Räumliches Hören in Theorie und Praxis: Wie ergänzt man Tiefe und Halligkeit mit künstlichem Nachhall ohne Beeinträchtigung der Deutlichkeit

The Theory and Practice of Perceptual Modeling - How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity

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

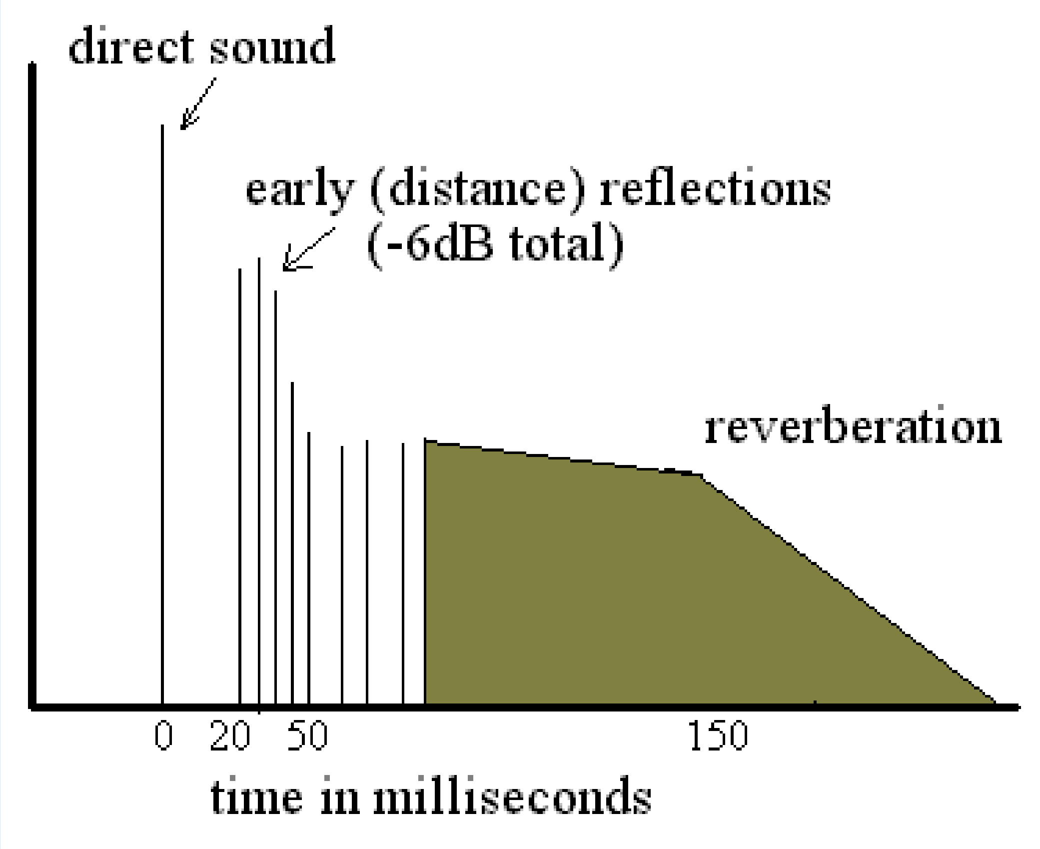
Main Message

- There is an ideal reverberation profile
  - This profile is required by human perception.
- The ideal profile is NOT provided by most acoustic spaces.
  - Most real rooms have too few very early reflections
  - and too many reflections in the 50 to 150ms time range
- Common microphone technique partially achieves the ideal profile,
    - but only for some instruments in a group.
    - And at the cost of a restricted listening area, reduced intelligibility, and excessive coloration.
- We can make a more natural recording by using artificial reflections and reverberation.

Second message: The need for a large listening area

- The great majority of your customers do not listen in the "sweet spot"
  - It is possible and necessary to make recordings that have a large listening area.
  - Large listening area requires amplitude panning.
    - We need significant separation between left and right, between the center and left, and between center and right.
    - Time delay panning is unacceptable if you want a large listening area.
  - And you must keep the reverberation in the four outer channels decorrelated.
- It is difficult to record three front channels from a single stand and achieve both good three channel separation and reverberant decorrelation.
- It is impossible to record four decorrelated reverberant signals from a single stand.

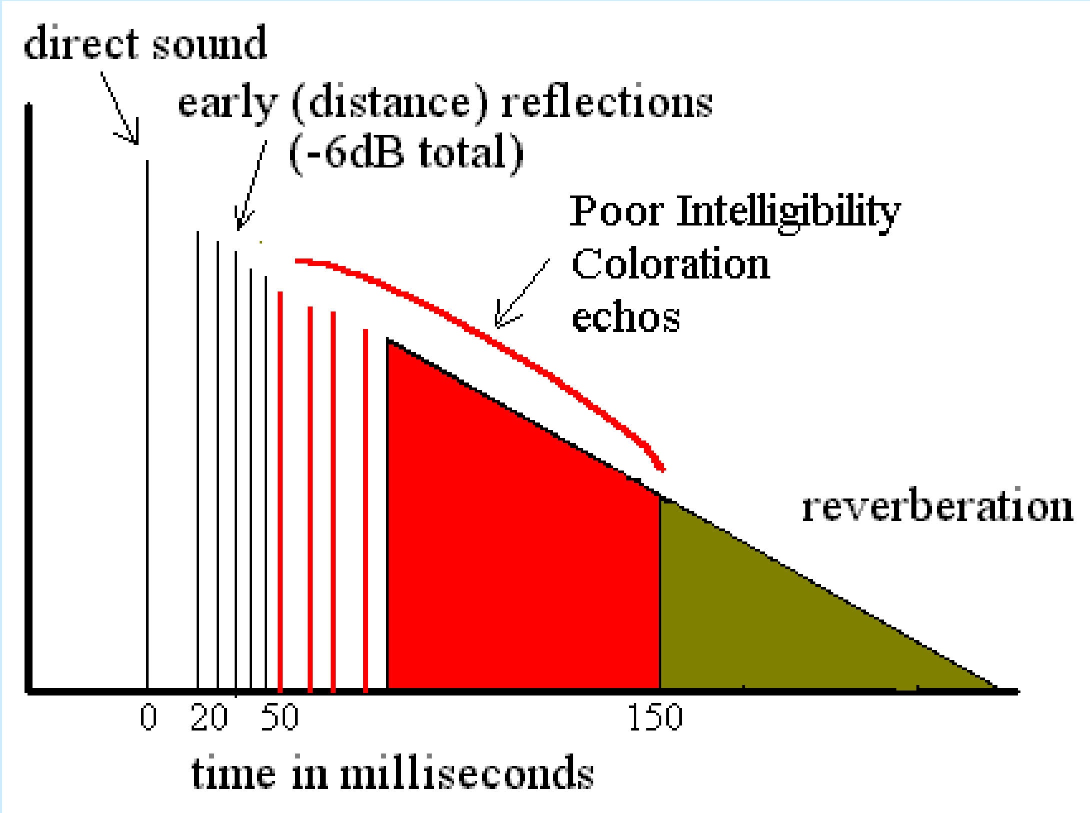
The Ideal Reverberation



– has 20ms to 50ms reflections with a total energy -4dB to -6dB

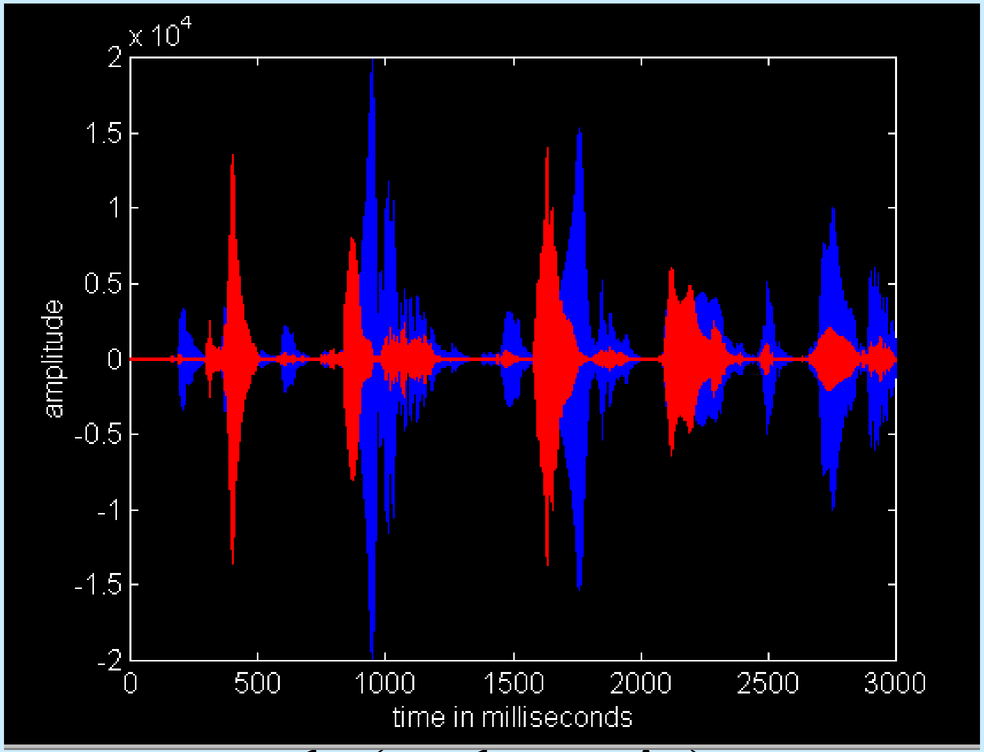
– has relatively little energy from 50 to 150ms.

Most real rooms



– Have exponential decay
– to get enough early reflections and reverberation, we get coloration and poor intelligibility.

Why do we want low energy at 50-150ms?

