The Physics and Psycho-Acoustics of Surround Recording

David Griesinger

Lexicon

dgriesinger@lexicon.com

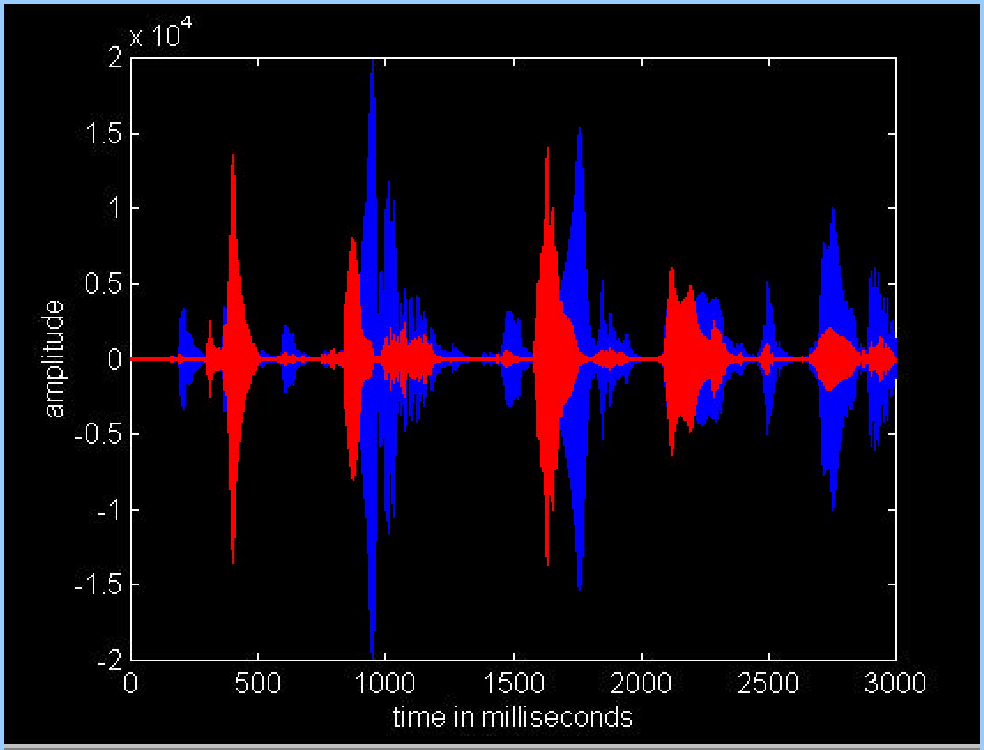
www.world.std.com/~griesngr

Major Goals

- To show how physics and psycho acoustics combine to produce <u>absolute</u> standards of quality in sound recording.
  - And for live sound in opera houses and concert halls!
- To show how to know when a recording meet these standards, and why.
- To show how to create a high quality recording in practice.
- To play as many musical examples as possible!

Human sound perception – Separation of the sound field into foreground streams.

- Engineers are entranced with frequency response, distortion, and time delay.
  - But human perception works differently.
  - Human brains evolved to understand speech, not to measure sound systems.



Third-octave filtered speech.

Blue 500Hz. Red 800Hz

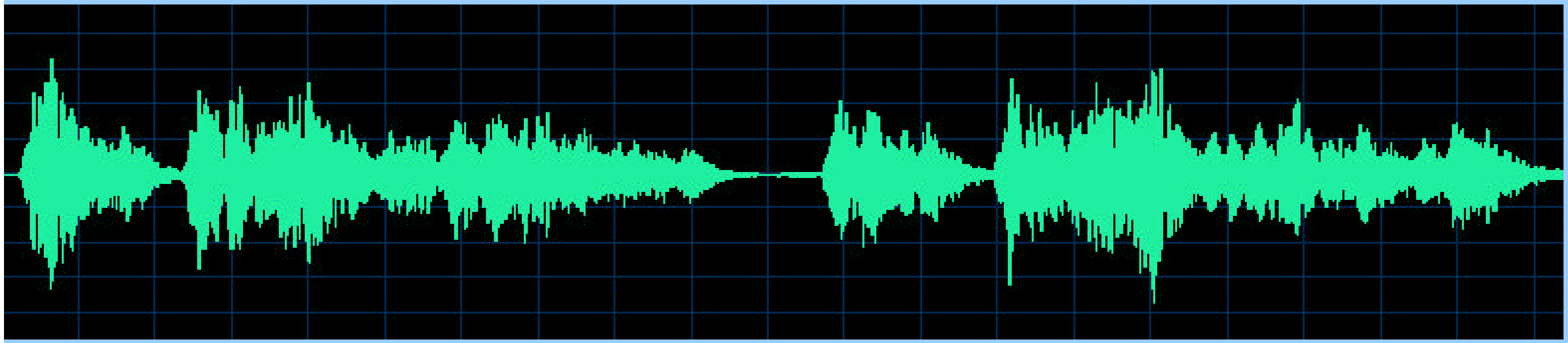
Speech consists of a series of *foreground sound events* separated by periods of relative silence, in which the *background sound* can be heard.

The primary function of human hearing is *stream formation*

- Foreground sound **events** (*phones or notes*) must be separated from a total sound field containing both foreground and background sounds (*reverberation, noise*).
  - Foreground events are then assembled into streams of common direction and/or timbre.
  - A set of events from a single source becomes a **sound stream, or a sound object**. A stream consists of many sound events.
    - Meaning is assigned to the stream through the higher level functions, including phoneme recognition and the combination of phonemes into words.

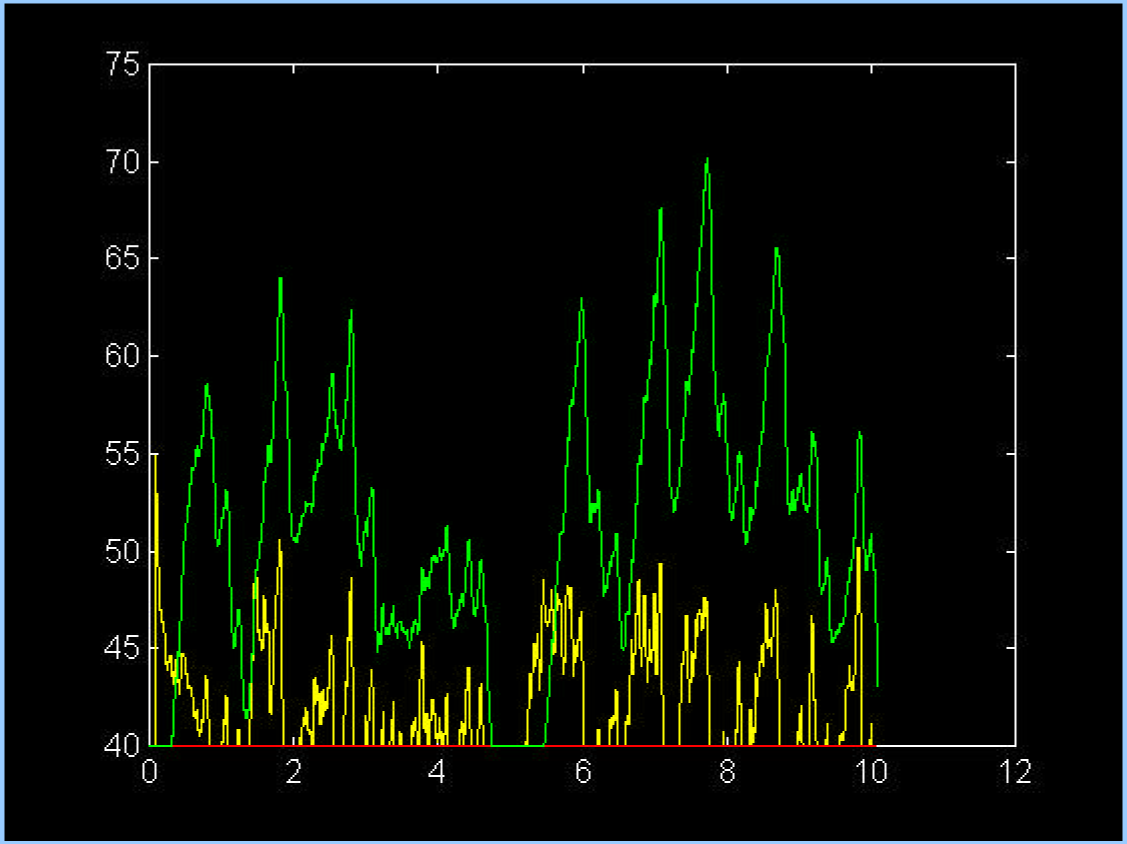
- Stream formation is essential for understanding speech
  - When the separation of sound streams is easy, intelligibility is high
    - Separation is degraded by noise and reverberation.
    - This degradation can be measured by computer analysis of binaural speech recordings.

Analysis of binaural speech



Reverb forward

Reverb backward



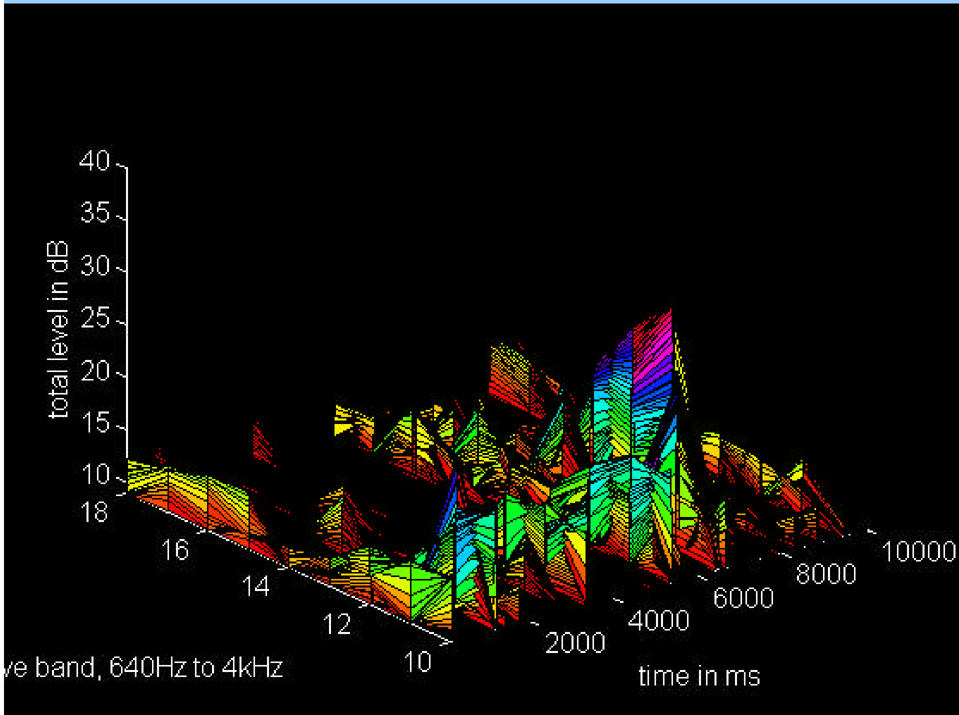
Analysis into 1/3 octave bands, followed by envelope detection.

Green = envelope

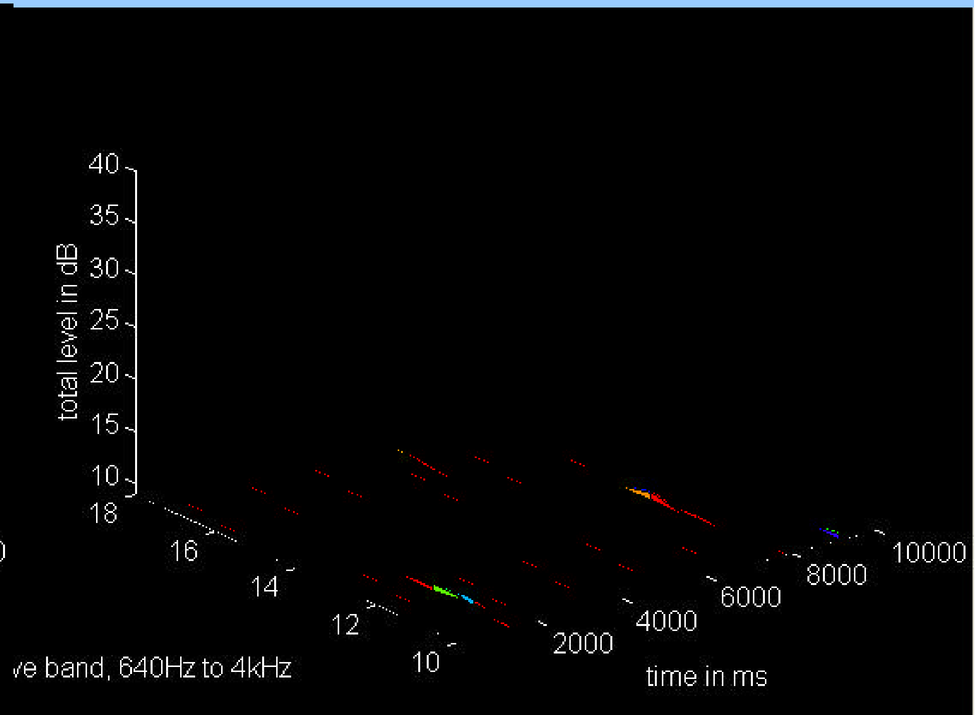
Yellow = edge detection

Analysis of binaural speech

- We can then plot the syllable onsets as a function of frequency and time, and count them.



Reverberation forward



Reverberation backwards

Note many syllables are detected (~30)

Notice hardly ANY are detected (~2)

RASTI will give an identical value for both cases!!

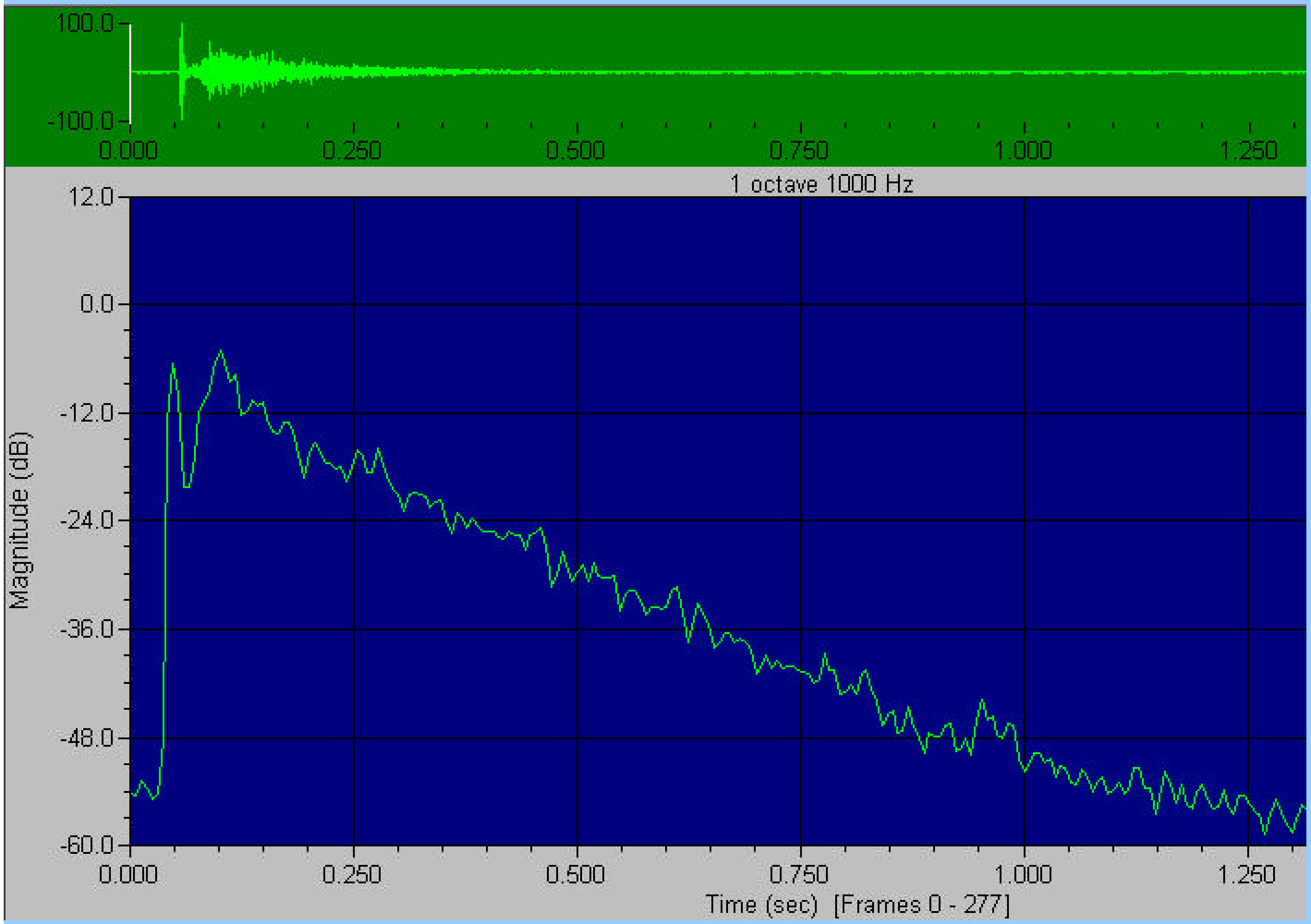
We also perceive *distance* and *space*

- We perceive fluctuations in the level **during** a sound event and 50 to 150ms after as a sense of **distance** and as a sense of space**.**
  - These fluctuations are usually caused by interference from reflected energy.
- If the reflections are spatially diffuse (from all directions) the fluctuations will be different in each ear.
  - Fluctuations that occur during the sound event and within 50ms after the end of the event produce both a sense of distance and the perception of a **space around the source.**
    - This is Early Spatial Impression (ESI)
    - The listener is outside the space – and the sound is not enveloping
    - But the sense of distance is natural and pleasant.
  - Spatially diffuse reflections later than 50ms after the direct sound produce a sense of **space around the listener.**
    - This can be perceived as envelopment. (Umgebung)

Distance Perception

- Reflections during the sound event and up to 150ms after it ends create the perception of **distance**
- But there is a price to pay:
  - Reflections from 10-50ms do not impair intelligibility.
    - The fluctuations they produce are perceived as an acoustic “halo” or “air”around the original sound stream. (ESI)
  - Reflections from 50-150ms contribute to the perception of distance – but they degrade both timbre and intelligibility, producing the perception of sonic **MUD**.
- We will have many examples of mud in this talk!

