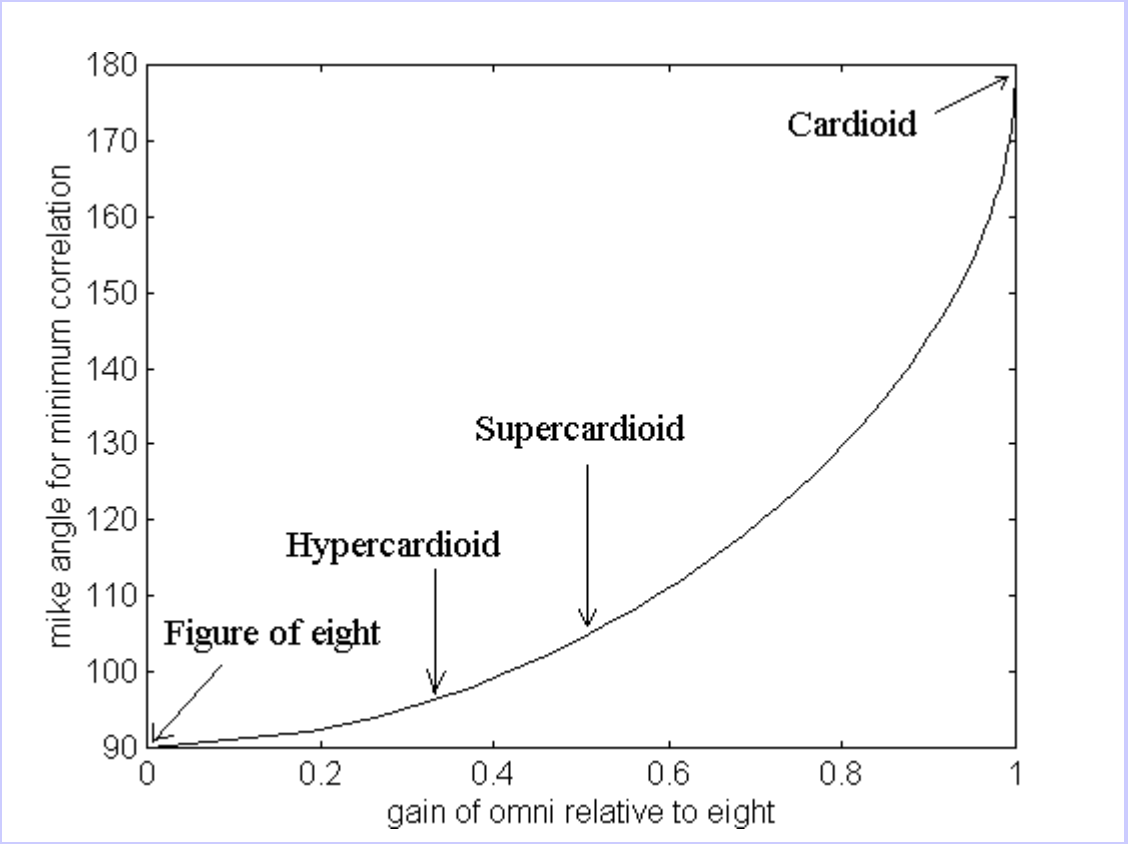
What technique is left?

- Fortunately, most commercial engineers are skilled in using multimicrophone technique.
- Perhaps in practice stereo technique is not too common outside of schools and broadcast stations.
  - They will have to catch up to the rest of us.