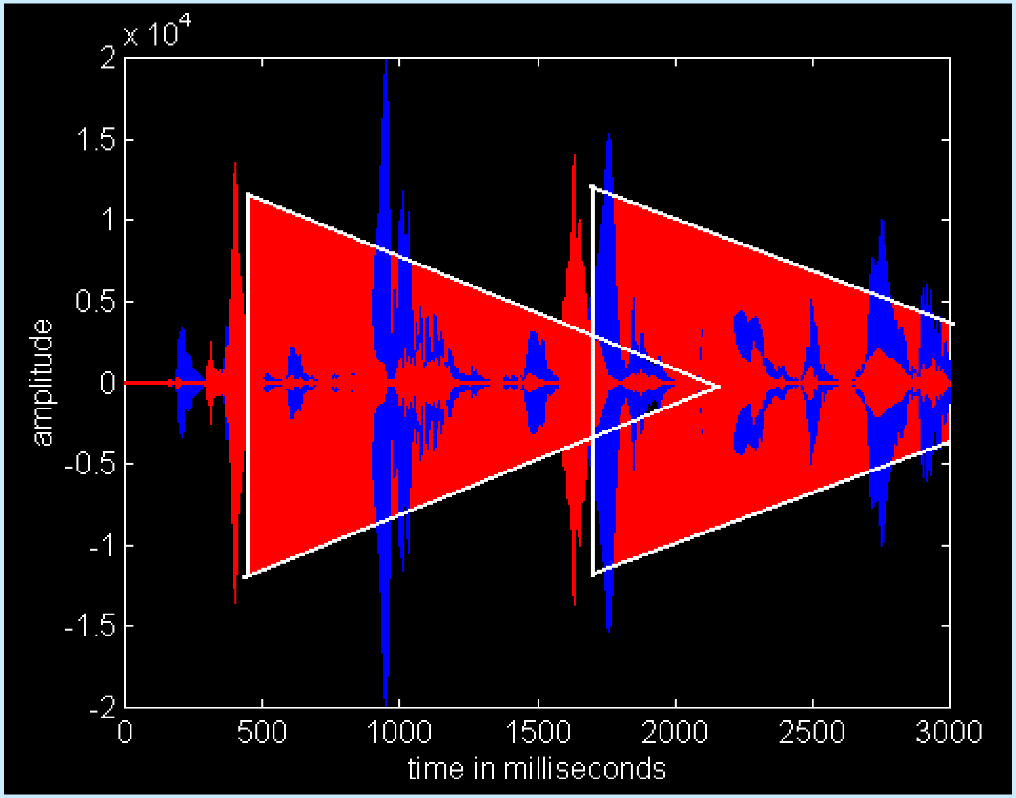


Third-octave filtered speech.

Blue 500Hz.
Red 800Hz

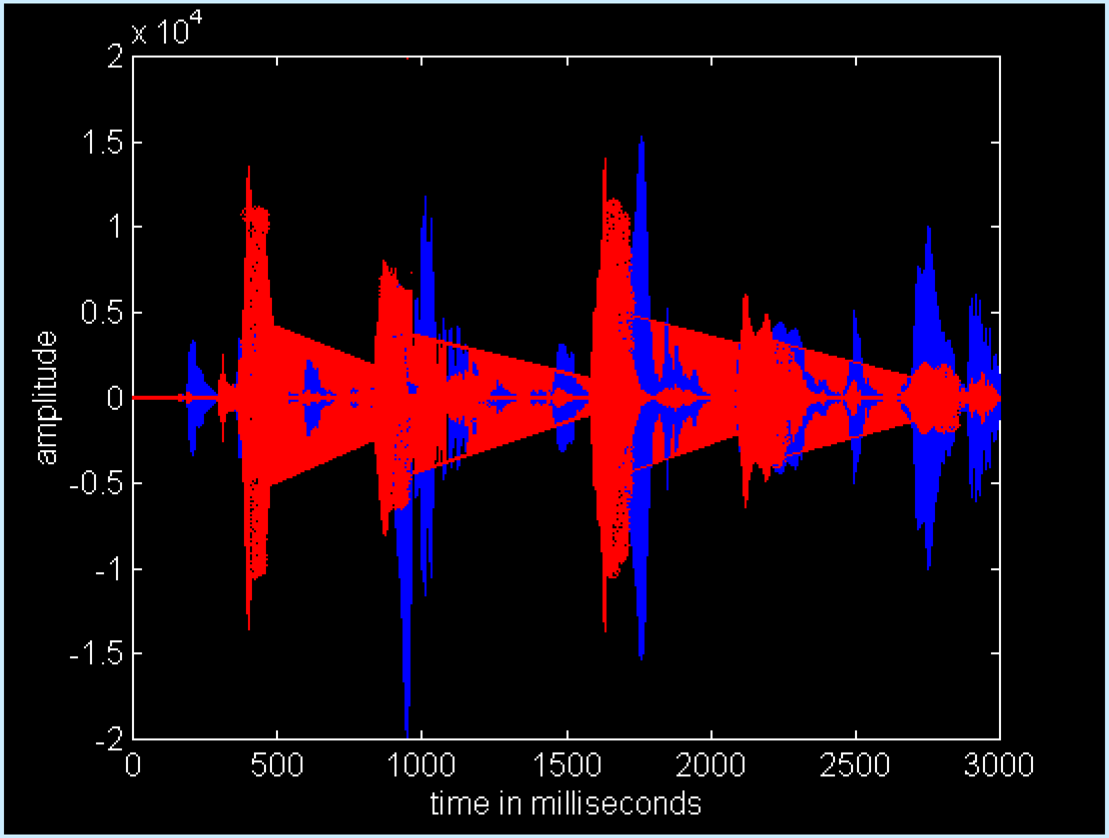
- Because speech (and music) are composed of streams of sound events (notes)
  – with ~ 200ms spaces between each event.
  – a series of such events form a perceptual stream.
  – reflections at 50-150ms make separation of events impossible.

Why exponential decay is problematic.



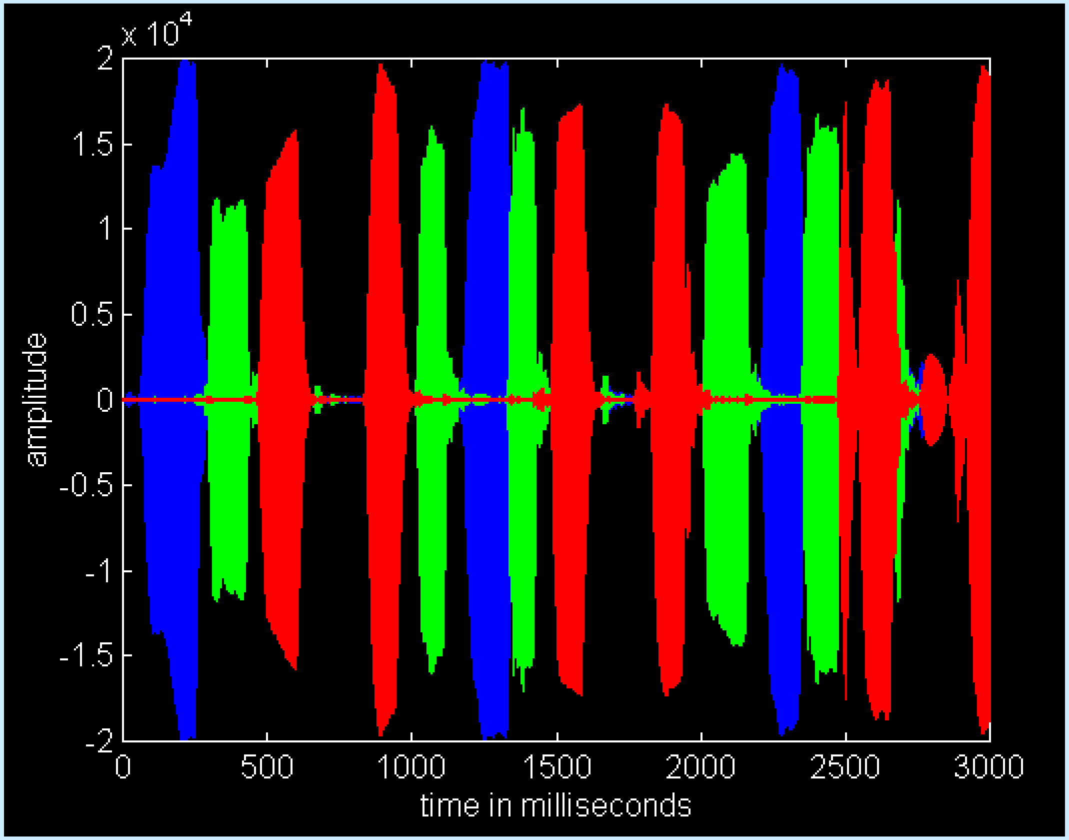
- Adding reverb with exponential decay masks many sound events.
- Comprehension becomes impossible.

Why our "ideal" decay is better



- Strong early reflections combine with the direct sound,
  - But give a sense of blend and space.
  - Intelligibility stays high because the reverberation does not obscure the spaces between sound events.

A little Bach in 1/3 octaves



Bockflute musik

Red = 500Hz, Green = 630Hz, Blue = 800Hz

Although the notes sound continuous, on the basilar membrane each note forms a sound event with plenty of space for reverberation to be heard.

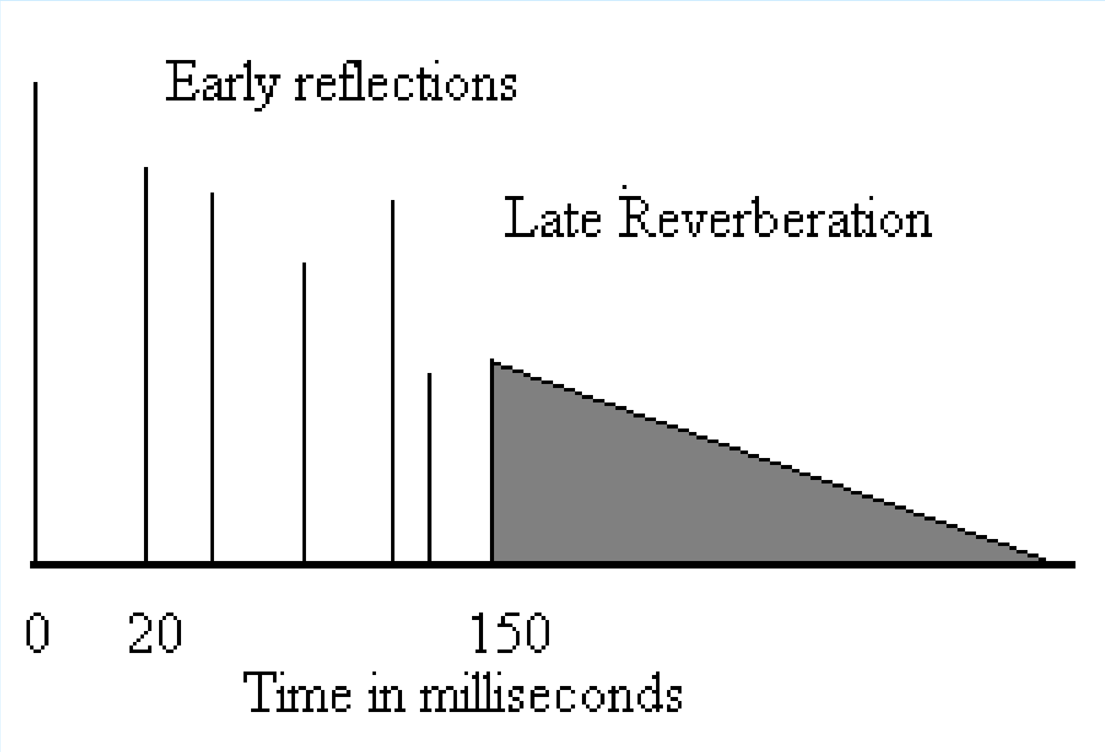
Spatial Hearing

- Horizontal localization and the perception of the spatial properties of a room BOTH rely on the same neural circuitry.
- We detect both through the ITD (Interaural Time Differences) and IID (Interaural Intensity Differences).