Example of reflections in the 50-150ms range



Balloon burst in an opera house. Forestage to stalls row 10.

Note the HUGE burst of energy about 50ms after the direct sound. The 1000Hz 0ctave band shows the combined reflections to be 6dB stronger than the direct sound.

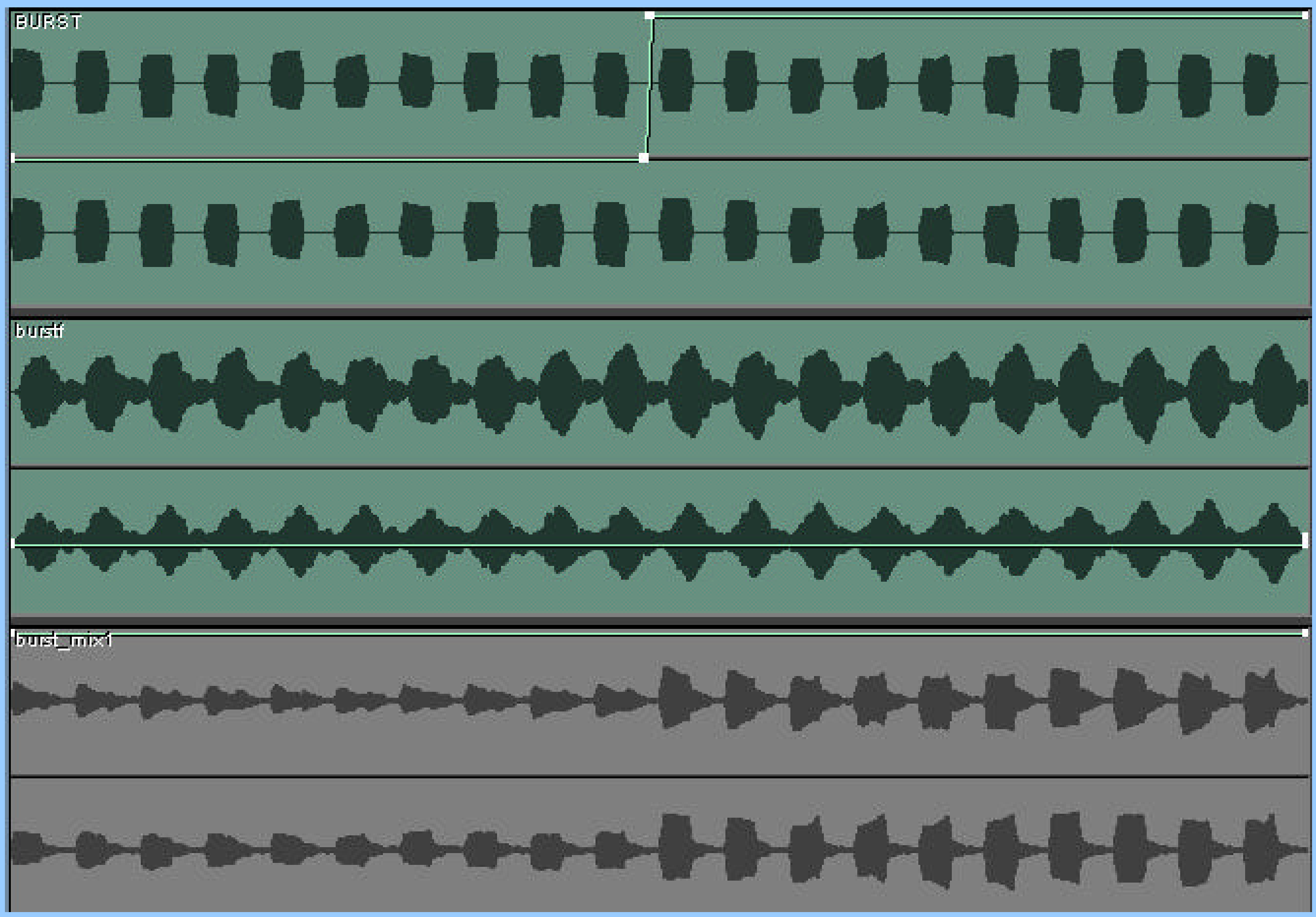
The sound clip shows the result of this impulse response on speech.

The result (in this case) is a decrease in intelligibility and an increase in distance

Human Perception – the background sound stream

- We also perceive the background sound in the spaces between the individual sound.
  - The background stream is perceived as continuous, even though it may be rapidly fluctuating.
  - The background stream is perceived at absolute level, not as a ratio to the foreground sound.
  - Perception of background is inhibited for 50ms after the end of a sound event, and reaches full sensitivity only after 150ms.

Example of foreground/background perception (as a cooledit mix)



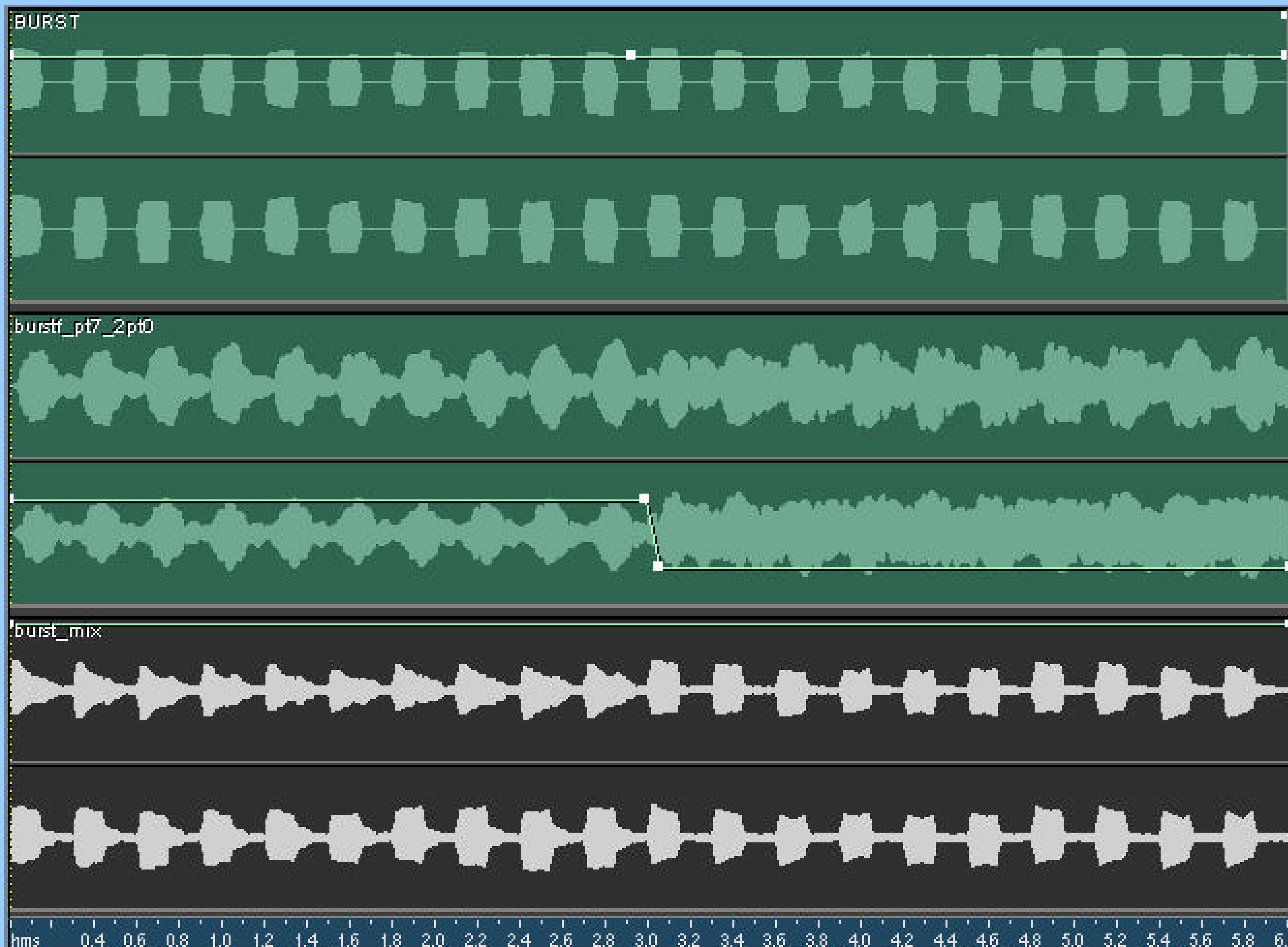
Series of tone bursts (with a slight vibrato) increasing in level by 6dB

Reverberation at constant level

Mix with direct increasing 6dB

Result: backgound tone seems continuous and at constant level

Example of background loudness as a function of Reverberation Time



Tone bursts at constant level, mixed with

reverberation switching from 0.7s RT to 2.0s RT, and reducing in level ~8dB

Output – perceived background is constant! (But the first half is perceived as farther away!)

Note the reverb level in the mix is the same at 150ms and greater. One gets the same results with speech.

Summary: Perceptions relating to stream separation

- First is the creation of the foreground stream itself. The major perception is **intelligibility**
- Second is the formation of the background sound stream from sounds which occur mostly 150ms after the direct sound ends. The perception is **reverberance**
- Third is the perception of Early Spatial Impression (ESI) from reflections arriving 10-15ms after the end of the direct sound. The perception is the presence of **distance** and **acoustic space.**
- Fourth is the timbre alteration and reduction of intelligibility due to reflections from 50 to 150ms after the end of the direct sound event. The perception is **MUD** and **distance.**

- Intelligibility, Reverberance, distance, and mud are of MAJOR importance in sound recording.
- They are also of HIGHEST importance in Opera Houses.

Binaural Examples in Opera Houses

It is very difficult to study opera acoustics, as the sound changes drastically depending on:

1. the set design,
2. the position of the singers (actors),
3. the presence of the audience, and
4. the presence of the orchestra.

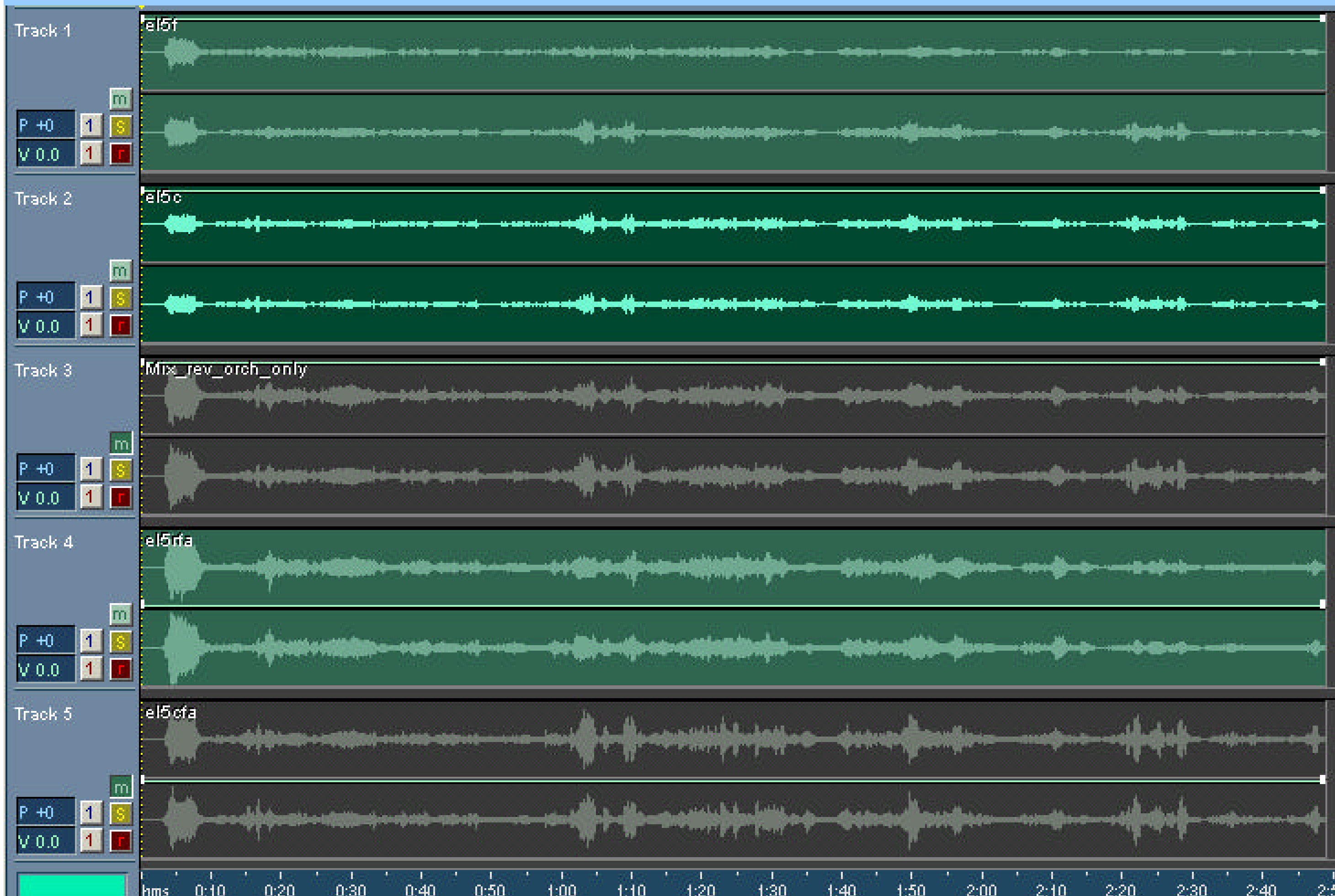
Binaural recordings made during performances give us the only clues.

Here is a sound bite from a famous German opera house: Note the excessive distance of the singers, and the low intelligibility

And here is an example from another famous German opera house: Note the increase in intelligibility and the improvement in dramatic connection between the singer and the audience.

Synthetic Opera House Study

- We can use MC12 Logic 7 to separate the orchestra from the singers on commercial recordings, and test different theories of balance and reverberation.
- From Elektra – Barenboim. Balance in original is OK by Barenboim.