Decorrelation of reverberation:

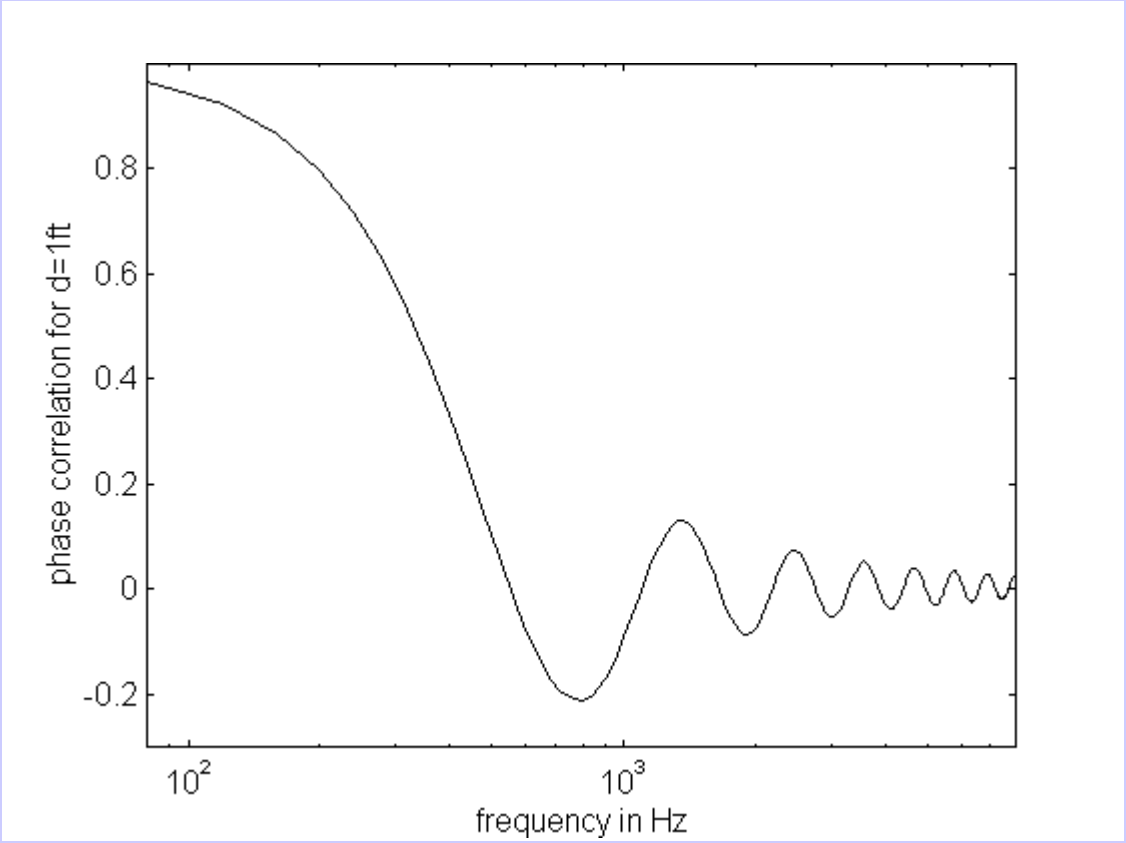
- How do we record the reverberation so that all four lateral signals are independent?
  - Pairs of microphones are independent if they are set at a particular angle that depends on their directivity.
  - Two pairs of microphones are independent of each other when they are separated by the critical distance in the hall.