- But human hearing (and all human perception) only responds to stimuli that are NOT CONSTANT.

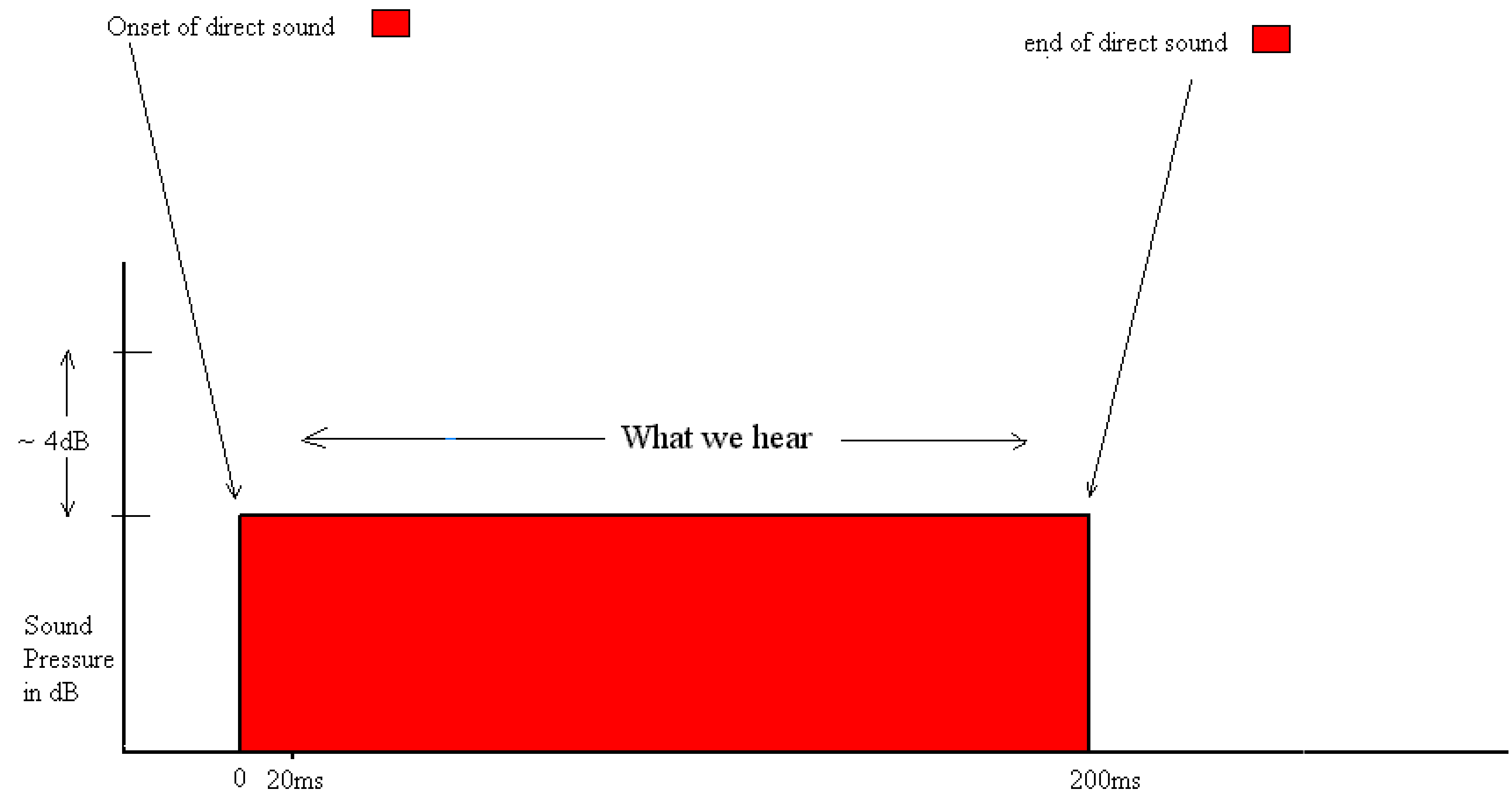
- The START of stimuli are important for localization
- The ENDS of stimuli are important for room perception

Impulse response is not what we hear



- The impulse response is the sound of a pistol, not the sound of music.

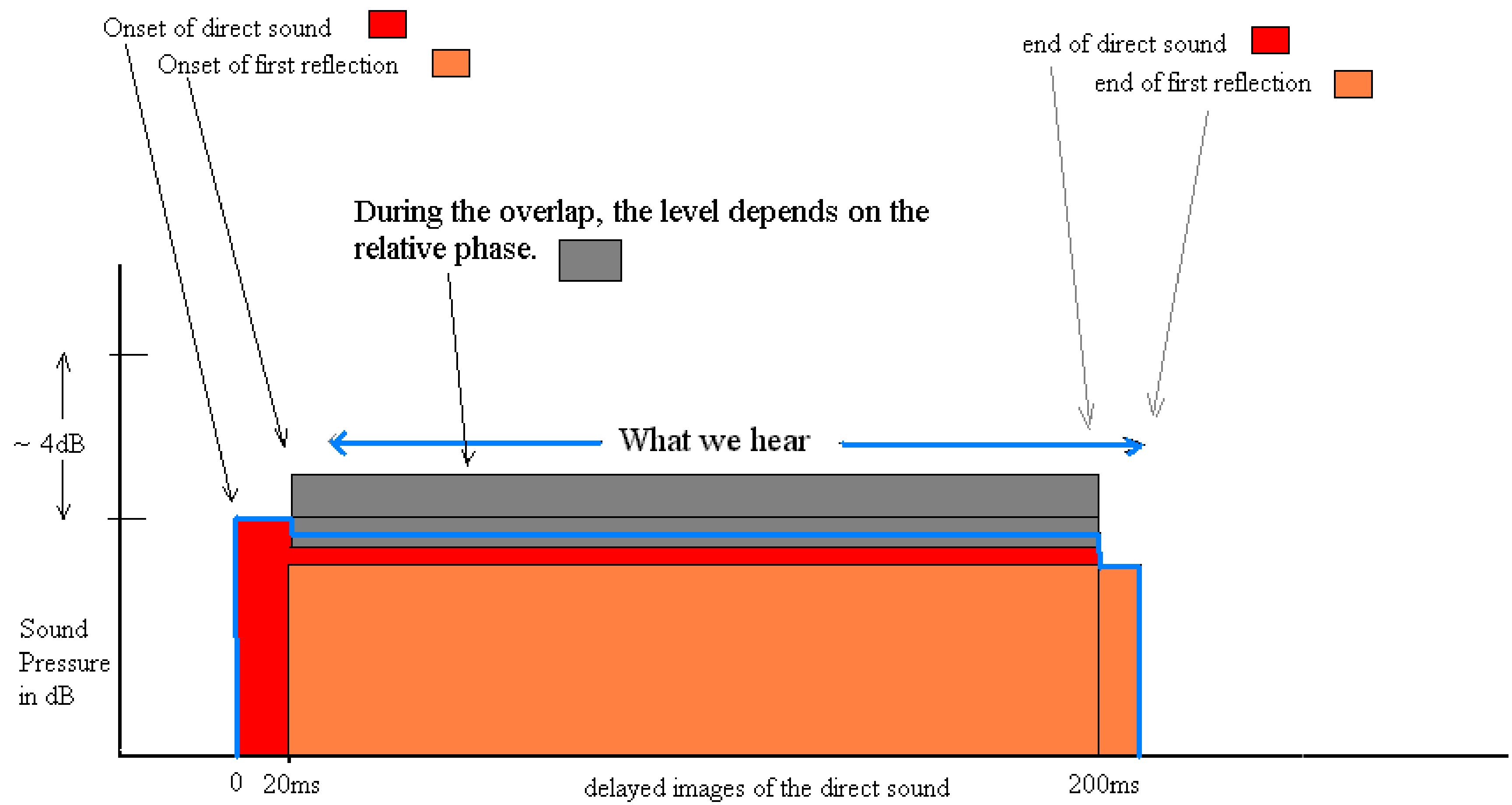
What we hear with notes:



Sound pressure at one ear from a single 200ms note

- We start with a 200ms note, showing the direct sound only.
- Notice it has a clear beginning and a clear end.

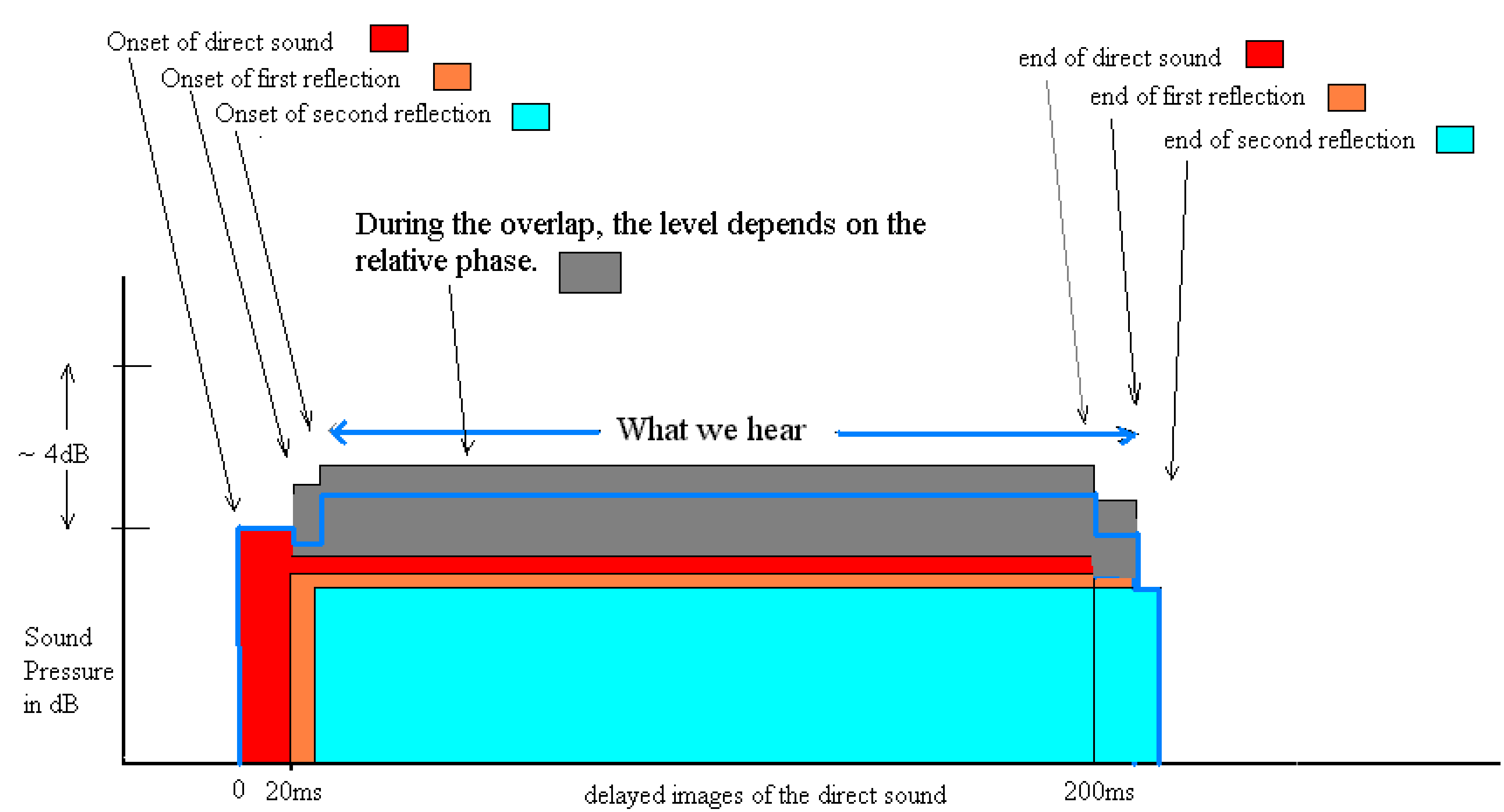
What we hear with notes:



Sound pressure at one ear from a single 200ms note

- Now add the first reflection. The level may go up or down, depending on the phase.
- Notice the direct sound is not corrupted for the first 20ms.