Original

Orchestra Left&Right

Vocals

Downmix - No reverb on the singers

Reverb from orchestra

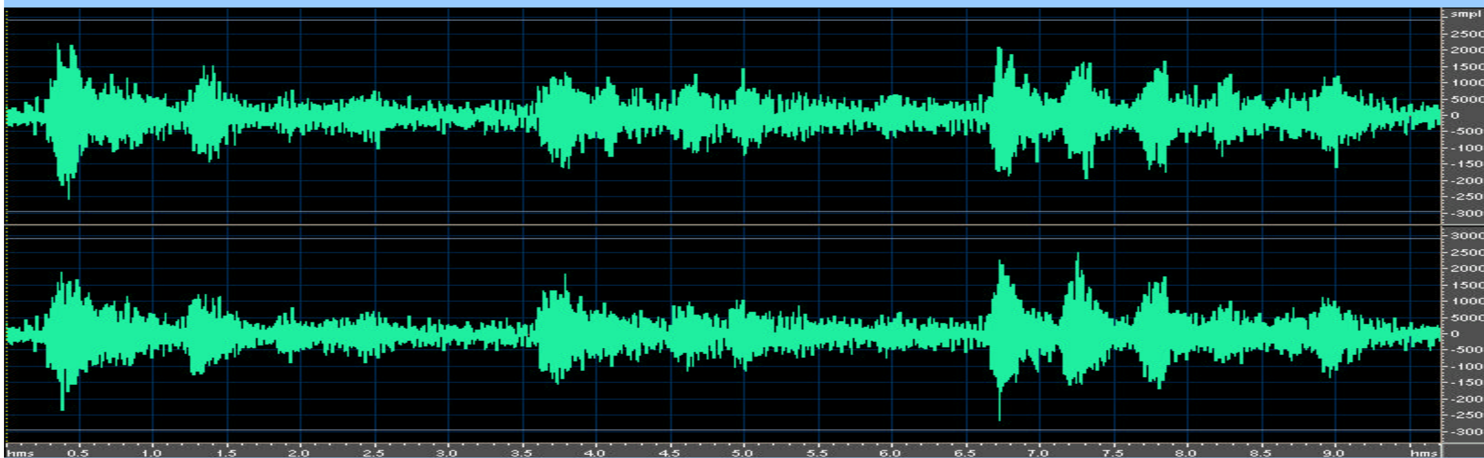
Reveb from singers

Downmix with reverb on the singers.

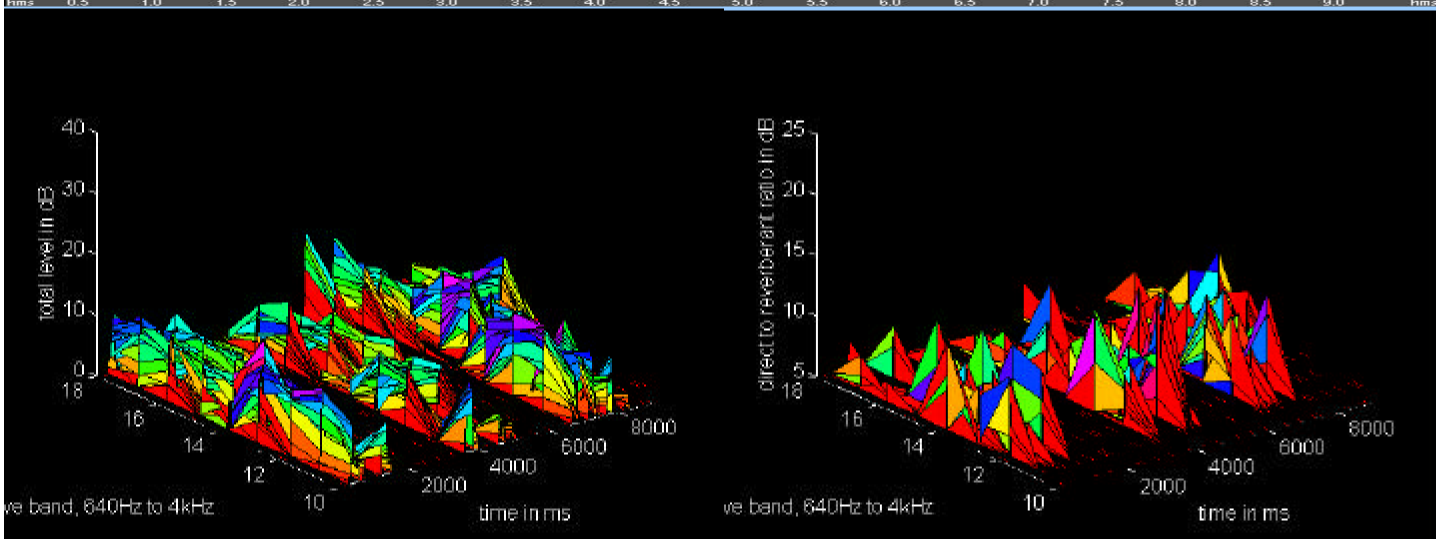
Localization

- Localization is related to stream formation. It depends strongly on the beginning of sound events.
  - IF the rise-time of the sound event is more rapid than the rise-time of the reverberation
  - Then during the rise time the IID (Interaural Intensity Difference) and the ITD (Interaural Time Difference) are unaffected by reflections.
    - We can detect the direction of the sound source during this brief interval.
    - Once detected, the brain **HOLDS** the detected direction during the reverberant part of the sound.
    - And gives up the assigned direction **very reluctantly.**
  - The conversion between IID and ITD and the perceived direction is simple in natural hearing, but complex (and unnatural) when sound is panned between two loudspeakers.
    - Sound panning only works because localization detection is both robust and resistant to change.
    - A sound panned between two loudspeakers is profoundly unnatural.

Detection of lateral direction through Interaural Cross Correlation (IACC)



Start with binaurally recorded speech from an opera house, approximately 10 meters from the live source.

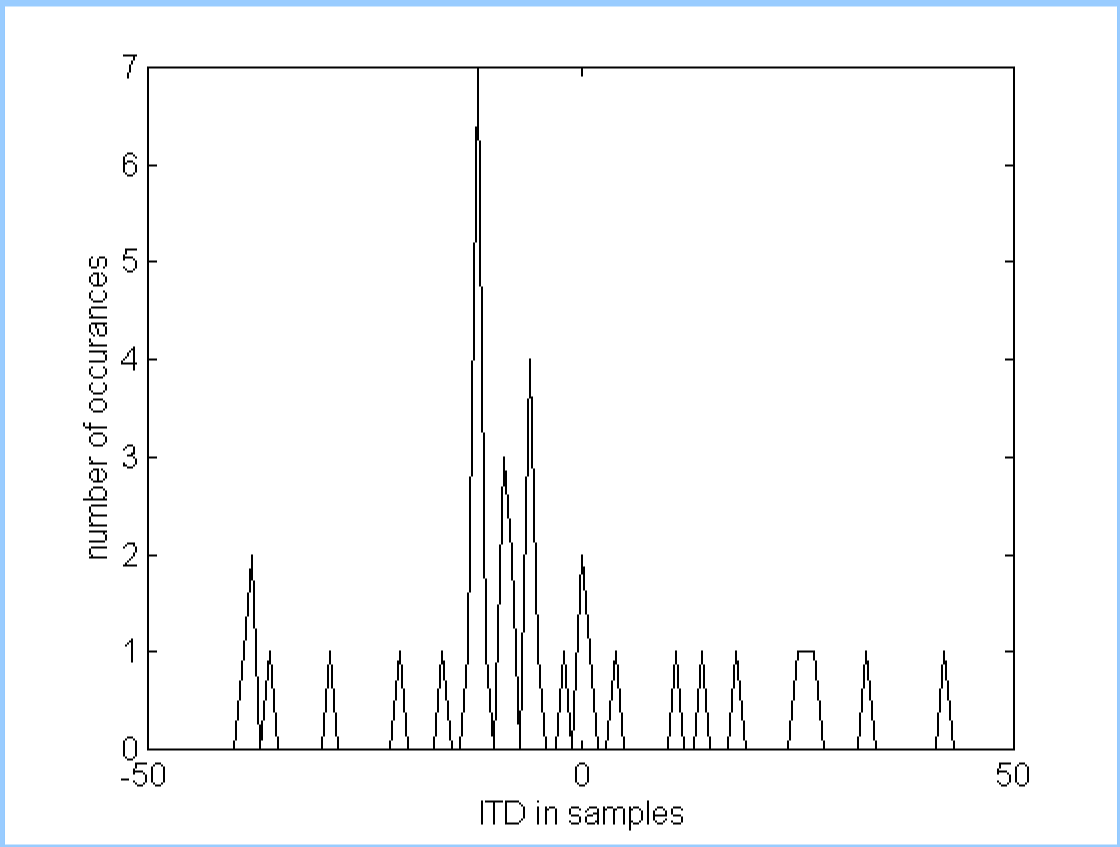


We can decompose the waveform into 1/3 octave bands and look at level and IACC as a function of frequency and time.

Level (x = time in ms y=1/3 octave bands 640Hz to 4kHz) IACC

Notice that there is NO information in the IACC below 1000Hz!

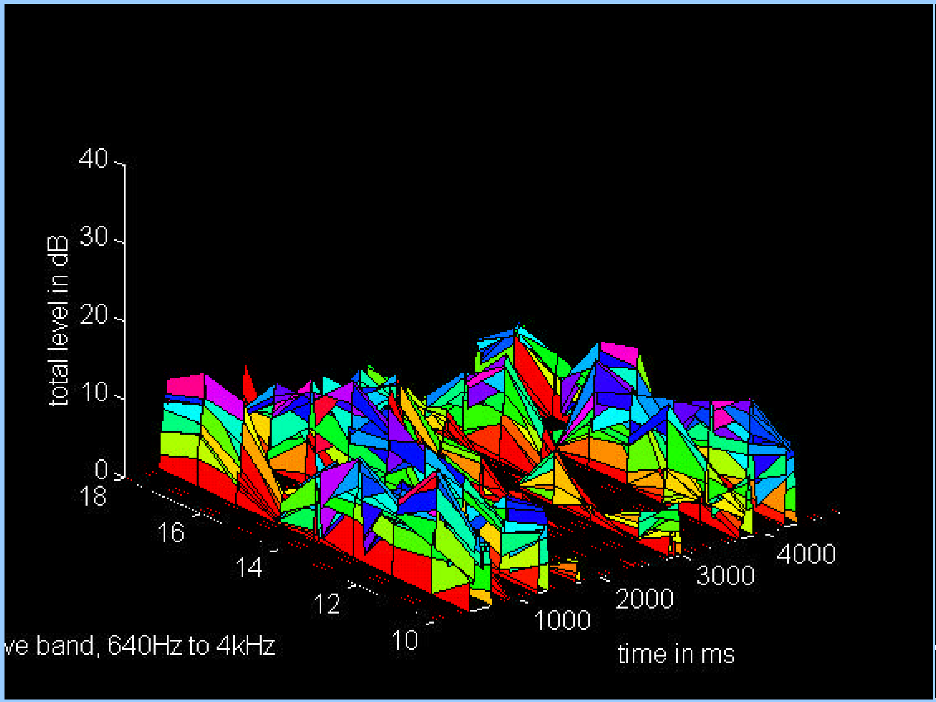
Position determination by IACC

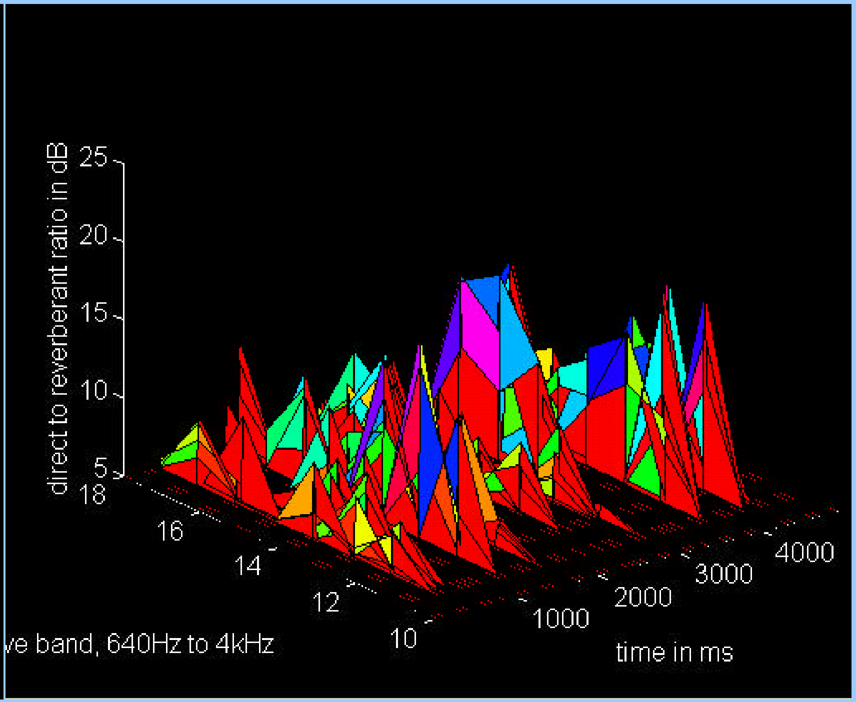


We can make a histogram of the time offset between the ears during periods of high IACC.

For the segment of natural speech in the previous slide, it is clear that localization is possible – but somewhat difficult.

Position determination by IACC 2



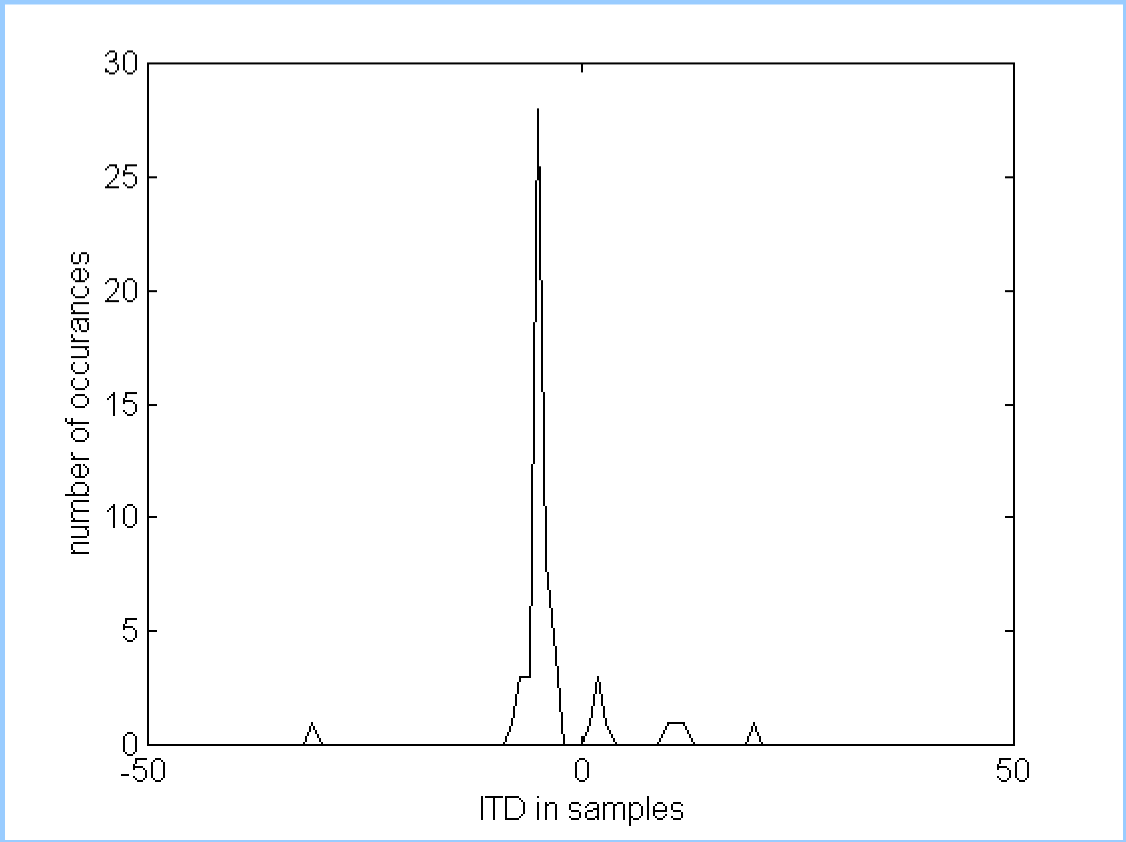


Level displayed in 1/3 octave bands (640Hz to 4kHz) IACC in 1/3 octave bands

We can duplicate the sound of the previous example by adding reverberation to dry speech, and giving it a 5 sample time offset to localize it to the right.

As can be seen in the picture, the direct sound is stronger in the simulation than in the original, and the IACCs - plotted as 10*log10(1-(1/IACC)) - are stronger.