Optimum mike angles



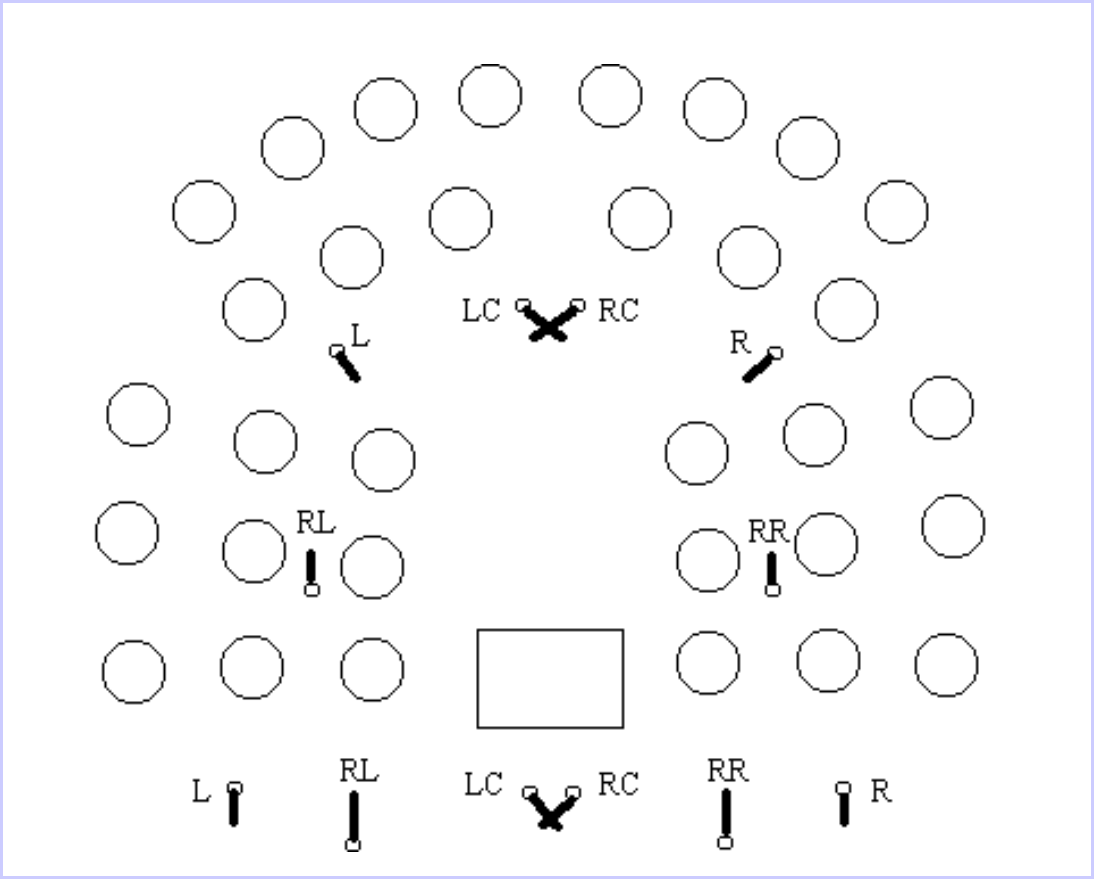
Reverberation is decorrelated in a microphone pair if the mike angle is correctly chosen. These angles are generally also optimum for the front image.

Decorrelation with distance:



Correlation between two omnidirectional microphones separated by 25cm in a reverberant field. Note the high correlation at 200-300Hz, and a negative correlation at 800Hz. Small distances are not sufficient for decorrelation!

Orchestra in Surround



Two supercardioid pairs are used for the center of the ensemble. They are mixed left-to-center and right-to-center. Four cardioid microphones point rearward, in front and behind the conductor. These create the rear channels.

Market for Surround

- Surround receivers are the dominant force in the US market
- Almost all surround recordings are cinema
- DVD based audio surround disks are very expensive to produce.
- A new market for matrix surround is developing in automobiles

Surround in Cars

- the auto is ideal for surround
  - particularly a matrix surround system with a very wide listening area.
  - If decorrelation can be preserved or enhanced the tiny playback space can be enlarged.
  - The difference between standard stereo and L7 surround in a car is night and day.
    - The distance of the front image expands beyond the windshield
    - The walls seem to disappear.
- By 2005 there will be >200,000 on the road.

Broadcast to Cars

- The number of cars with a high quality surround matrix is likely to increase steadily.
- This represents a large market for surround broadcast material - which is compatible with stereo.
- Music, advertising, film - all can be broadcast in matrix surround.

Conclusions 1:

- 1. Surround is coming - because it sounds better
- 2. There are several competing surround formats - Dolby AC3 and matrix surround dominate the market.
- 3. The advantages of surround are largely lost if the reverberation is correlated.
- 4. Recording for a large listening area and high decorrelation between channels requires specific techniques.

Conclusions 2:

- 5. Available two channel to surround matrices differ significantly in design and performance
- 6. A market for two channel matrix surround music and broadcast is developing in automobiles.