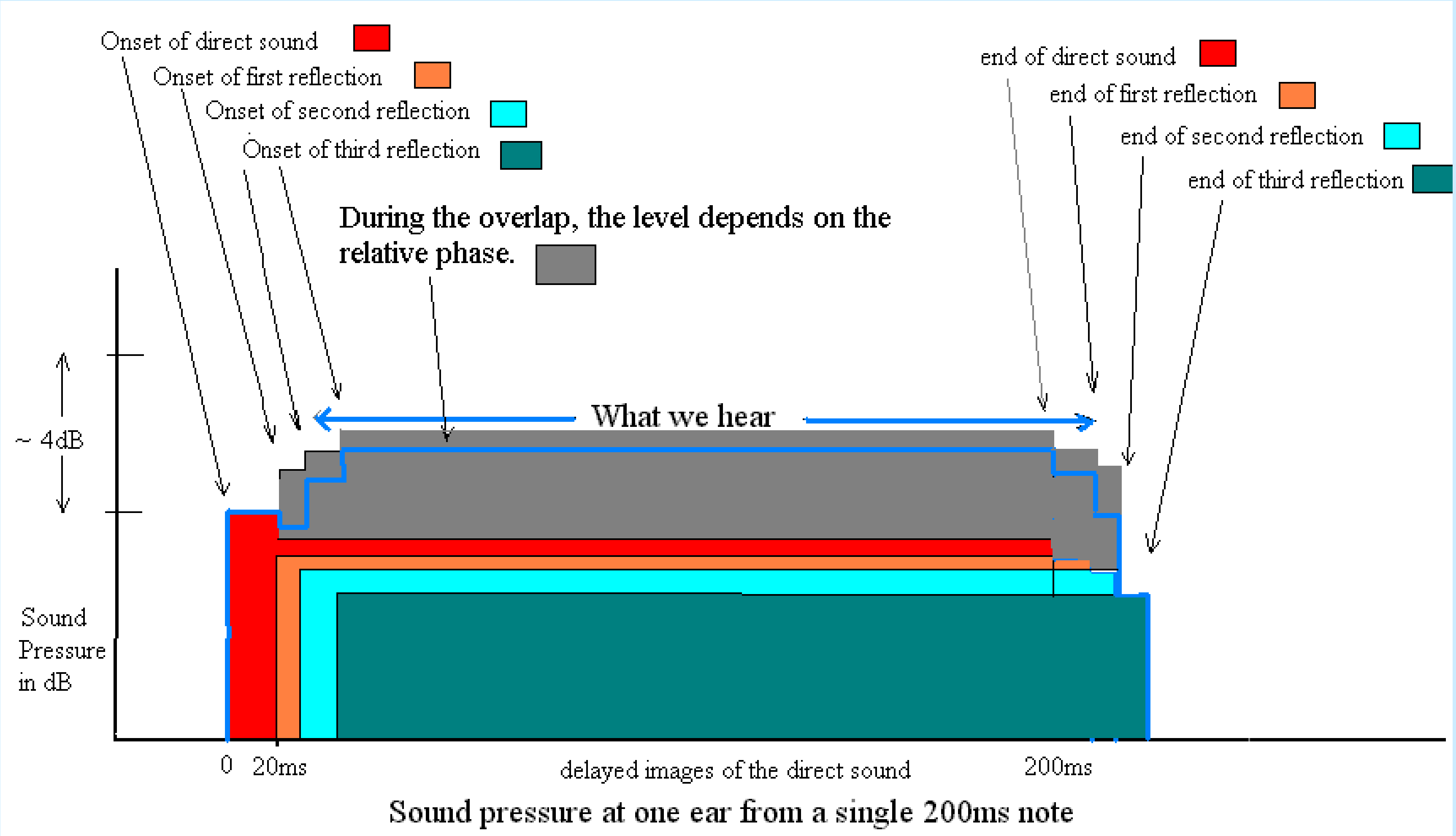
What we hear with notes:



Sound pressure at one ear from a single 200ms note

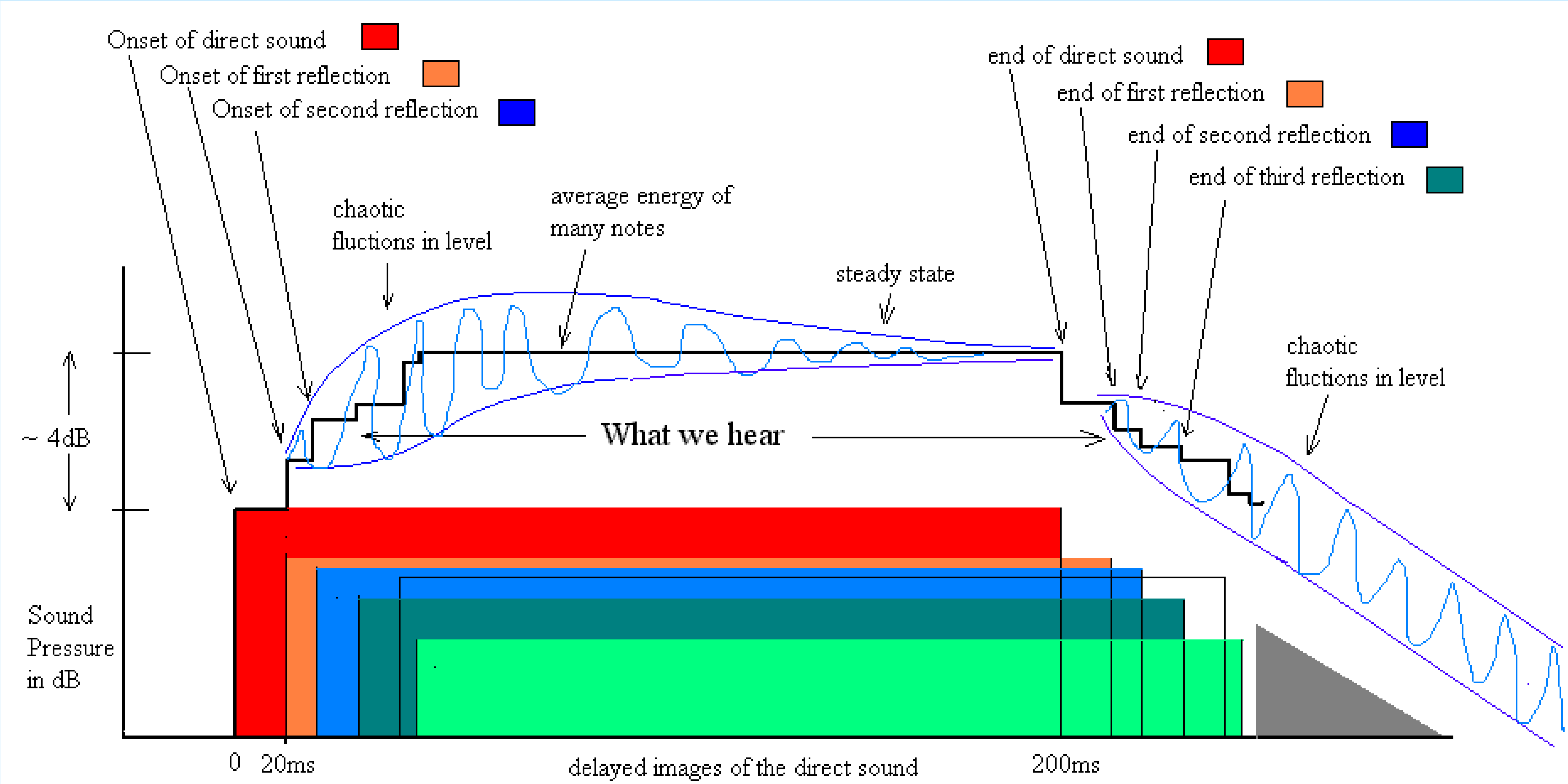
- Now add the second reflection. Once again, the level may go up or down, depending on the phase. If the reflection is lateral, the other ear may have a different phase, and a different level.

What we hear with notes:



- Now add a third reflection. Once again the level changes seemingly randomly.

What we hear with notes:



Sound pressure at one ear from a single 200ms note

- In the end, the level is uncorrupted until the first reflection, and then fluctuates until the room reaches steady-state.
- When the direct sound ends, the fluctuations start again.

Spatial Fluctuations

- We do not hear the fluctuations themselves. But the fluctuations in level and phase are <u>different</u> in the two ears.
- It is the <u>difference</u> in the fluctuations between the two ears that gives rise to the spatial properties of the sound.
- Fluctuation differences <u>during</u> 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 <u>after</u> the end of the event are heard as envelopment.
  - And as a clue to the size of the space.

Detection of reflections

- At low levels early reflections are detectable *only* by their effect on interaural fluctuations.
  - We can not detect the exact direction.
  - We can not detect the time of arrival (if it is less than 50ms).
  - We can not form an “image of the room” through early reflections!
- Low level early reflections have only TWO perceptions
  - a room impression (raumlichkeit)
  - distance (tiefe).
- High level early reflections cause coloration.

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.
  - The image of legato strings (with a slow rise-time) is broadened by early reflections.
  - Speech and solo instruments remain sharply localized
  - because there are no reflections during the rise-time.