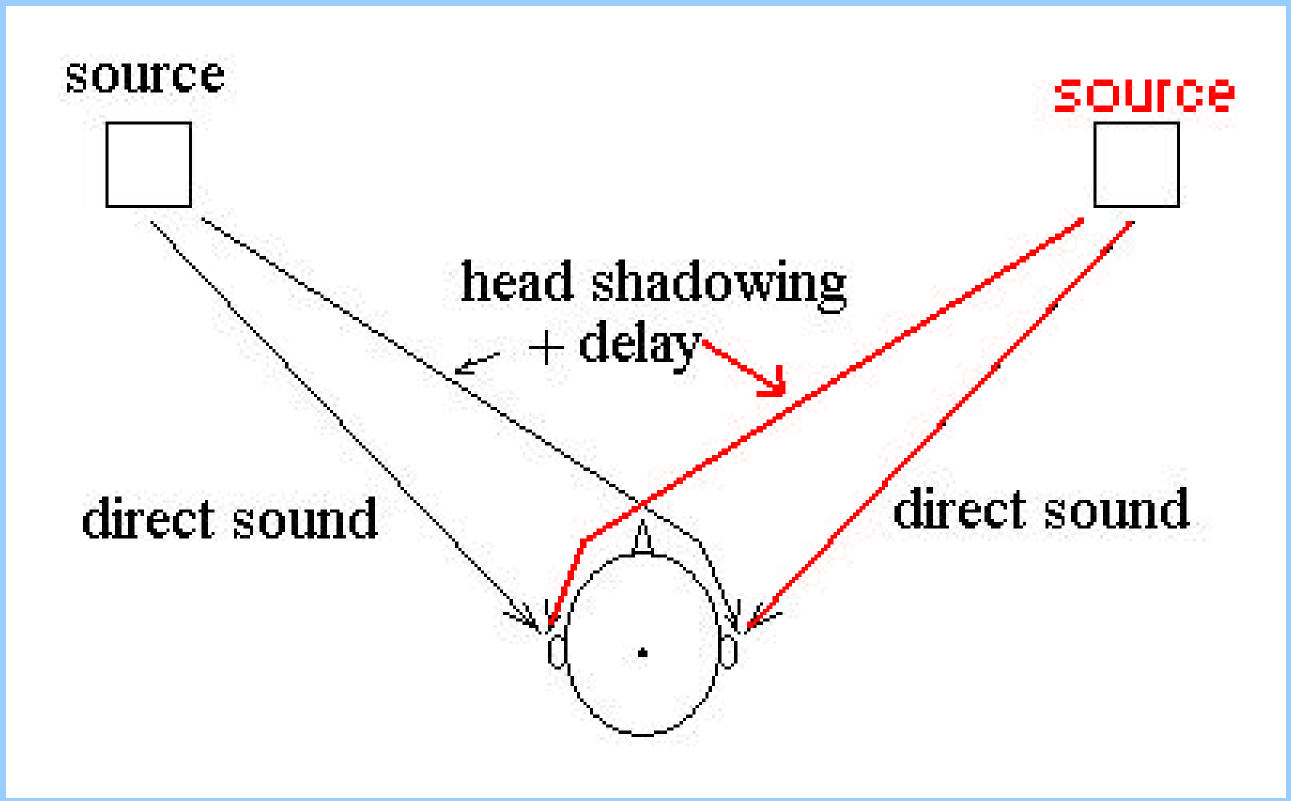
Position determination by IACC 3



Histogram of the time offset in samples for each of the IACC peaks detected, using the synthetically constructed speech signal in slide 2.

Not surprisingly, due to the higher direct sound level and the artificially stable source the lateral direction of the synthetic example is extremely clear and sharply defined.

The physics of two-channel panning



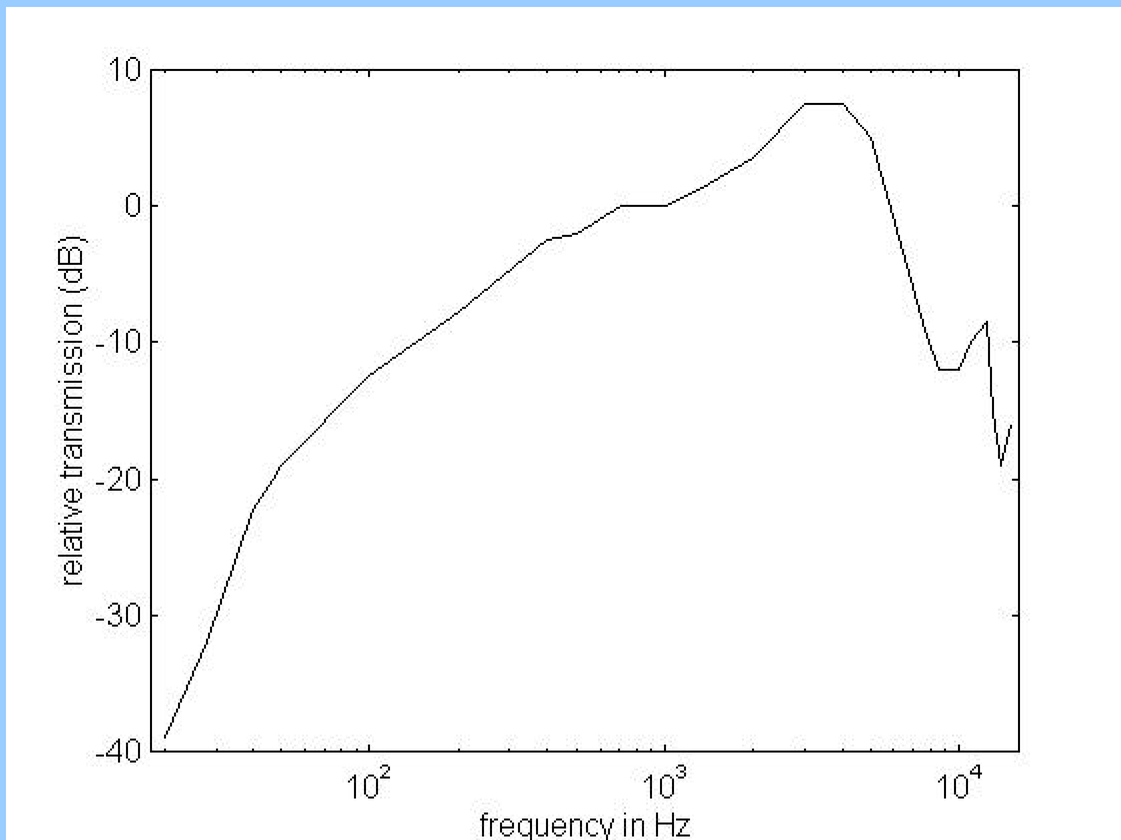
The pressure at each ear is the sum of the direct sound pressure from one speaker and the diffracted sound pressure from the other.

These two signals interfere with each other, producing a highly frequency dependent signal.

Consequences of panning physics

- A two channel pan is entirely different from the localization process in natural hearing.
  - Localization always depends on the interaural time delay (ITD) and the interaural intensity difference (IID).
  - In natural hearing the ITD and IID vary due to head shadowing alone.
    - Between 500Hz and 1500Hz the ear relies increasingly on IID rather than ITD, and the precise phase of the ear signals becomes inaudible.
  - In a two channel pan, ITD and IID vary due to INTERFERENCE.
    - The interference is entirely PHYSICAL. It occurs in the air at the entrance to the ear canals, and it happens at all frequencies, even HF.
    - The interference can be measured, and it can be calculated.
  - The result of the interference is that the ITD and IID become highly frequency dependent.
    - For broadband sources the brain must make a “best guess” about the true azimuth of the source.
    - The localization of narrow band souces can be bizarre.

The frequency transmission of the pinnae and middle ear



From: B. C. J. Moore, B. R. Glasberg and T. Baer, “A model for the prediction of thresholds, loudness and partial loudness,” *J. Audio Eng. Soc.*, vol. 45, pp. 224-240 (1997).

The intensity of nerve firings is concentrated in the frequency range of human speech signals, about 700Hz to 4kHz. With a broad-band source, the ITD and IID at these frequencies will dominate the apparent direction.

The influence of expectation, visual position, and past history on current position

- We discovered that the apparent position of a sound source is highly influenced by its visual position, expected position and its past history.
  - Localization in a natural environment is almost always dominated by the visual field, or a memory of a visual field.
  - Expectation of azimuth and elevation is particularly important in sound recording.
- In panning experiments we can alter the bandwidth or the band frequency of a known source type, like speech. This change of frequency mix will drastically alter the IDT and IID.
  - And yet a source which appears to be located at a particular position with one mix of frequencies will remain in that position when the frequency mix is changed.
- Alternating the presentation from left to right by switching the speaker channels breaks this hysterisis.
  - The subject is asked to estimate the width between sound images which alternate left and right,
  - rather than the position of images presented consistently on one side or the other.

Apparent width of broadband syllabic sources

- Broadband syllabic sources are consistently perceived as narrow in width.
  - Although when they are broken up into 1/3 octave or critical bands the apparent direction of sound may vary over a wide range.
- The neurological process of separating sound into streams assigns a “best guess” position to the entire stream, rather than separating the perception into several streams in different directions.
  - Once the direction of a stream has been assigned, the hearing apparatus is quite reluctant to change it.
- ESI – which is often present around syllabic sources, is most often described as “source width” by untrained listeners.
  - On careful listening the source will be found to be narrow, but surrounded by a local acoustic halo.

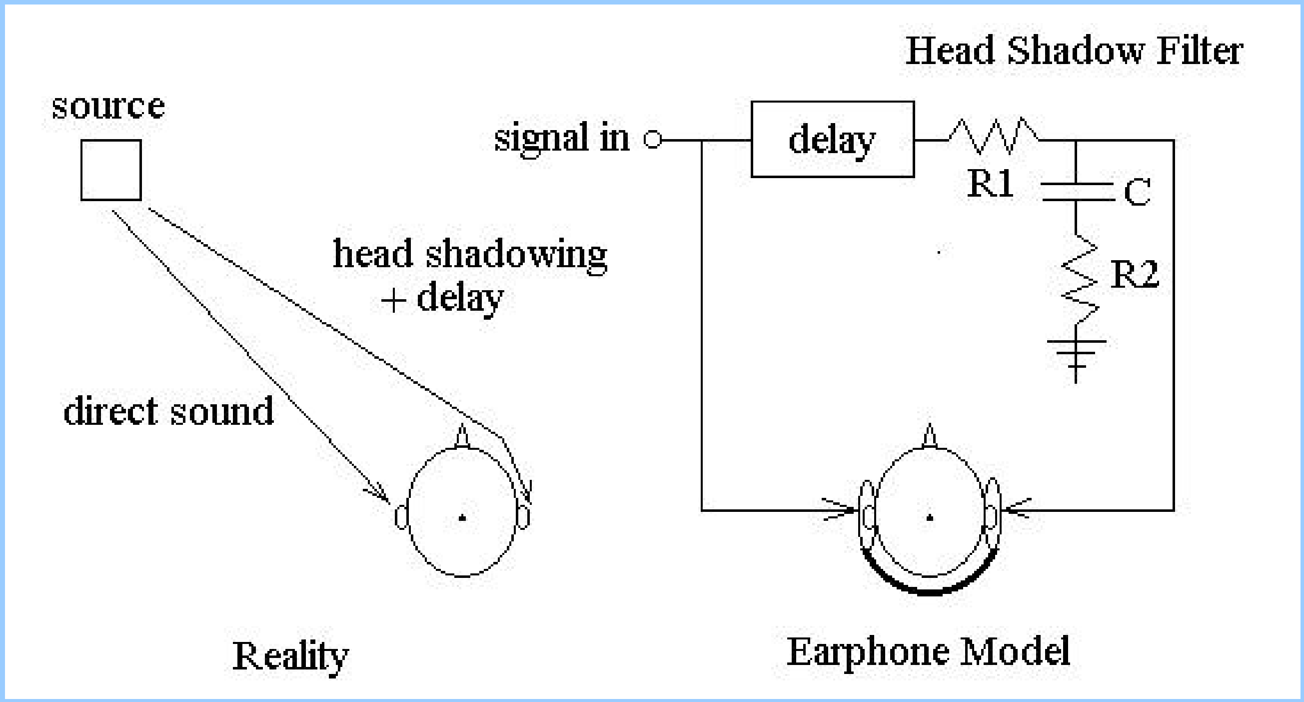
Apparent source width (ASW)

- ASW is consistently cited in the literature as important in musical acoustics.
- But human physiology insists that the apparent width of all *syllabic* sources – speech, music with clear notes, etc – will be sharp.
- The apparent width of *continuous* sources – legato string sections, pink noise, etc. can increase when ITD and IID become inconsistent or unstable.
  - ASW may be a useful concept for such sources, but these sources are not very interesting acoustically.

Apparent width of broadband continuous sources

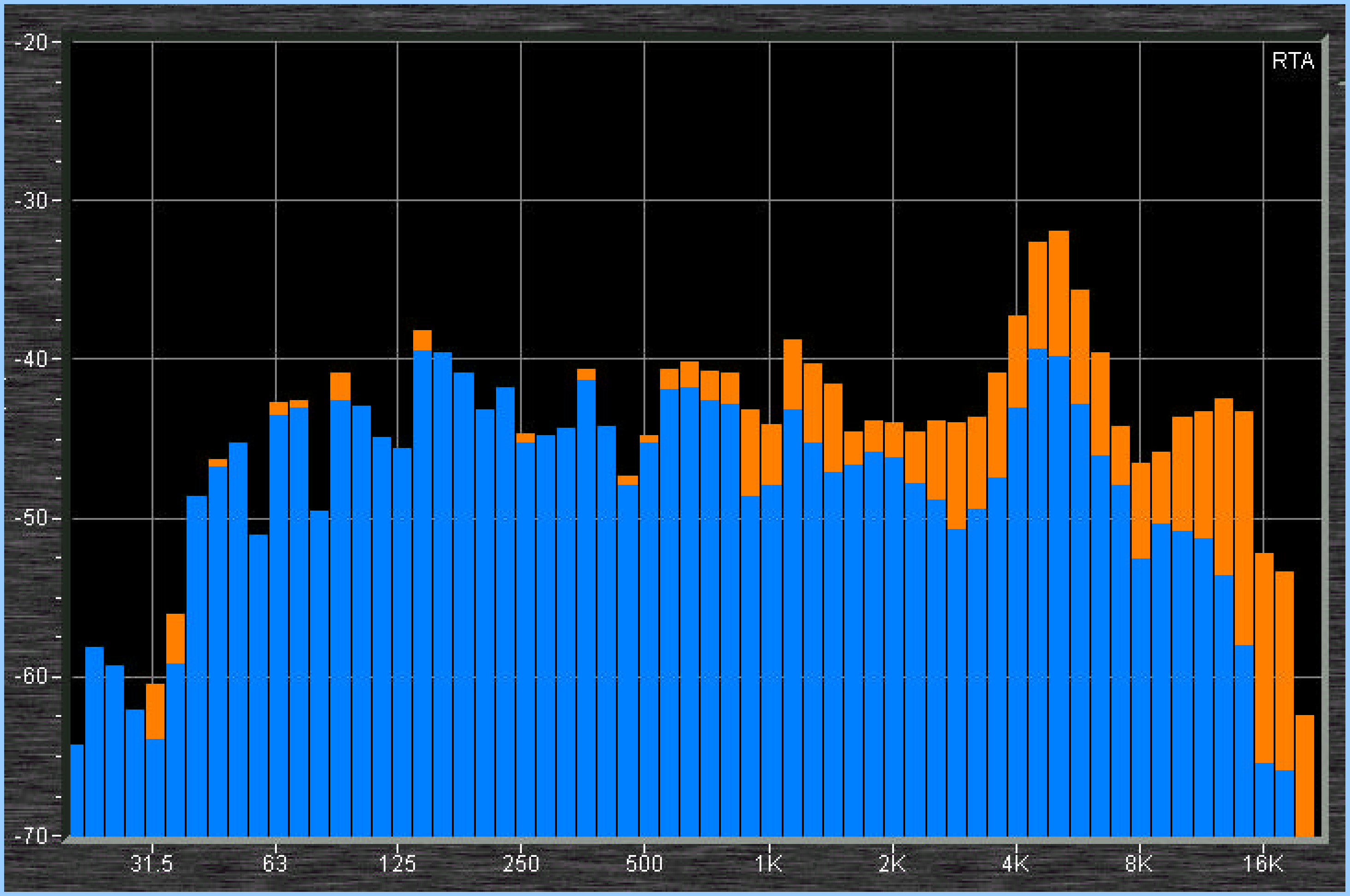
- Reflection can increase the apparent width of continuous sources.
- Such sources are often physically wide in any case.
  - Examples might include an orchestral string section or a chorus.
- In this case fluctuating or inconsistent ITD and IID cause the apparent width of the source to broaden.
  - In acoustics, a common experimental model consists of a loudspeaker which plays pink noise or legato strings.
  - This example (and the concept of source width itself) is not very useful.

Experimental Models



Diffraction around the head can be modeled with a combination of delay and a shelving filter. Typical filter values can be determined from dummy-head data, with a –3dB point of 1500Hz, and a shelf depth of 6dB.