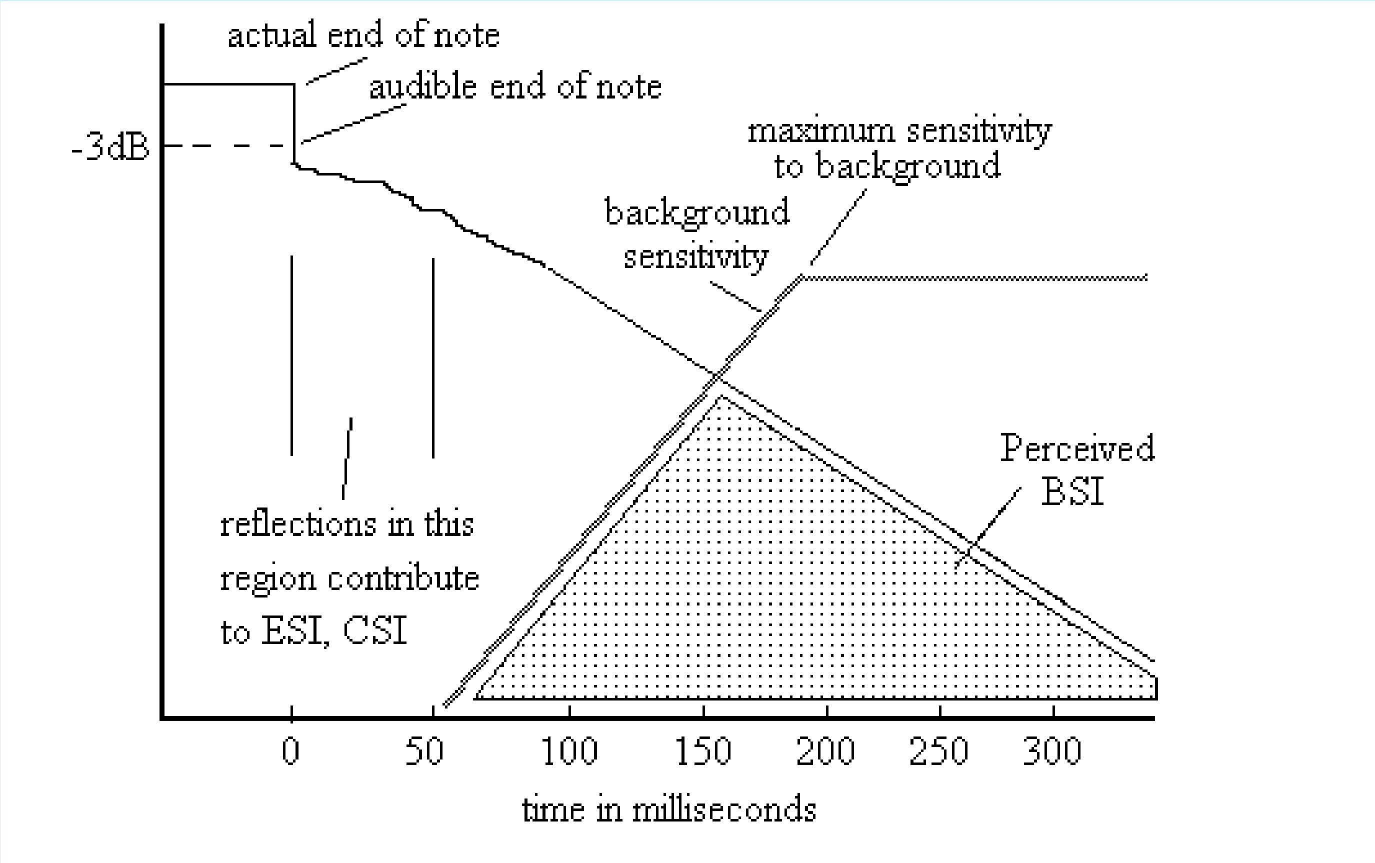
Background Streams

- 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.

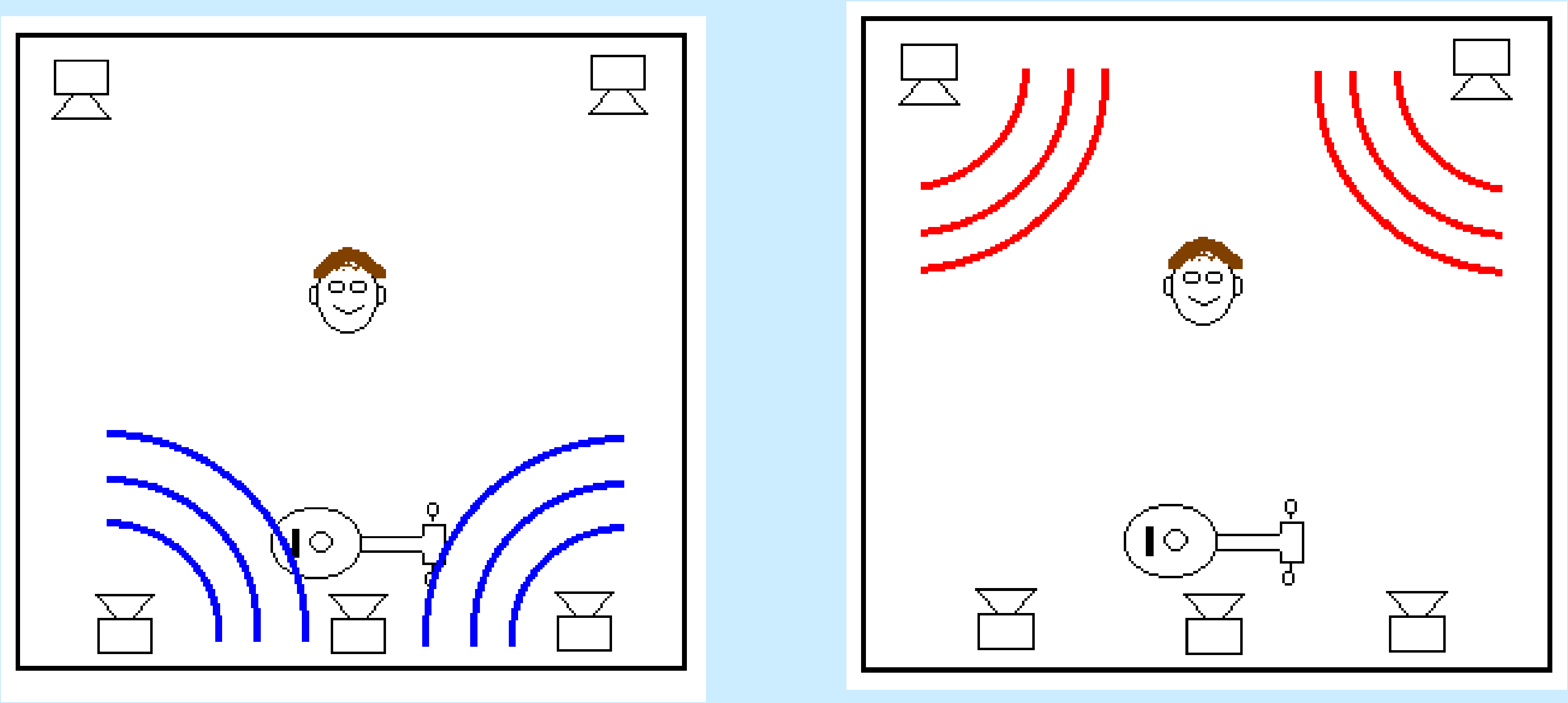
Separating the background takes time.



Haas Effect

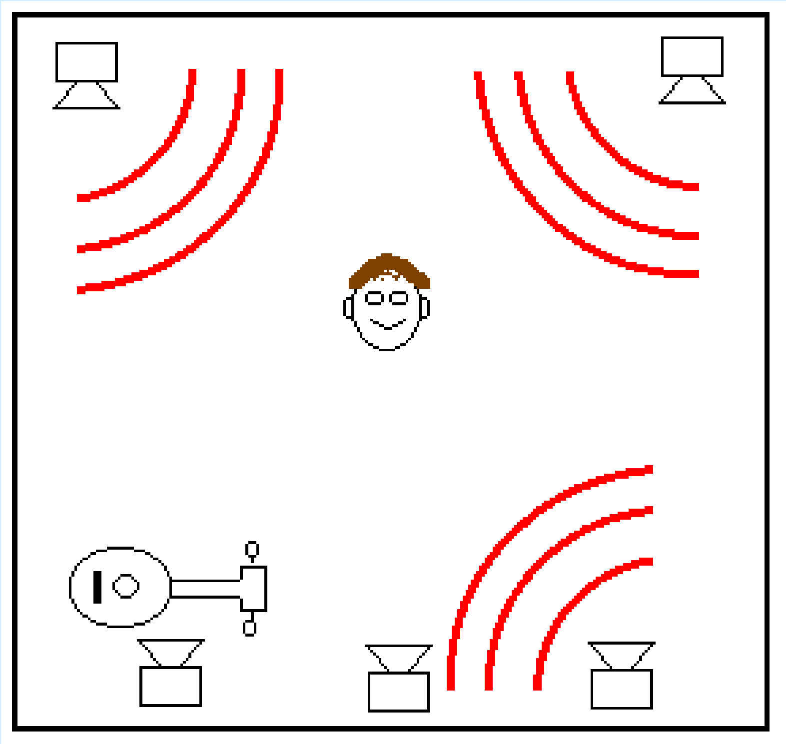
- The Haas effect arises because the neural circuits wait 50ms after the apparent end of a sound event to be sure it is really over.
- Sound which arrives during this waiting period is assigned to the previous event
  - it is not heard as a separate event,
  - although an “acoustic” impression is created, the impression is centered on the direction of the sound, regardless of the direction of the reflection.

Direction of early reflections



- For reflections that arrive 20ms to 50ms after the end of a sound,
- It is not possible to detect if they come from the front or the rear.
- But it is more natural if they come from both front and rear.

When the sound comes from the right, the early reflections should come from all three other directions.



- Reflections from the same direction as the source are either inaudible or undesirable.