Dummy Head measurement

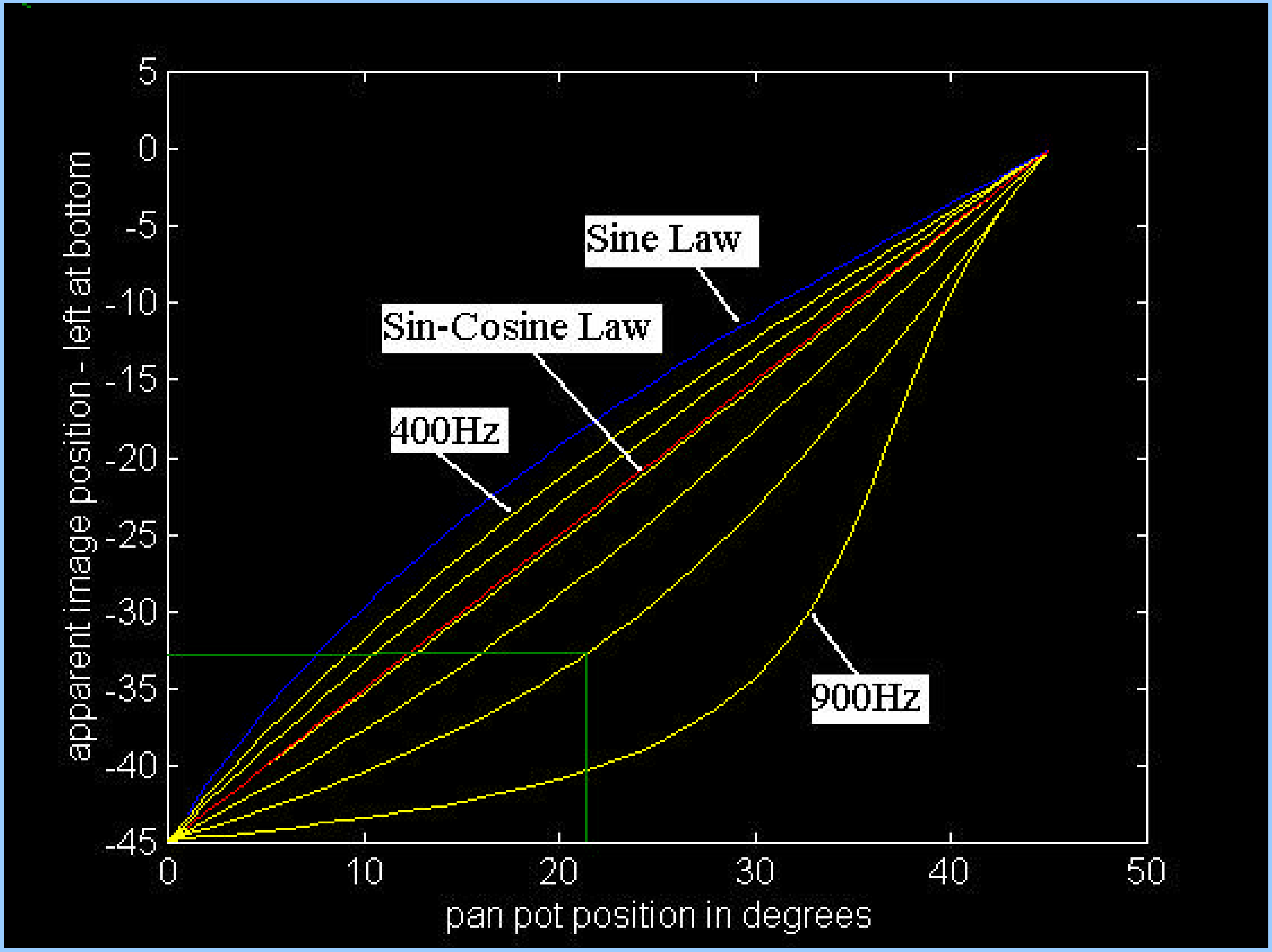


Modified Neumann KU81 head. Sound source is front (0 degrees) for the orange curve. Source angle is 45 degrees right for the blue curve.

Using the model

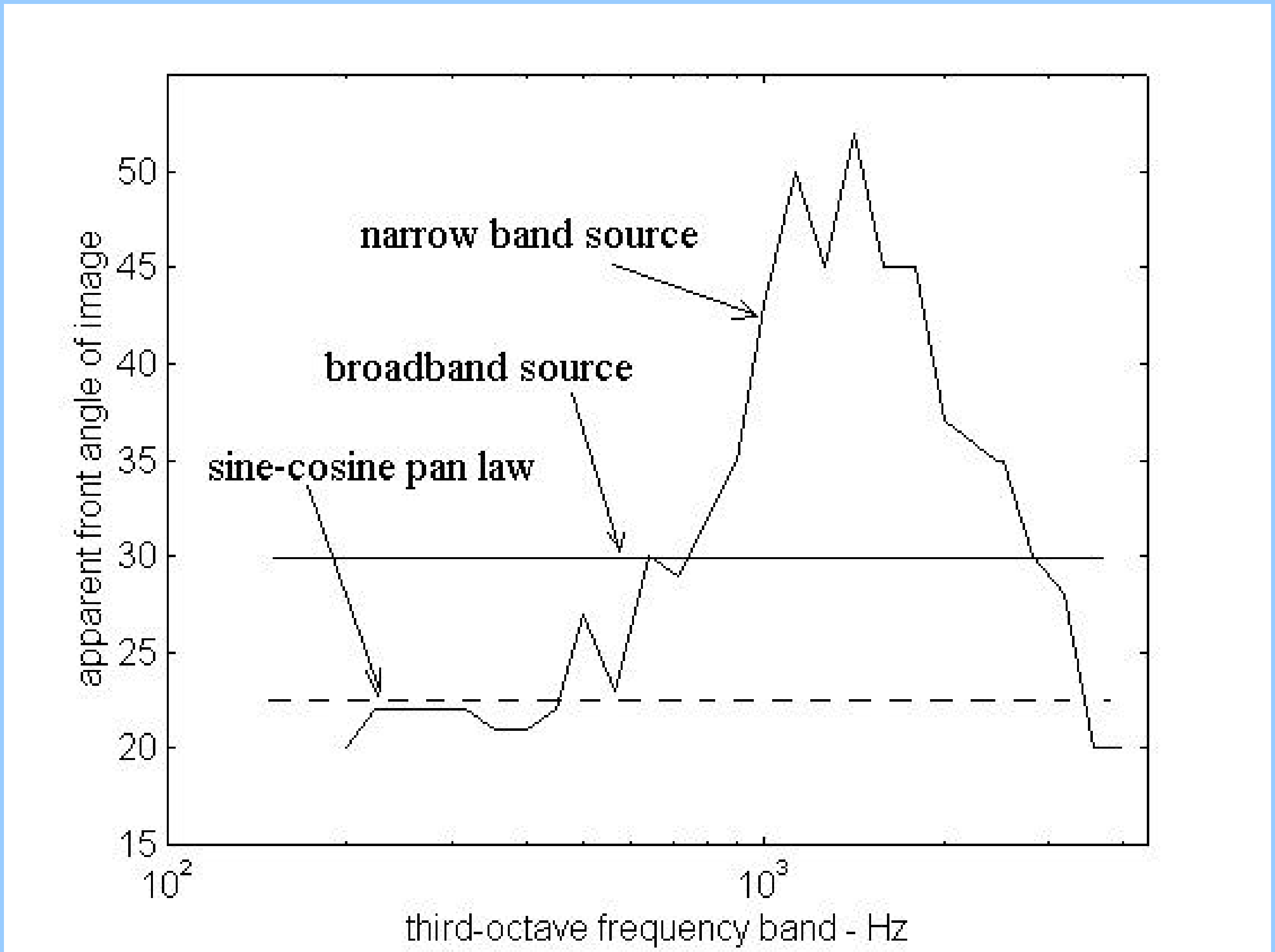
- Because direction determination is strongly weighted by the pinnae and middle ear, it is important to duplicate the frequency response at the ear drum.
  - The model assumes a flat frequency transfer between the model outputs and the eardrum of the listener. The HRTF (head related transfer function) from the ear drum to the left (or right) speaker position is not included.
  - This means the earphones in the experiment must be matched to the HRTF's of the individual listener.
  - The best way to do this is by putting a probe microphone on the eardrum, and matching the earphone response to the measured response of a loudspeaker at the listening position.
- Results for poorly matched HRTFs will not correspond to natural hearing.
  - But will preserve the relative positions of different sources and are still quite useful.
- The model allows us to analyze the interference at the entrance to the ear canals, and relate what we find to the apparent source direction.

Results from the model – using ITD only



Results assuming an interaural spacing of 21cm: The apparent position of low frequencies matches the sine law. The sine/cosine law is accurate at 600Hz, and above 600Hz the sound image moves rapidly to the left.

Listening experiment results at High Frequencies (ILD + ITD)



Apparent position of 1/3 octave filtered speech as a function of frequency.

Although the original signal is panned to a position of 22 degrees, at 1700Hz the image is outside the loudspeaker basis, (+-45 degrees.)

Note that the position of the broadband source is strongly pulled outward by the increased width of the frequencies above 1kHz.

Further observations:

- The model allows us to generate signals that combine different IDTs and ILDs. Matlab can be used to remove either the delay component or the amplitude difference between the two ear signals.
  - The resulting signals can then be tested for perception with earphones.
- We find that, as expected, above 1kHz the direction determination is dominated by the amplitude difference (ILD), but that the ITD of the <u>envelope</u> of the signal still contributes to the apparent position.
- A signal that varies in ITD only is perceived as closer to the center than a signal that varies in both ITD and ILD.
  - And is less reliably localized.

Conclusions on two channel panning

- The apparent position of **narrow band** speech or syllabic music sources in a typical two channel pan is HIGHLY dependent on frequency
  - A source panned half way between center and left with a conventional pan pot can appear slightly left of center to full left, depending on the selection of 1/3 octave frequency band.
  - At frequencies above 700Hz the apparent position of these sources is usually further from the center than would be expected from the sine/cosine law.
- Images from **broadband** speech and syllabic music sources are consistently perceived as further left or right from the center than would be predicted by a sine/cosine pan law.
  - A more accurate pan law can be made by simply expanding the scale on the pan control, so half-left or half-right is closer to center.
- The discrepancy can be explained by the dominance of frequencies between 700Hz and 4kHz in human hearing.
  - These frequencies are consistently perceived as wider than low frequencies in a pan which is half-way between center and either right or left.

Conclusions, continued

- The hearing mechanism appears to simply weight the apparent position of each frequency band by the intensity of nerve firings in that band when assigning an azimuth to a particular sound stream.
  - Thus frequencies above 500Hz are particularly important
- Syllabic sound sources are perceived as narrow in lateral width regardless of the positional discrepancy between critical bands.
  - This observation seems to be a universal property of human perception.
    - Fluctuation in IID or ITD causes an acoustic impression around the source, not widening of the source itself.
  - Continuous sources, such as pink noise or legato string sound, can be broadened by positional discrepancy between bands, or rapidly fluctuating IID and ITD.
- The expected position of a sound stream, and the past history of its position, have a strong influence on perception.
  - A sound source is usually prohibited from leaving its visual position.
  - A sound source is highly resistant to moving to a new position once its original position has been established.

Conclusions on Panning for microphone technique

- Sound localization with amplitude panning is NOT NATURAL
  - It is strongly frequency dependent.
  - Broad band sources appear narrow, but are localized with difficulty.
- Sound localization with time delay panning is even LESS NATURAL
  - Time delay panning is useful only in the sweet spot, and this restriction is not commercially viable.
  - Low frequencies are always in the middle of the head.
  - High frequencies also frequency dependent.

So – how can we evaluate our recordings?

- Recording standards are NOT arbitrary.
- Experience judging student recordings shows:
  - Everyone on the jury panel agrees on the ranking of the submitted recordings.
  - When asked to explain their reasons everyone agrees on the comments.
  - The juries do NOT agree on the best method to achieve good results. In fact, with regard to microphone type and position they can be quite closed-minded.
  - But when they do not know how the recordings were made, they agree on their quality. This is very refreshing.
- Recordings stand or fall on their ability to satisfy the needs of human hearing. This is constant among people, even among recording engineers.

Main Message:

- The recording venue is CRITICAL!!!
  - Large (>2000 seat) concert halls can make stunningly beautiful recordings.
    - A wide variety of techniques can achieve satisfactory results.
  - A technique that works well in a large concert hall will probably NOT work well in a hall with 1200 seats.
- It is the job of the engineer to make a stunningly beautiful recordings in the hall that happens to be available.
  - Working to a world-class standard in a small space takes both science and art.

Spatial Properties of Recordings

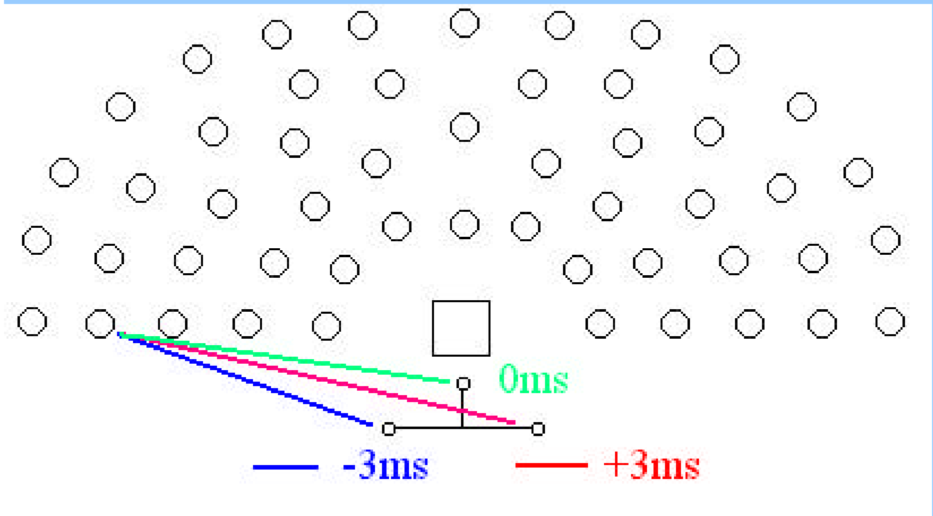
- All engineers know how to judge instrumental balance and tonal balance
  - So we will not talk about these perceptions.
- We are going to talk exclusively about
  - Intelligibility
    - not really a problem in recordings
  - Distance
  - Localization
  - Envelopment
- How can you learn to hear and to assess these perceptions?
- How can you achieve the best results?

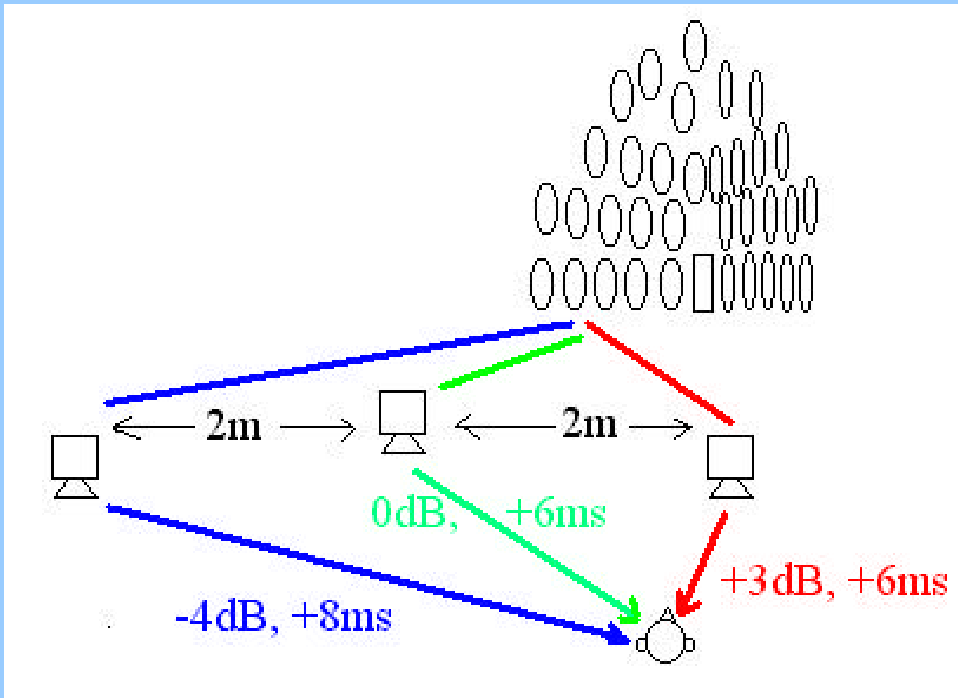
Training to hear localization

- The importance of ignoring the sweet spot
  - Most research tests of localization use a single listener, who is strictly restricted to the sweet spot.
  - Your customers will not listen this way!
    - And neither will the jury for your student submission. There are always at lest three jurors, and they move around.
  - A recording that only localizes well in the sweet spot will not make it past the first round of judging!
- How do you know if the recording will pass this test?
  - Move laterally in front of the loudspeakers. Does the sound image stay wide and fixed to the loudspeakers, or does it follow you?
  - Do the soloists in the center follow you left or right? If they do they are recorded with too much phantom center.
- Since most 5 channel recording methods are derived from stereo techniques almost all have too much phantom center.
- **A center image that follows a listener who moves laterally out of the sweet spot is the most common failing of even the best five channel recordings.**

» **Play example**

Example: Time delay panning outside the sweet spot.

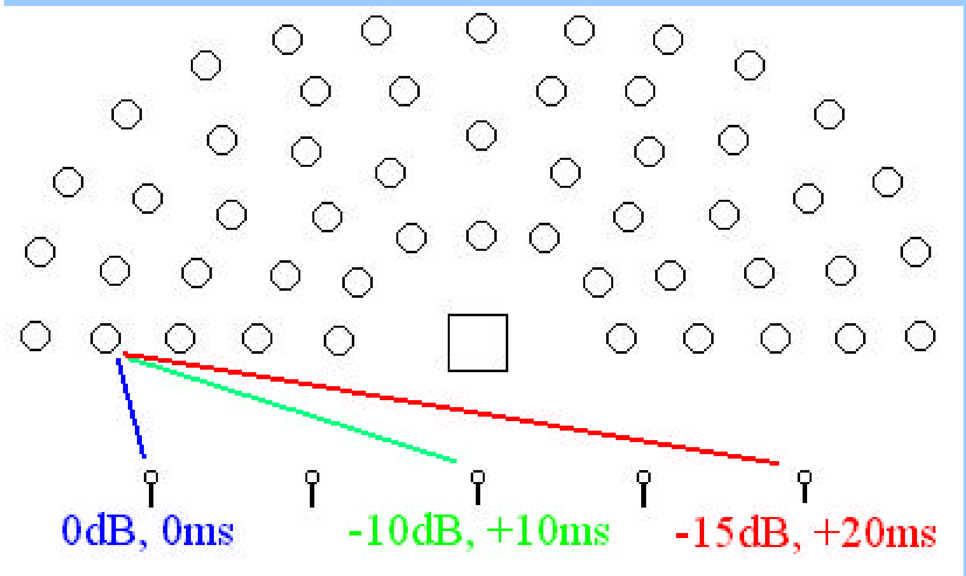


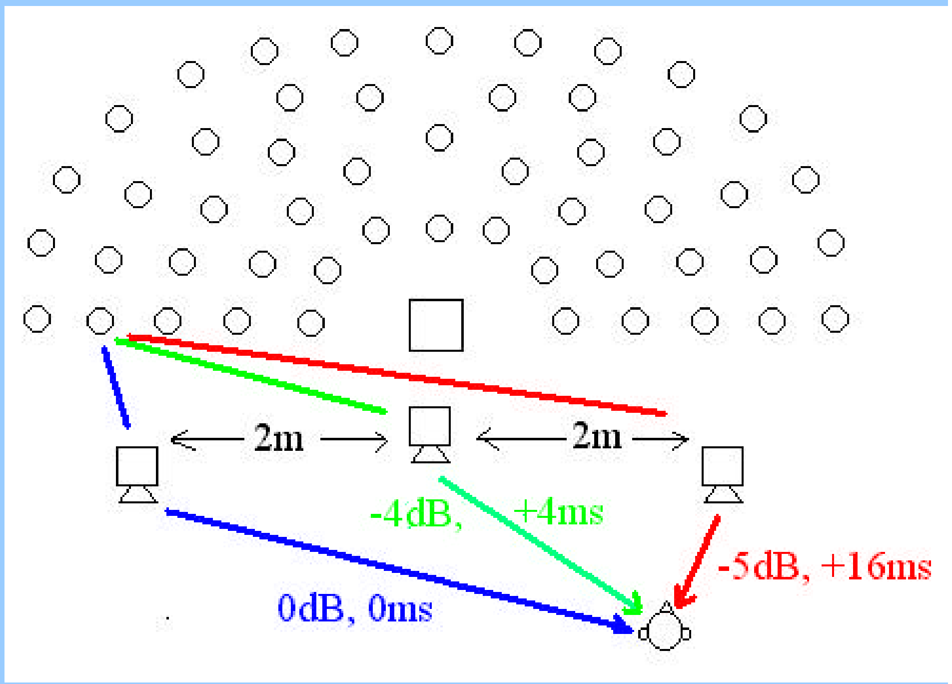


Record the orchestra with a "Decca Tree" - three omni microphones separated by one meter. A source on the left will give three outputs identical in level and differing by time delay.

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, even when the listener is not in the sweet spot.

Training to hear distance

- Closely miked sources often sound in front of the loudspeaker.
- They seem unnaturally forward and dry.
- Adding diffuse early reflections through all four lateral speakers puts them behind the loudspeakers and in a unified acoustical space of no perceivable size or shape.

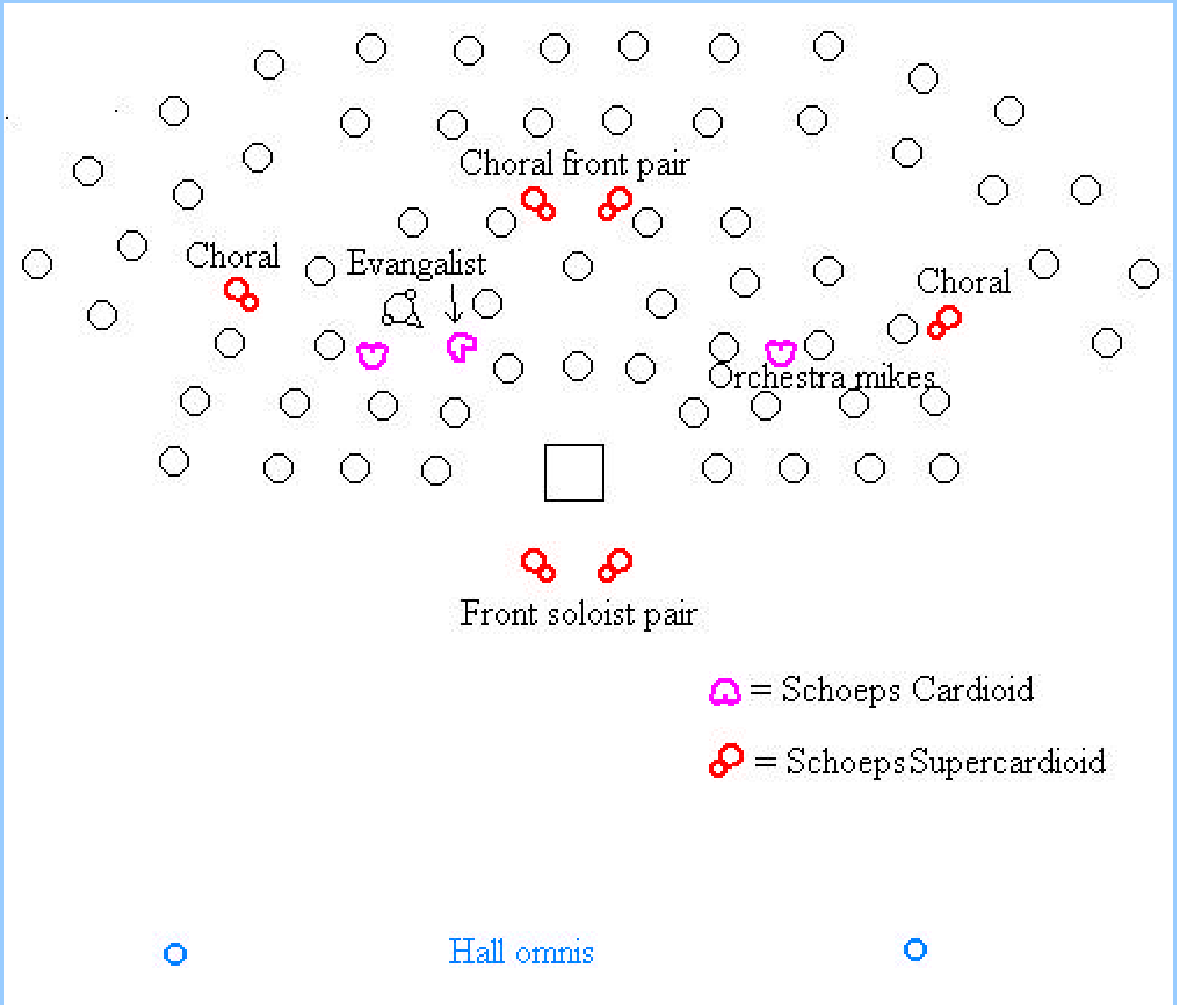
    » Play examples

Boston Cantata Singers in Jordan Hall

Major Characteristics

- Chorus is deep in an enclosing stage-house with significant reverberation.
- Small distances between microphones results in unwanted leakage.
- Microphones pointed into the stage house increase the amount of undesirable reverberation.
  - Thus the chorus mikes, which must face the chorus, are supercardiod to minimize reverberation pick-up.
  - And the orchestra mikes face the hall, not the stage house.
- Microphones in front do not pick up enough direct sound from the chorus to supply the sense of distance without also getting considerable mud.

Jordan Hall Setup



Solutions

- Add distance to the chorus at the mixing stage with controlled early reflections
- Minimize stage-house pickup wherever possible

Audio Demos

- Early reflections
- Late reverberation

Training to hear MUD

- It is relatively easy to train yourself to hear mud, but it is often very hard to avoid it.
- Mud occurs when the reverberant decay of the recording venue has too much reflected energy in the 50-150ms region of the decay curve.
  - This is true of nearly all sound stages, small auditoria, and churches.
- If you are recording in such a space with a relatively large ensemble, you are in trouble.

Example: John Eargle at Skywalker ranch

- John Eargle made a wonderful 5.1 channel DVD audio recording at the Skywalker ranch in Los Angeles.
- Skywalker is a large sound stage with controllable acoustics. It is not a concert hall.
- As a consequence the reverberation radius is relatively short. By my estimate (without having seen it) the radius is about 3.5 meters.
- It is very easy to record mud in such a space.
  - Many instruments are beyond the reverb radius.
  - Adding more microphones only increases the reverberant pickup.

Example: Revels Chorus in the Sonic Temple

Characteristics

- Main problem here was excessive reverberation level.
  - Solution was to add blankets – a LOT of them. 648ft^2
  - Here we list the measured reverberation times

Hz	blankets	empty
8000	0.6	0.9
4000	0.8	1.2
2000	0.9	1.4
1000	0.9	1.4
500	1.0	1.3
250	0.9	1.3
125	1.1	1.4
63	1.0	1.5

  - Reverb radius before the blankets: ~6 feet (2 meters)
  - Reverb radius after the blankets: ~8 feet (2.7 meters)

After the blankets

- Reverberation time drops below 1 second, the magic number for the early decay in Boston Symphony Hall
  - Recording the band is easy, as we can mike them all quite closely.
  - Recording the chorus is hard, as there are >20 singers, and we cannot get the microphones close enough to each.
    - Adding more microphones simply results in picking up more reverberation!
  - With the blankets we can record with adequate clarity using only four supercardioid microphones.
    - Once again we augment the early reflections in all outer channels using the Lexicon.
    - Late reverberation is also created using Lexicon late reverberation.

Reverberation Radius

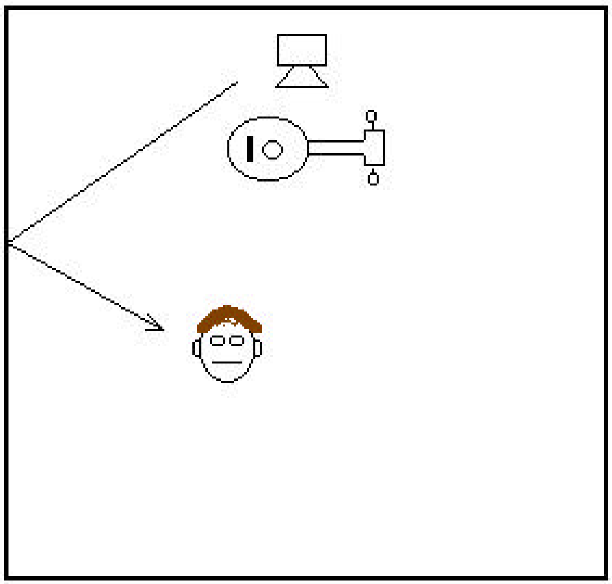
- The reverberation radius changed from 6’ to 8’ when we added blankets. ~2dB.
  - This is not a large enough change to account for the perceived difference in sound.
- But the change in the total reflected energy in the time range of 50-150ms (the undesirable time range) is much larger: 4.5dB.
  - This is a highly significant and desirable decrease!
- The decrease in the late reverberation (150ms and greater) is 6dB.
  - But we make this back up with the Lexicon.

Training to hear envelopment

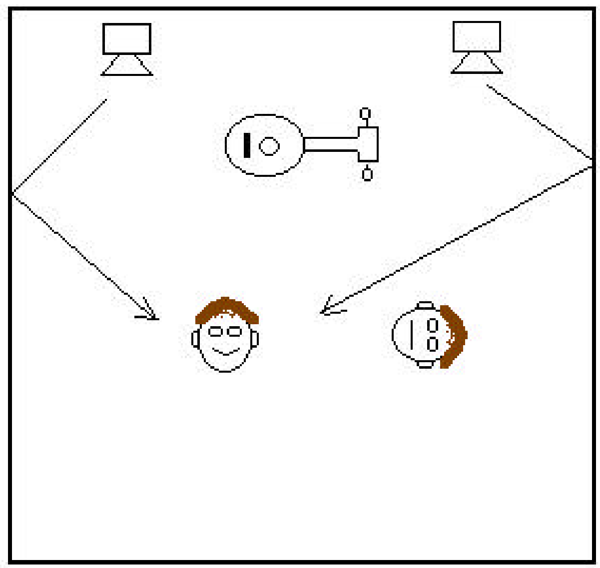
- Here it is essential that you move around the room, and that you face different directions!
- You must fill the WHOLE room with the sound of the original recording, and it must work when you face all directions.
- Reproducing space only in front, or only in the rear, will not get you the prize.

3/0 versus 3/2

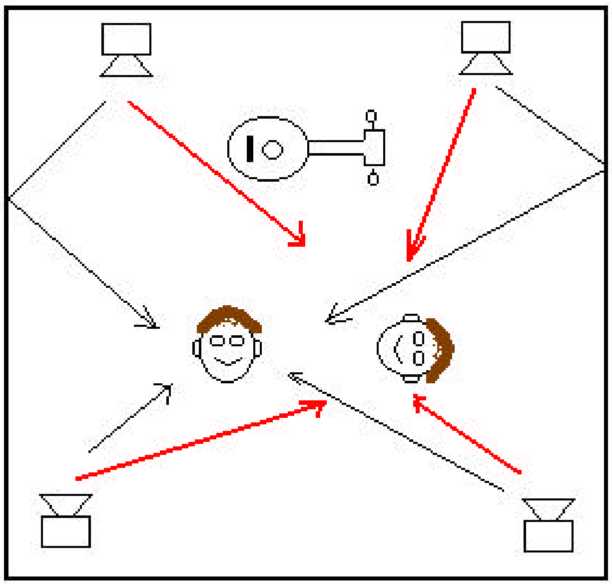
- OK, perhaps we need three speakers in the front, and amplitude panning in the front.
- Why do we need two additional speakers and channels?



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. And the sense of space is stronger in the front.



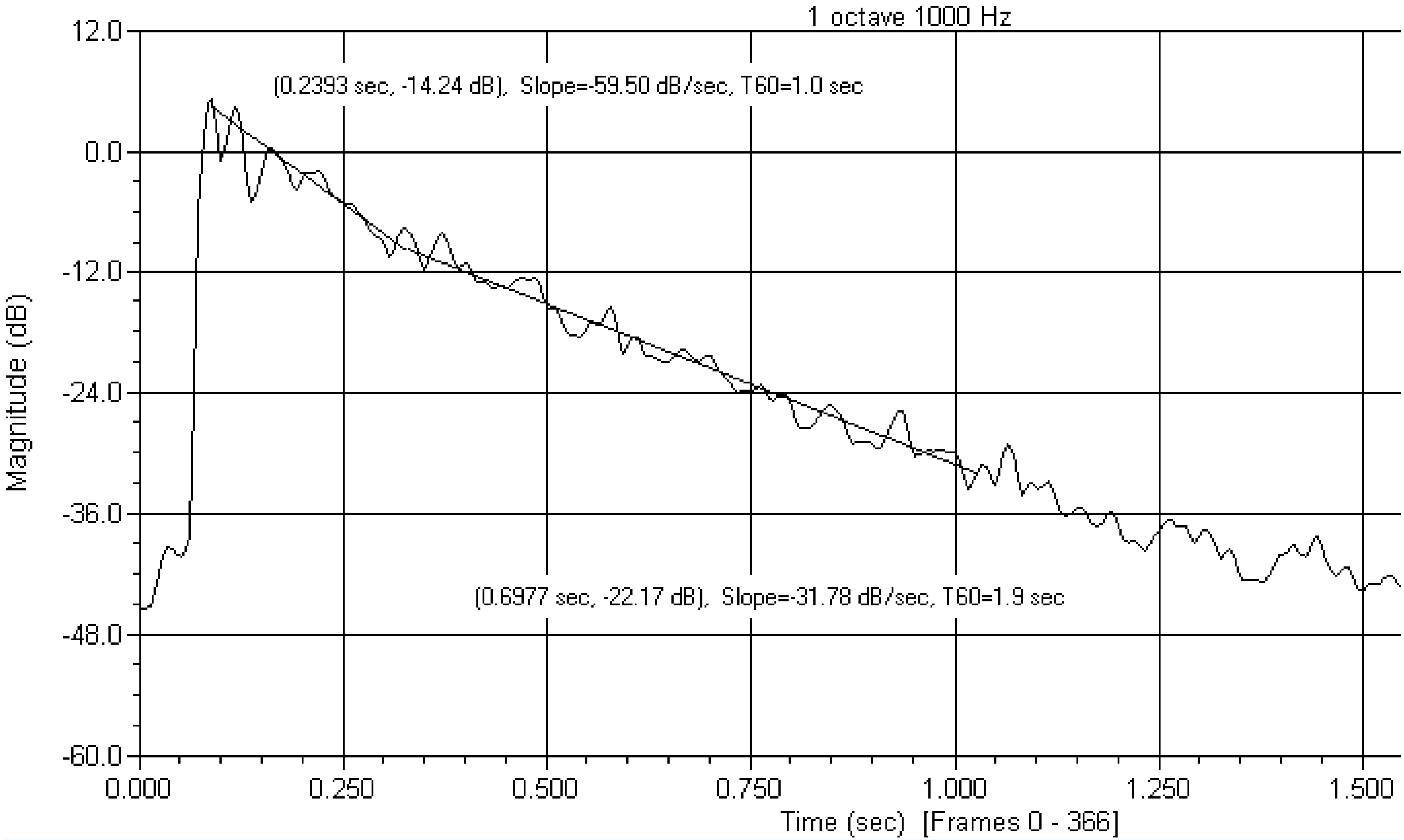
We need at least four speakers to reproduce a two dimensional spatial sensation that is uniform through the room.