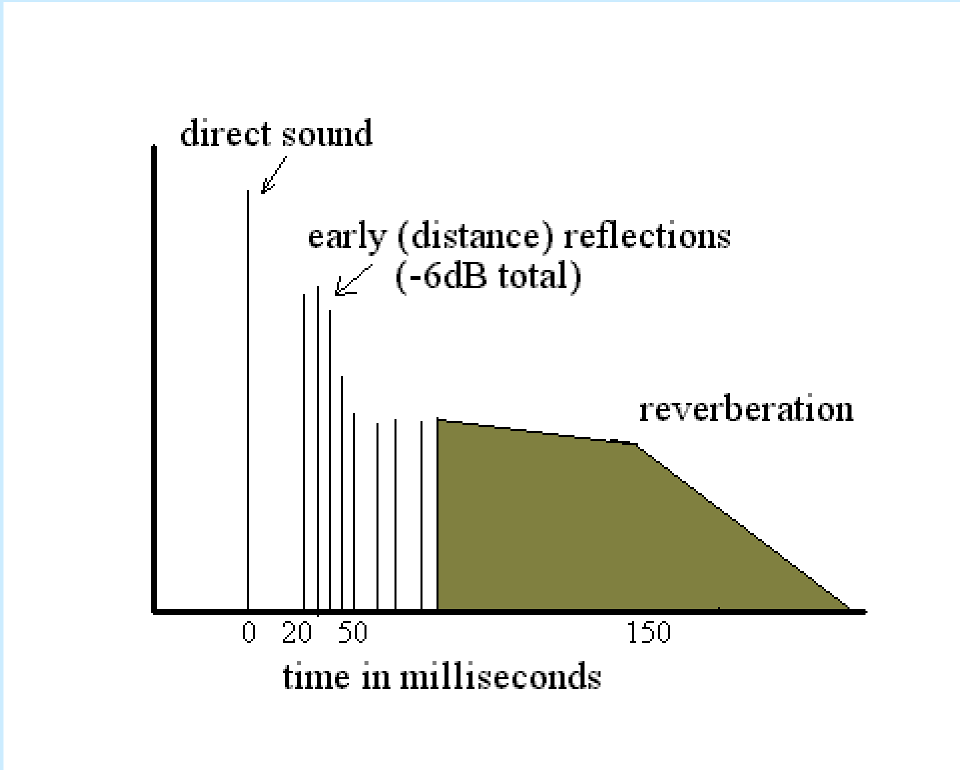
Reflections between 50ms and 150ms

- Reflections between 50ms and 150ms add a sense of distance, but at the cost of reduced intelligibility
- Reflections in this range sound “muddy”.
- They 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 again

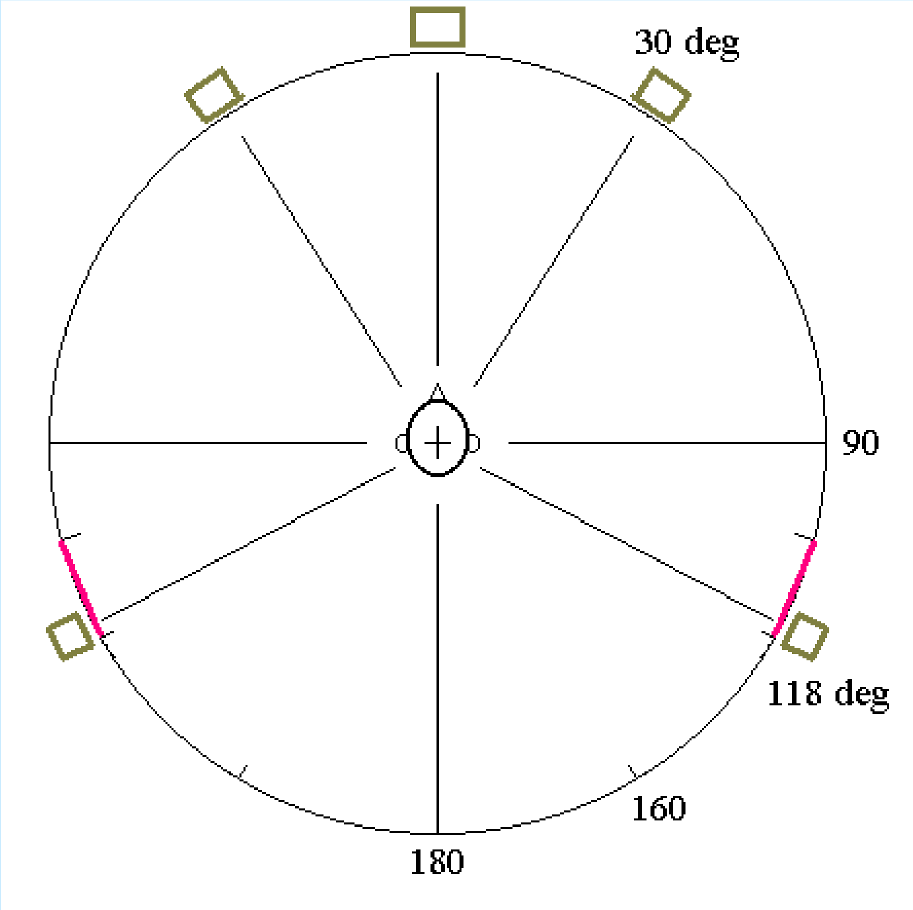


A strong early lateral field gives the sound distance and an acoustic integration. Reflections in the range of 50 to 150ms are minimized, while keeping reflections after 150ms strong.

How do we record?

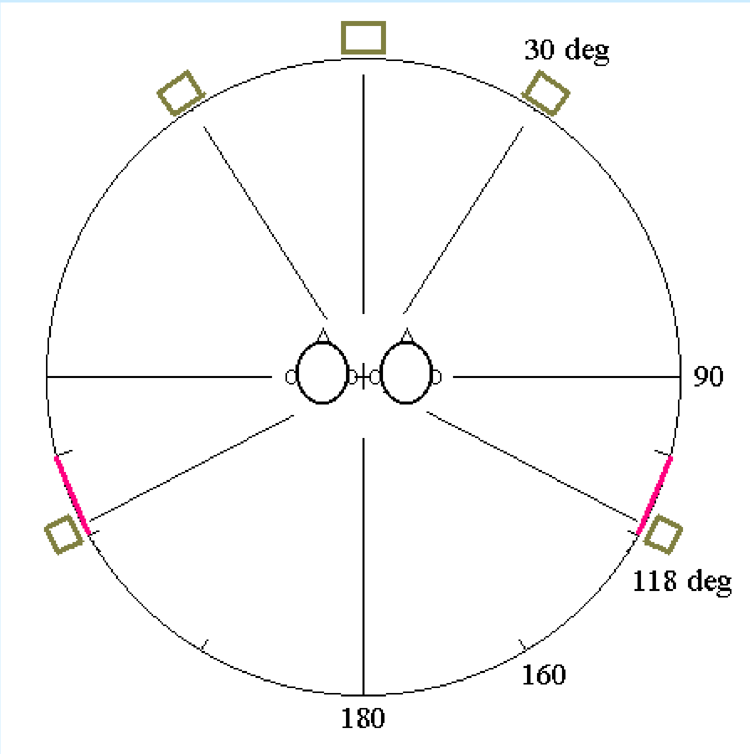
- Who Are Our Customers?
- How do they listen?
  - Do they listen by themselves at a single point
    - Or do they listen wherever they want
  - Do they listen in groups?
- We think we have answered these questions without even asking them.
- We have a STANDARD

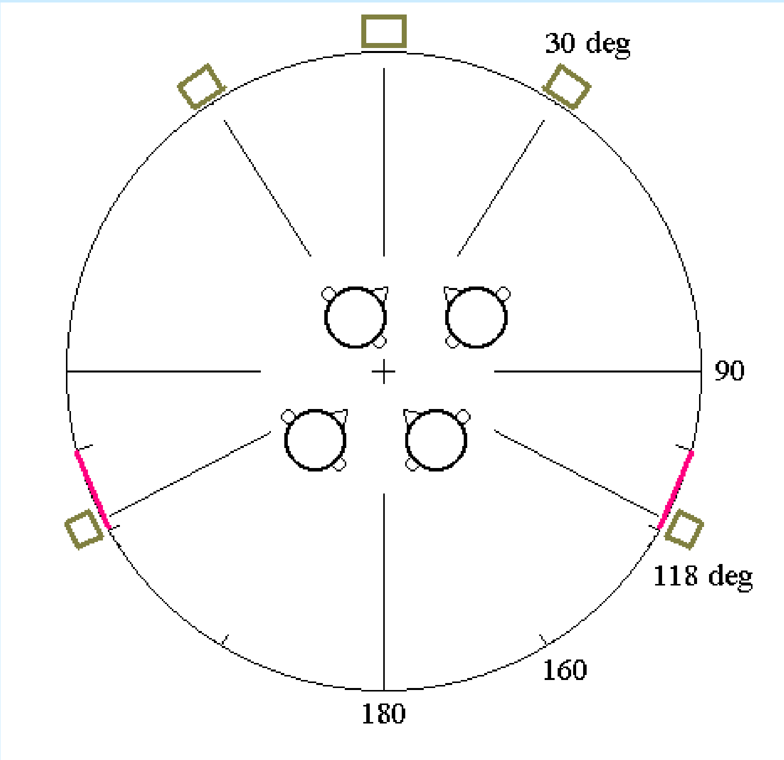
The Standard



- Is for a single listener at a single point
- There is no “listening area”
- This may be your standard but it is not mine!

Is the standard sensible?





- If we listen with friends
  - will someone - or everyone - be disappointed?
  - YES! They <u>will be</u> disappointed if the recording uses time delay panning!

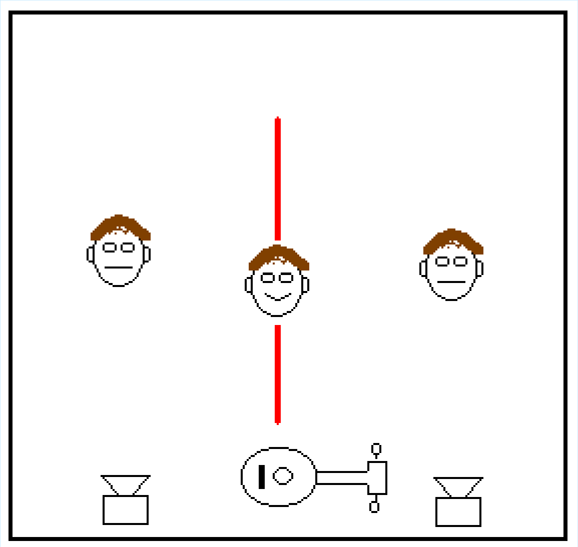
The Customer Speaks!

- Nearly all the surround systems purchased are for films.
- Films are a group experience.
  - People set up their systems any way they can listen with their families and friends.
- Like it or not - your customers do not use the standard speaker layout,
  - and you should not make recordings that require it!

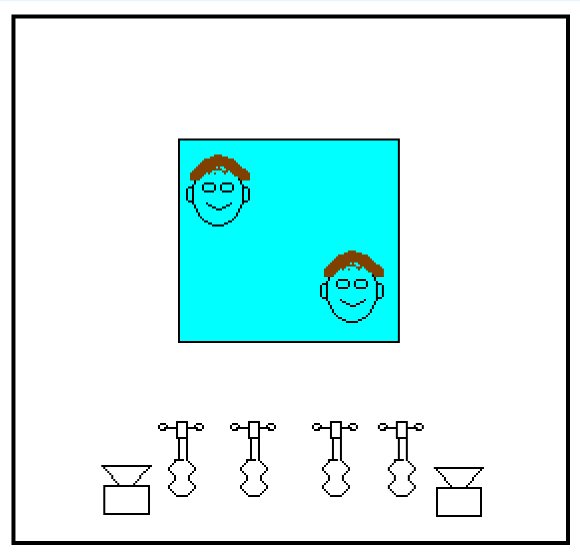
- We need to make recordings that sound excellent over a wide listening area!
  - If they sound a little better in the center that is OK.

We Can Do It!

- The secrets are: high amplitude separation, no time delay panning, and decorrelated reverberation.