Boston Symphony Hall

Boston Symphony Hall

- 2631 seats, 662,000ft^3, 18700m^3, RT 1.9s
  - It's enormous!
  - One of the greatest concert halls in the world – maybe the best.
  - Recording here is almost too easy!
  - Working here is a rare privilege
    - Sufficiently rare I do not do it. (It's a union shop.)
  - The recording in this talk is courtesy of Alan McClellan of WGBH Boston. (Mixed from 16 tracks by the presenter)
  - Reverb Radius is >20' (>6.6m) even on stage.
  - The stage house is enormous. With the orchestra in place, stage house RT ~1 sec

Boston Symphony Hall, occupied, stage to front of balcony, 1000Hz



Why is the impulse response relevant?

- Because the early decay (from the stage) is short enough to get out of the way before it muddies he sound.
- And the late decay (from the hall) is long enough to provide envelopment.

Boston Symphony Orchestra in Symphony Hall

Boston Cantata Singers in Symphony Hall.   March 17, 2002

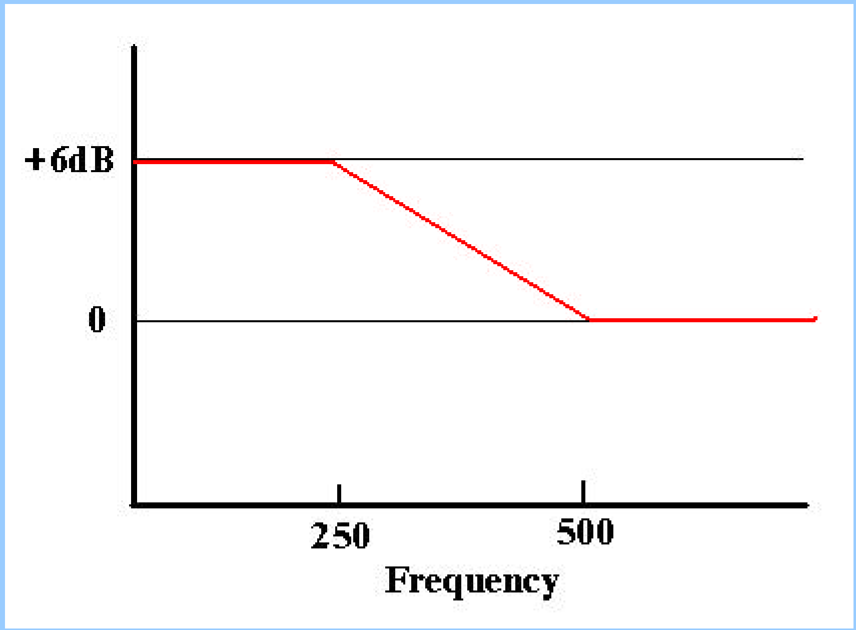
How can we reproduce envelopment in a small room?

- The reverberant field of a LARGE room can be reproduced in a SMALL room if:
  - We can excite a fluctuating sound VELOCITY across the listener’s head that mimics the fluctuating velocity in the original space.
  - To do this we MUST have at least two LF drivers on opposite sides of the listener.
  - If the listener is allowed to turn the head, we must have at least 3 independent drivers, and four is better!
  - All the LF drivers must be driven by independent (uncorrelated) reverberation signals, derived from a large, non-steady-state room.

Low frequencies are particularly important!

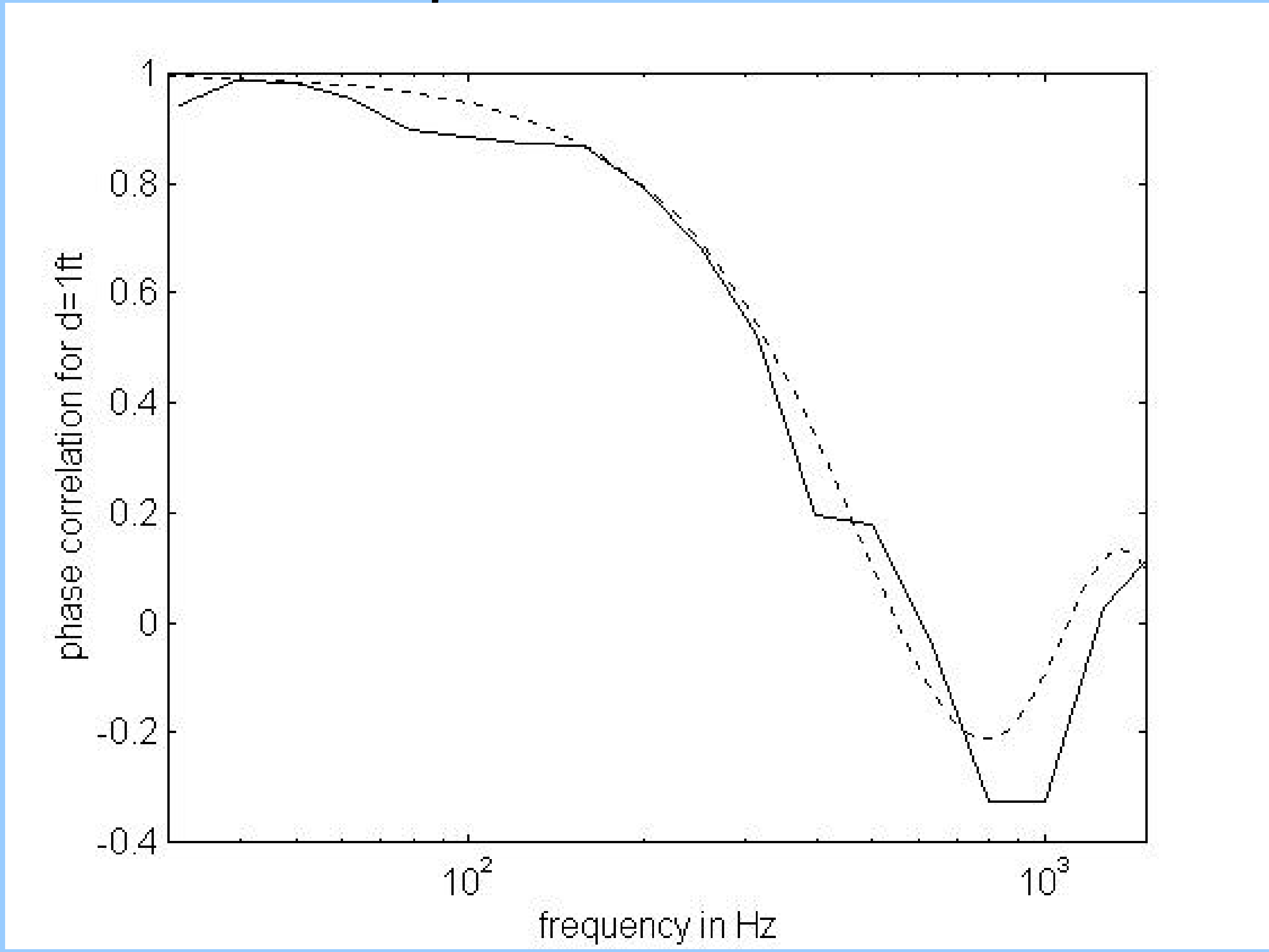
- In our concert hall and opera work it is frequencies below 300Hz where the major benefit is achieved.
  - The result is “inaudible” but highly effective in increasing the emotional power of the music.
- It is commonly believed that because we “cannot localize” low frequencies in a playback room we need only one LF driver
  - We can however easily hear the difference on reverberation.
- It is often the case that using a shelf filter on the rear channels can greatly improve the surround impression.

Shelf filter for rear channels



Applying a shelf filter to the rear channels increases subjective envelopment dramatically without drawing attention to the rear speakers.

Correlation in the omni “Hauptmicrophone” two omnis just behind the conductor.



___ = measured correlation; - - - = calculated, assuming d=25cm

Let’s build the hall sound

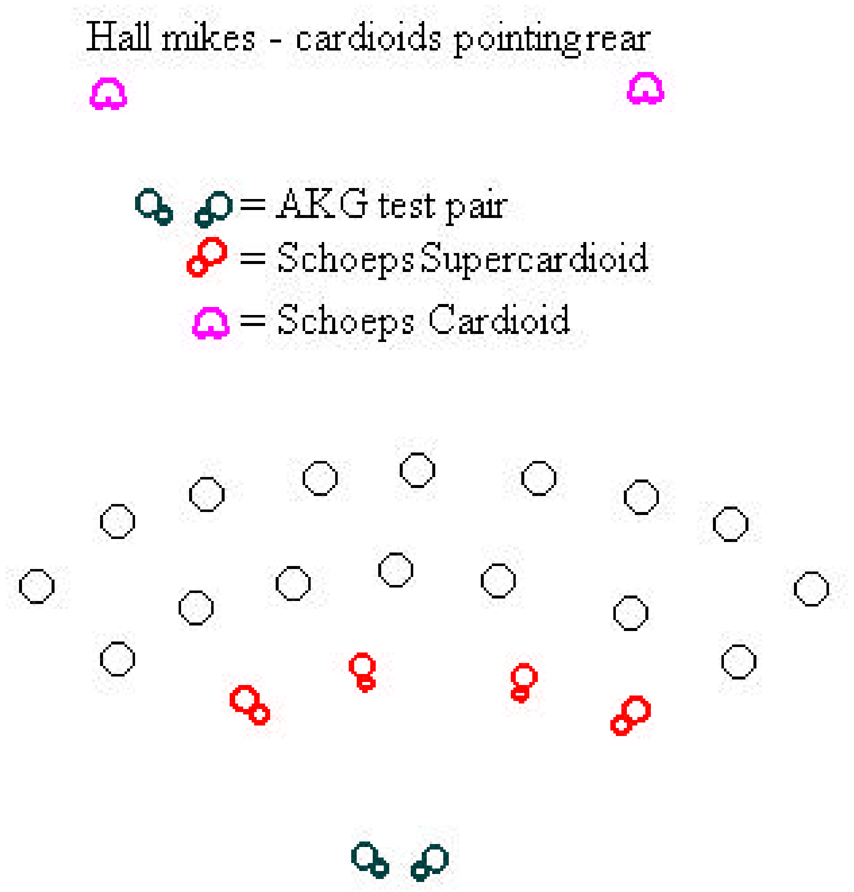
- We need decorrelated reverberation in both the front and the rear with equal level
- Test just the hall microphones to see if the reverberation is enveloping and uniform.
- Then add the microphones for the direct sound.
- Here there is too little chorus in the reverberation!!
- So we add hall (equally in all four outer speakers) from the Lexicon surround reverberator.

Oriana Consort in Swedenborg Chapel

Major Characteristics

- Hall has relatively low volume of 1450m^3 at the same time as medium RT ~1.5s
  - Low Volume and high RT means the reverb LEVEL will be very high!
    - We will have to keep the microphones close
  - Reverb time is a bit to short for this type of music.
  - With a small group it might be possible to use a microphone pair for a two channel recording.
    - But it might sound better if you did not.

Oriana Setup



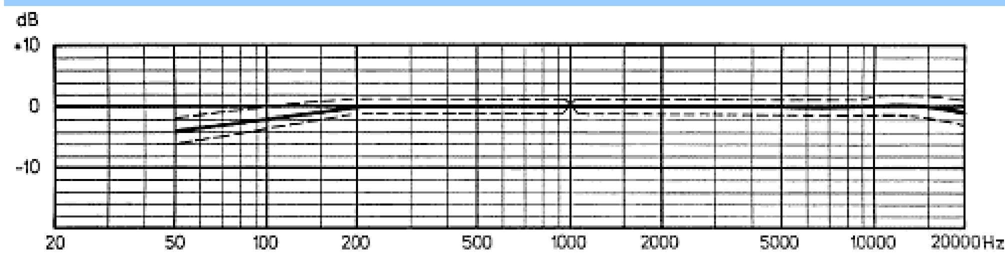
Surround Recording

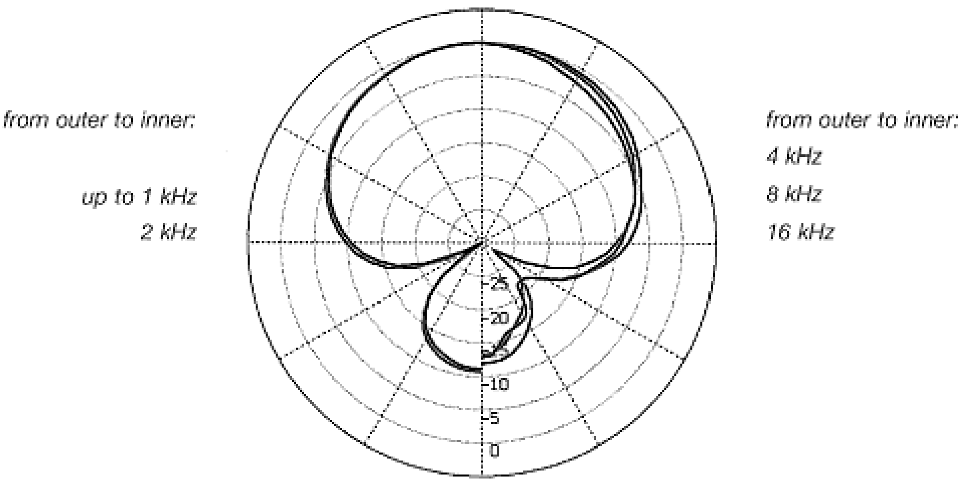
- The recording is created using the multimicrophone front array, (equalized)
  - Augmented with an early reflection pattern from Lexicon in all four outer speakers.
- The surround environment is created using the rear microphones (equalized for the bass roll-off) for the rear channels.
  - And Lexicon late reverberation for the front,
    - And some in the rear also.

The Microphone Pair

A venerable pair of multi-pattern microphones

Another possibility





Pressure Gradient Microphones

- Pressure gradient microphones are a combination of an omni and a figure of eight.
- When the two are mixed with equal on-axis sensitivity, a cardioid results.
  - Reduce the gain of the omni by 6dB and you have a Supercardioid.
  - Reduce the gain of the omni by 10dB and you have a Hypercardioid.

Problem:

- The figure of eight in nearly all available microphones has a bass roll-off, typically at about 120Hz. (Depends on diaphragm size.)
  - When we combine this with the omni – which (may) be inherently flat at LF, and:
    - The overall sensitivity decreases at LF
      - The mike sounds weak in the bass compared to an omni
    - The pattern may become omni directional at low frequencies
      - This is particularly true for large dual-diaphragm mikes.

Solution

- One solution to this problem is to “correct” it by measuring the microphone at a distance of 1M from the sound source!
  - A spherical sound wave increases the LF velocity of the sound at 6dB/octave when the distance to the sound source approaches ½ the wavelength.
  - A one meter measurement distance exactly compensates for the inherent roll-off of the velocity transducer, and an apparently perfect microphone results.
- A more satisfactory solution would be to equalize the figure of eight pattern electronically before combining it with the omni.
  - The “Soundfield” microphone does this.
- One can also roll off the omni response (electronically or mechanically) to match the figure of eight.
  - Mr. Wuttke (Schoeps) takes this approach.