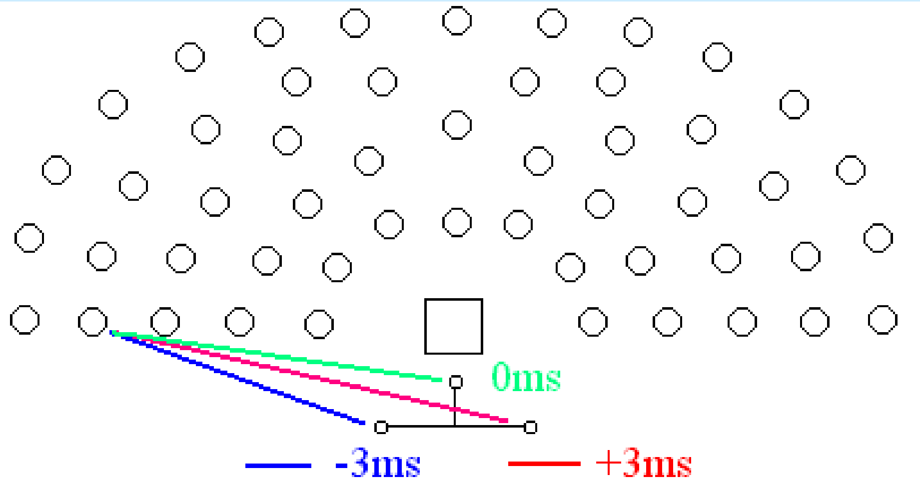
Two channel recordings with a strong center image localize well only on a line between the speakers.

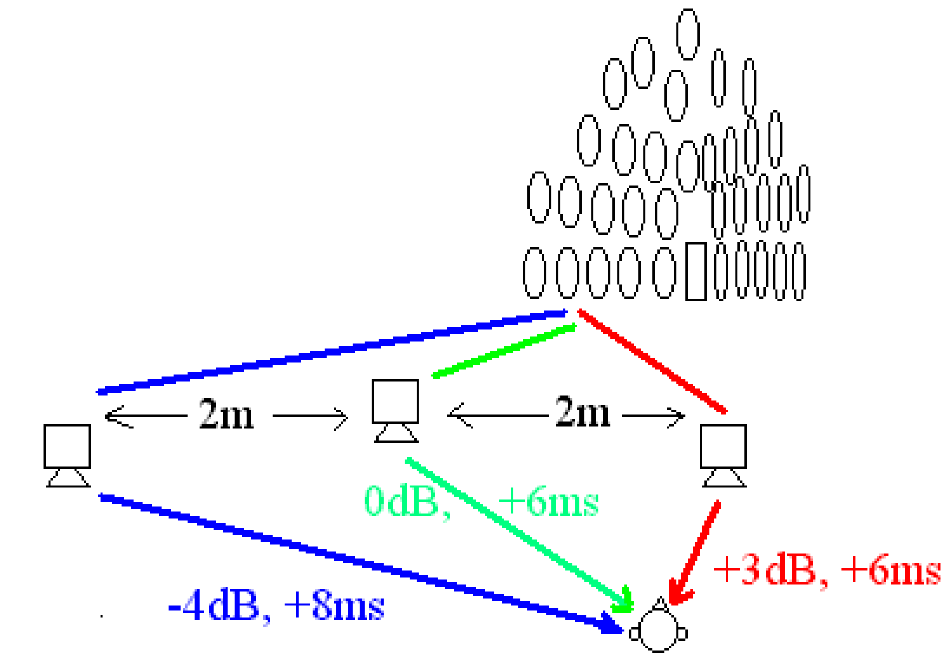


Two channel recordings with a broad source sound good over a wide area if they have high amplitude separation.

Adding a hard center channel improves the listening area for both sources.

Time delay panning outside the sweet spot.

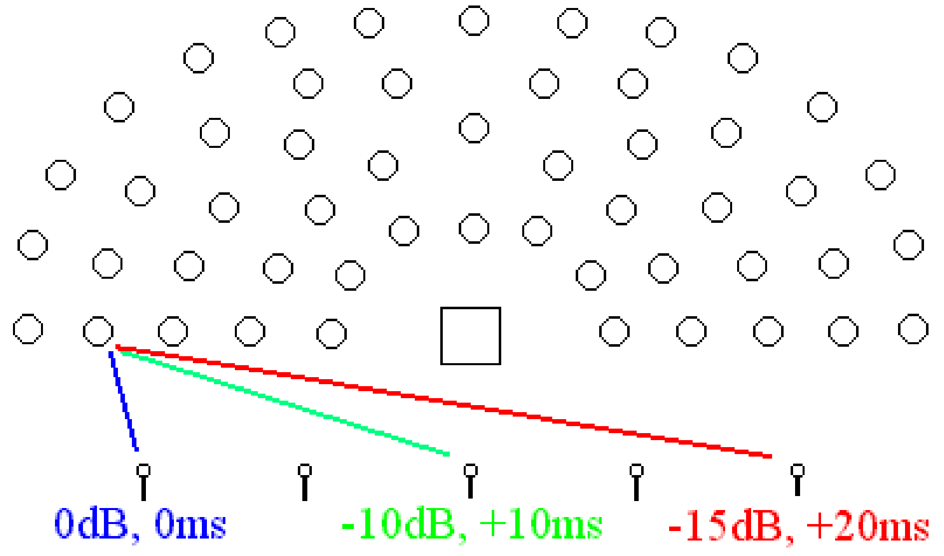


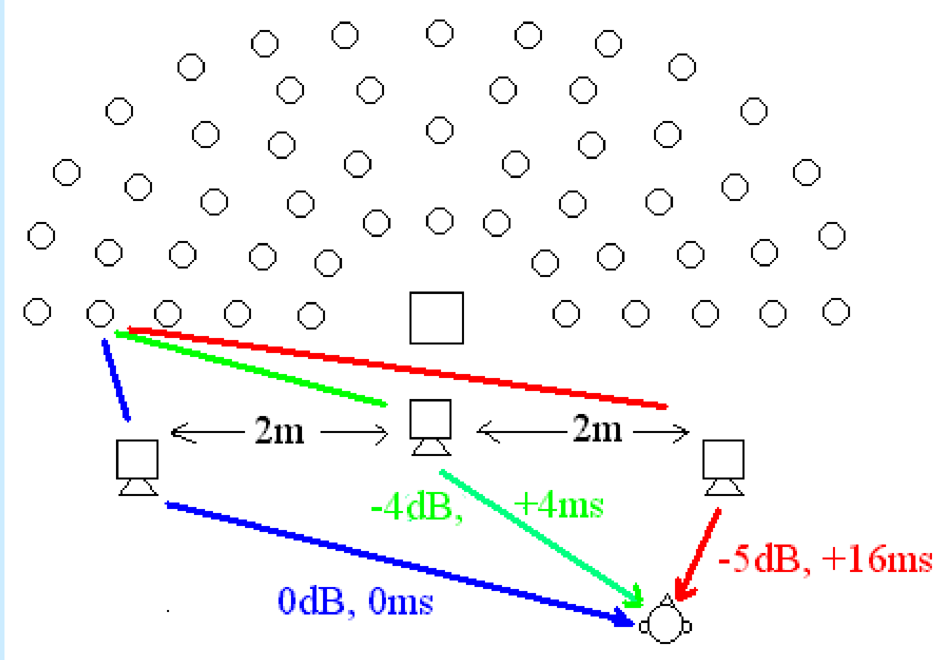


Record the orchestra with a "Decca Tree" - three omni microphones separated by one meter. A source on the left will give three outputs identical in level and differing by time delay.

On playback, a listener on the far right will hear this instrument coming from the right loudspeaker. This listener will hear *every* instrument coming from the right.

Amplitude panning outside the sweet spot.



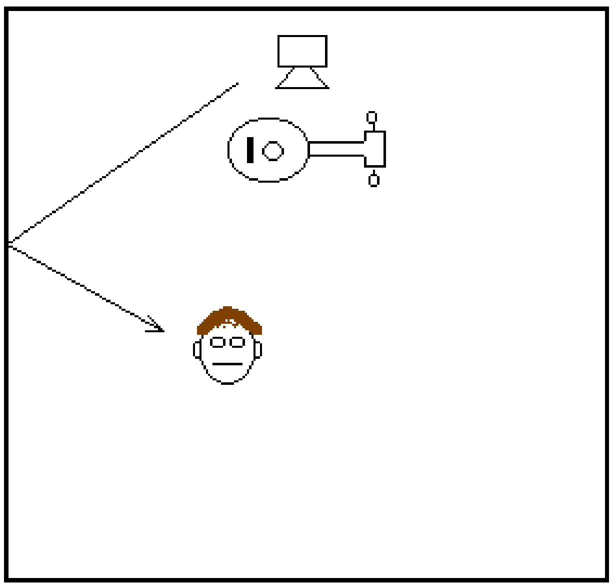


If you record with three widely spaced microphones, an instrument on the left will have high amplitude and time differences in the output signals.

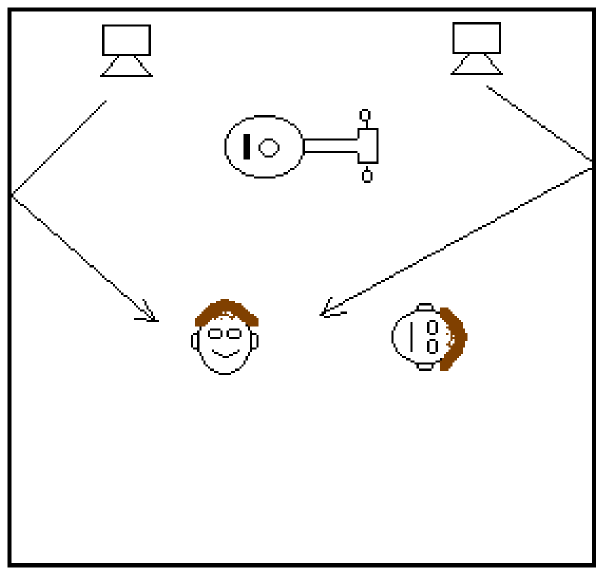
A listener on the far right will hear the instrument on the left. Now the orchestra spreads out across the entire loudspeaker basis.

3/0 versus 3/2

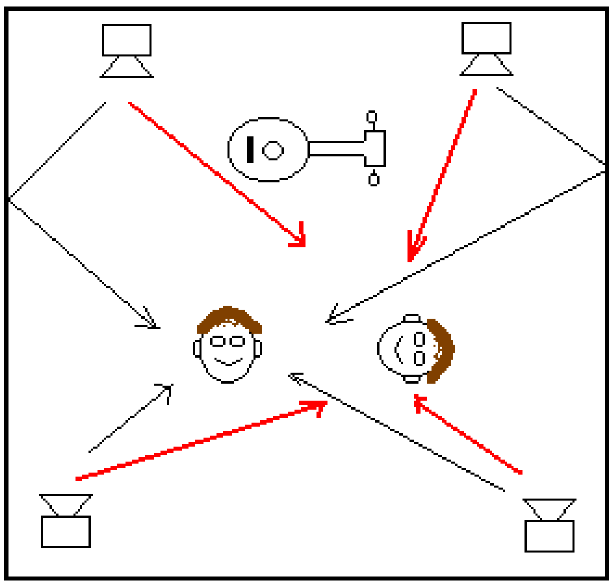
- OK, perhaps we need three speakers in the front, and amplitude panning in the front.
- Why do we need two additional speakers and channels?



Mono sounds poor because it does not reproduce the spatial properties of the original recording space.

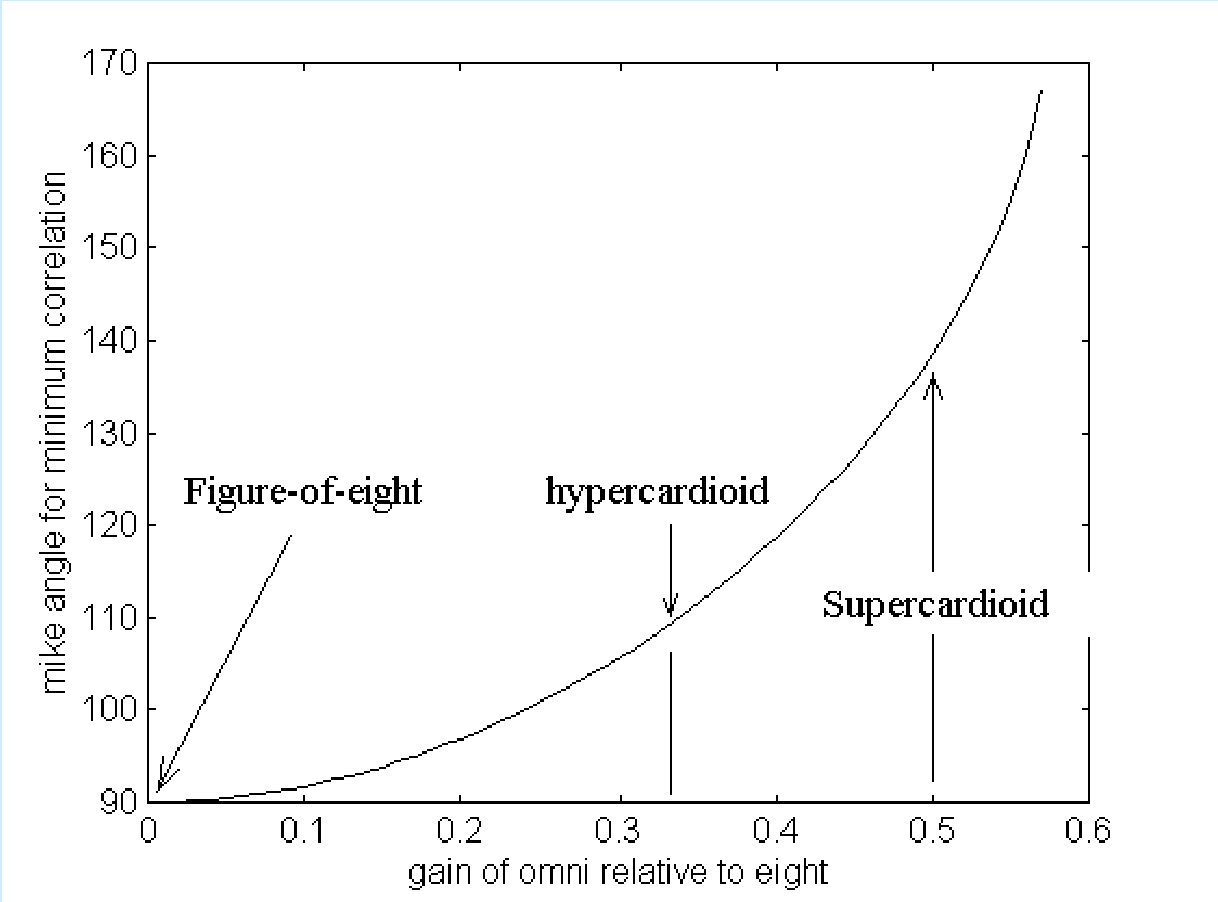


With decorrelated reverberation a few spatial properties come through, but only if the listener faces forward.



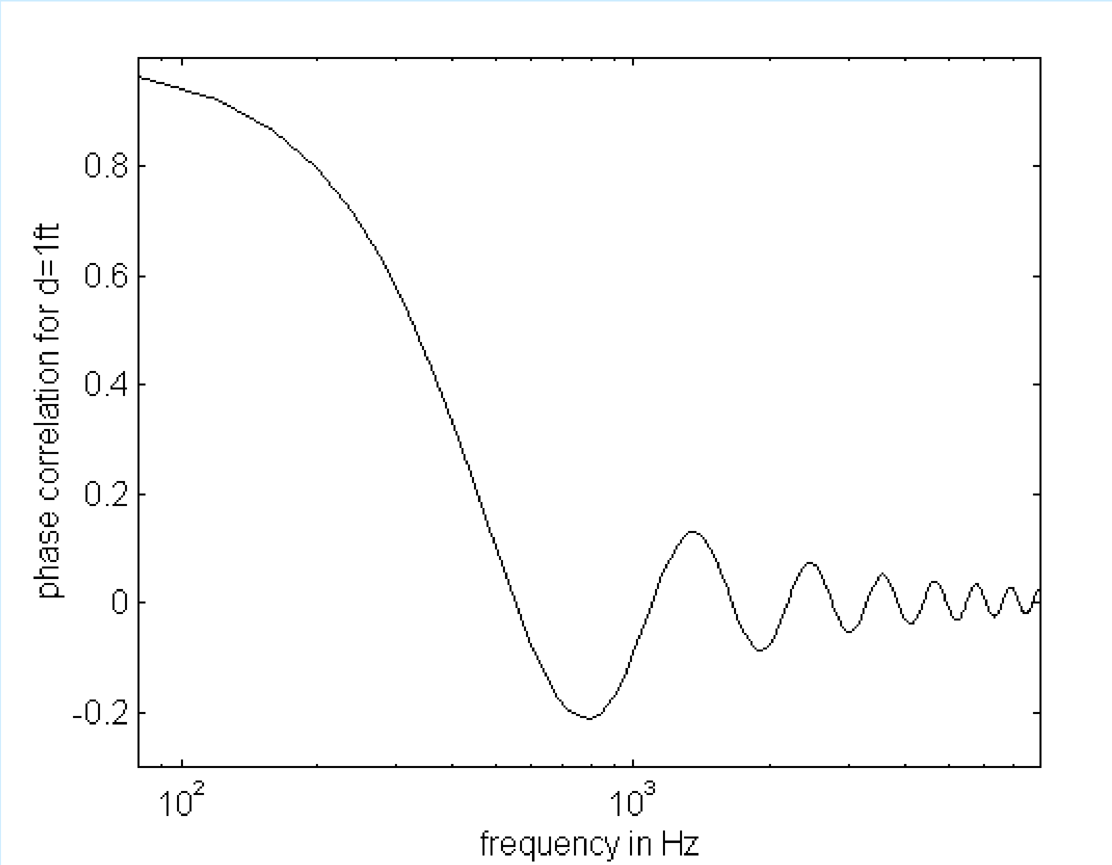
We need at least four speakers to reproduce a two dimensional spatial sensation.

Decorrelation for directional microphones



- Sorry, this diagram in “Production Partner” is in error!
- It is NOT possible to achieve decorrelation with cardioid microphones!

Decorrelation for omnis as a function of distance



- Notice high correlation below 300Hz, and negative correlation at 800Hz.
- Frequency and distance are inversely proportional.

Many of these goals can be achieved from two channels, with a matrix.

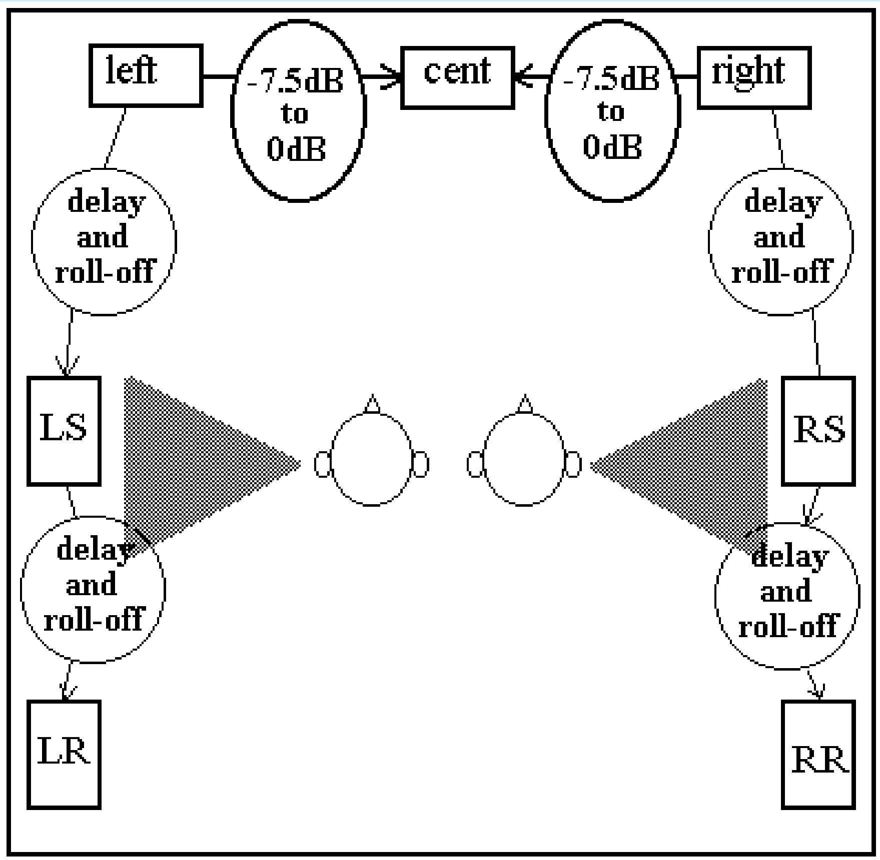


Diagram shows matrix decoding with NO directional steering.

- The center level is reduced so the front image is preserved.

- Delay and roll-off are used to derive two separate rear channels, while preserving the left/right decorrelation in the original recording.

When there is a directional signal, the matrix changes to enhance the desired direction.

The end result is good frontal imaging over a wide listening area, and high envelopment in the room.

Conclusions!

- 1. There is an ideal reverberation for recordings, and it is NOT provided by normal room acoustics of recording spaces.
- 2. Microphone technique has developed by trial and error to provide the ideal profile
  - but it often does not succeed
  - and at a cost of a very small sweet spot, coloration, and muddiness.
- 3. We can do better. With amplitude panning and deliberate control of early reflections, we can make recordings with a large listening area, excellent depth, and high envelopment.