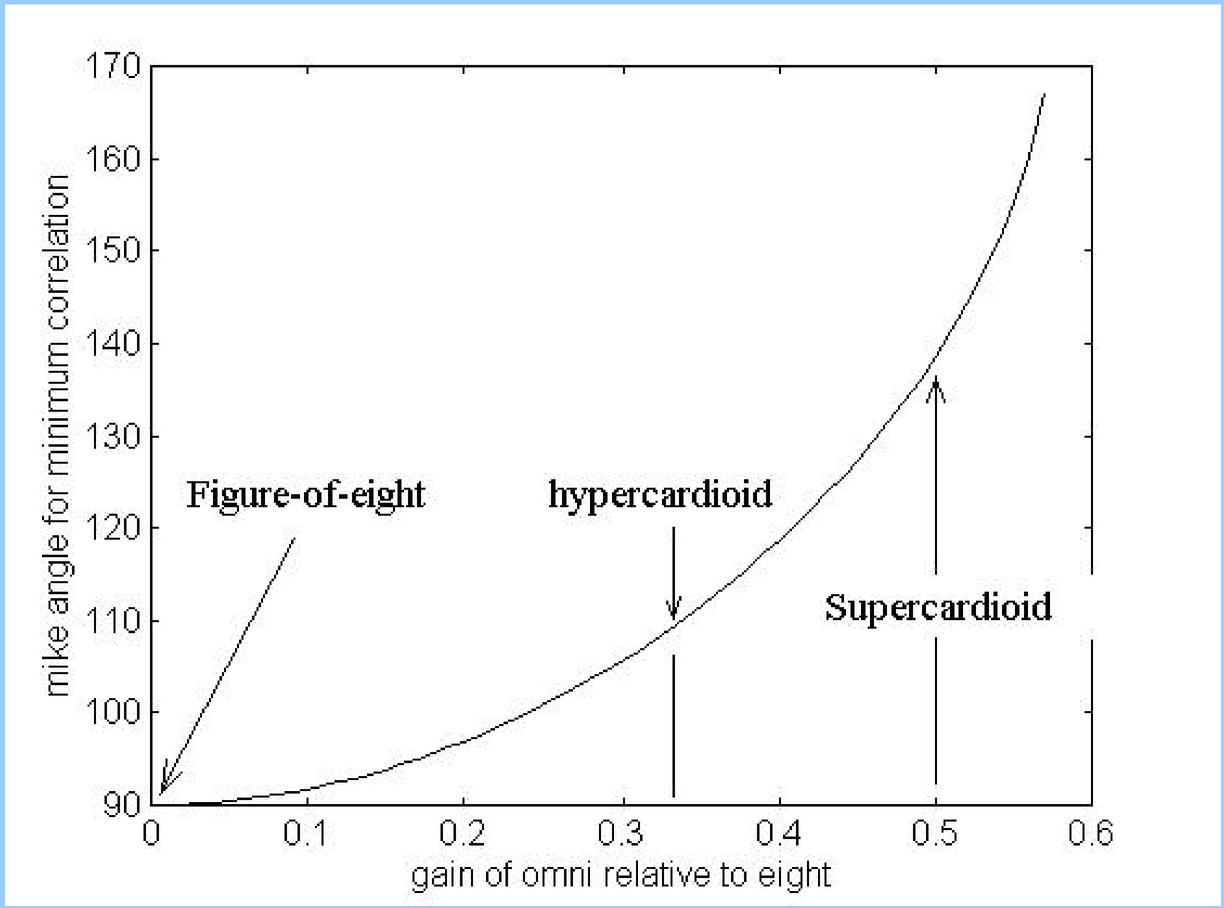
Consequences

- Nearly all available directional microphones either roll off the bass response,
  - Which can be compensated for at the mixing desk
- Or they become omnidirectional at low frequencies,
  - Which usually cannot be compensated.
- Or they do both.
  - The venerable microphones shown earlier do both.
- The consequence for a ORTF – style pair is
  - The low frequencies will be generally weak
    - Which can be compensated.
  - The low frequencies may be monaural
    - Which is more difficult to compensate.
    - But which can be fixed with a Blumlein shuffler circuit
- Be sure to equalize the LF when you use directional microphones!

Correlation of reverberation

- Remember we are (desperately) trying to keep the reverberation decorrelated.
  - We can do this with a coincident pair if we choose the right pattern and the right angle

Ideal angle as a function of microphone pattern for decorrelated reverberation in a coincident pair.

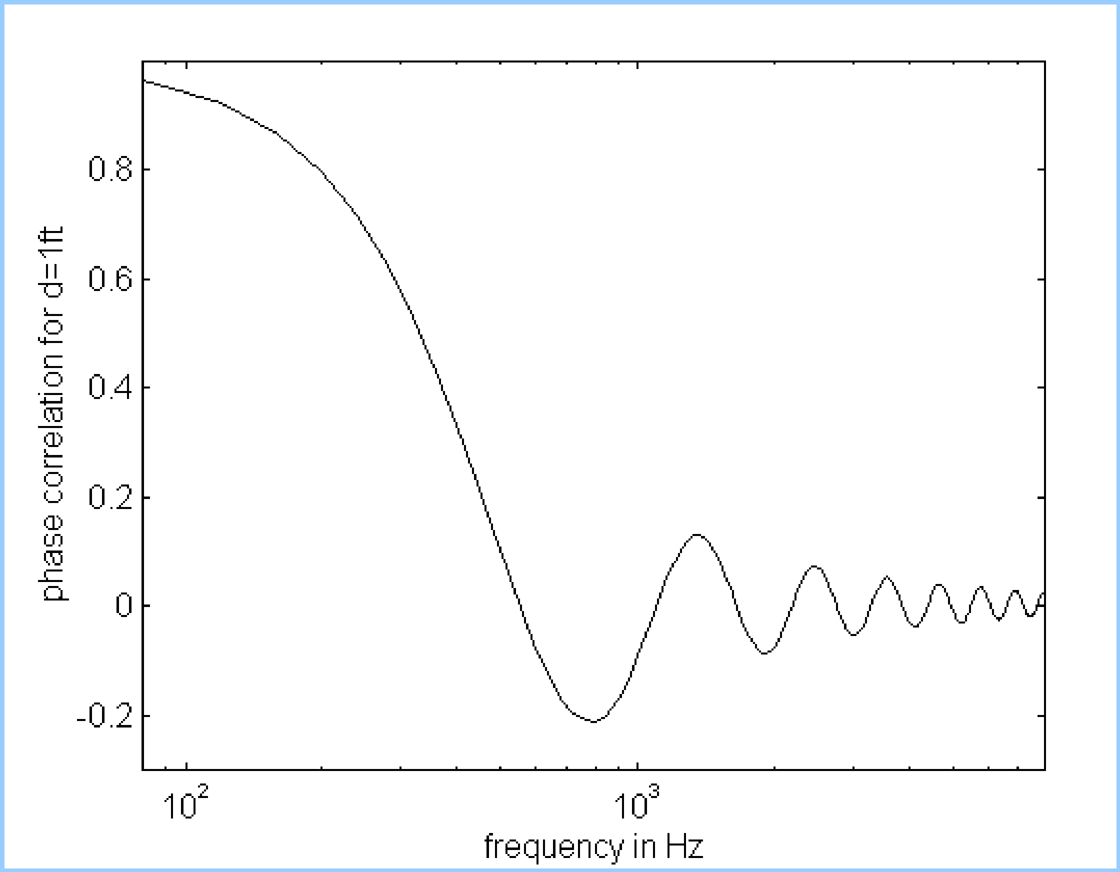


- It is NOT possible to achieve decorrelation with cardioid microphones!

Correlation through distance

- Normal ORTF technique with cardioid microphones reduces the correlation at HF by adding distance.
- But the trick does NOT work at LF,
- And LF correlation is exceedingly audible

Correlation of two omnidirectional microphones in a reverberant field as a function of microphone separation.

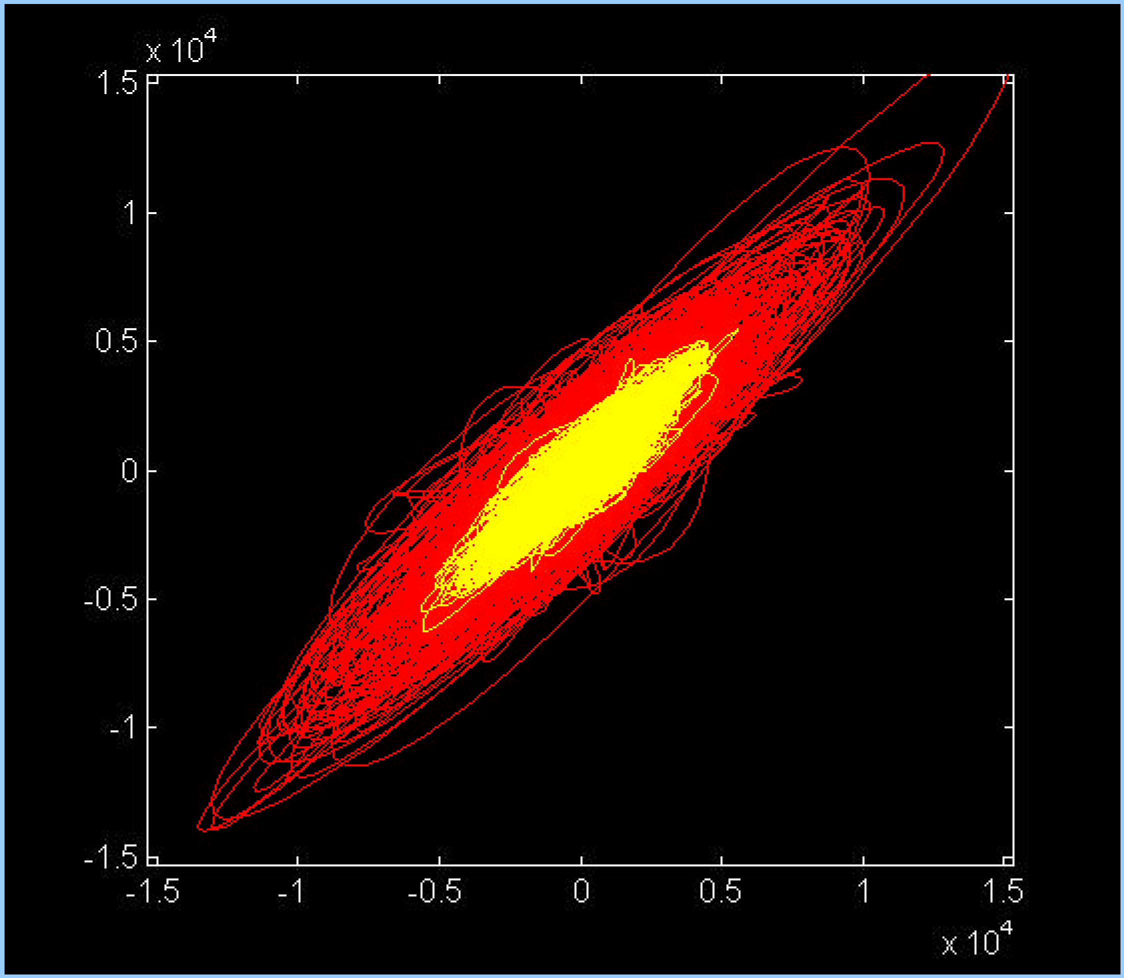


- Notice high correlation below 300Hz, and negative correlation at 800Hz.
- Frequency and distance are inversely proportional.

Audio Demos

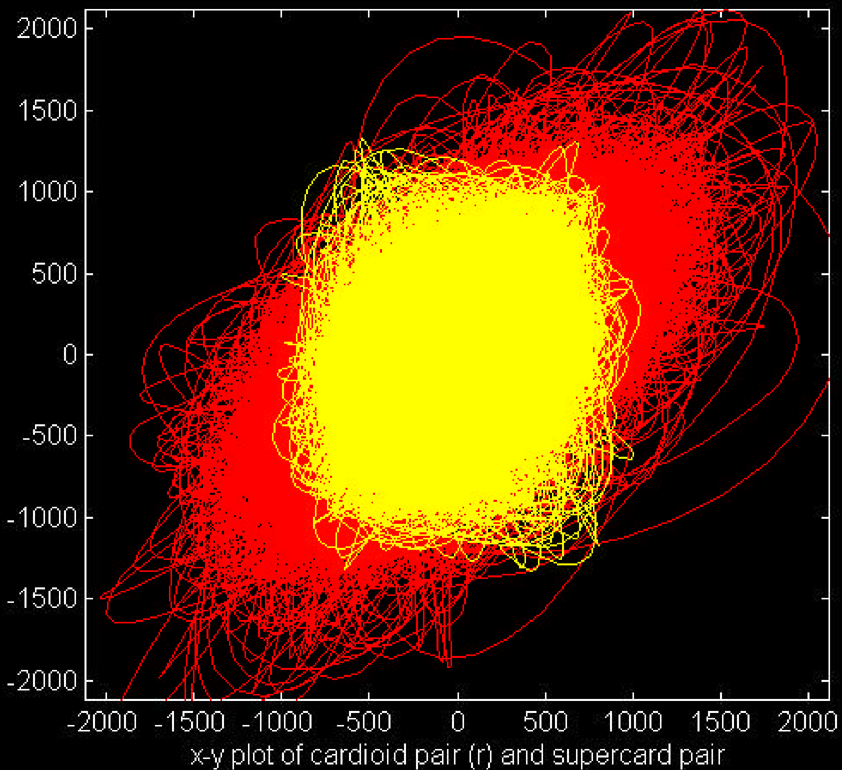
- Omni pair
  - Slavery Documents
- Cardioid Pair
  - AKG large diaphragm mikes with Oriana
- SuperCaridoid pair.
  - AKG large diaphragm mikes with Oriana
- Multimiked front image.
  - Oriana with four Schoeps Supercardioids

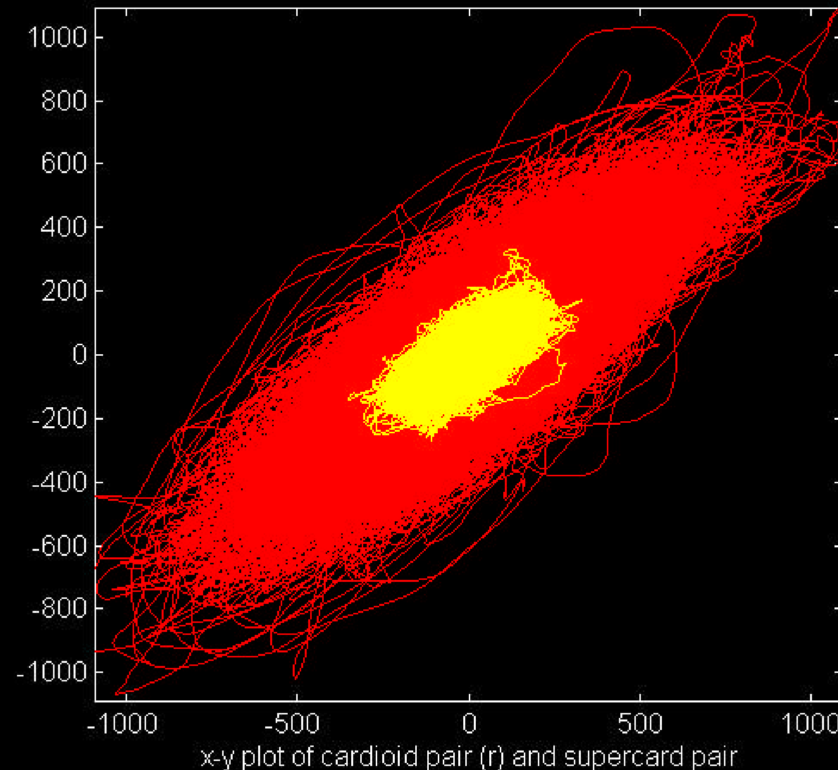
We can use a goniometer



X-Y plot of the omni front pair in Slavery Documents. Red trace is Low pass filtered at 200Hz, Yellow trace LP filtered at 100Hz.

Goniometer with AKG pair

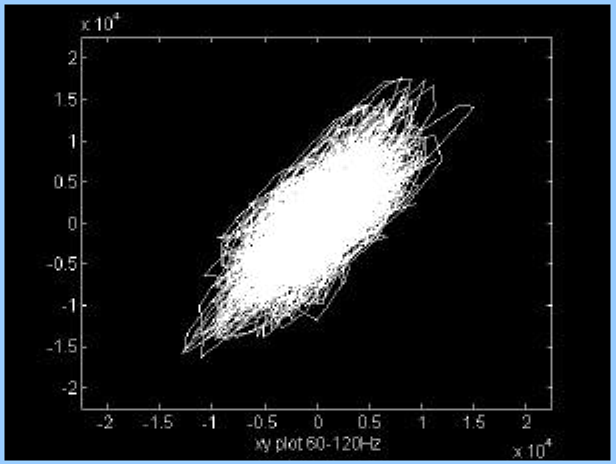




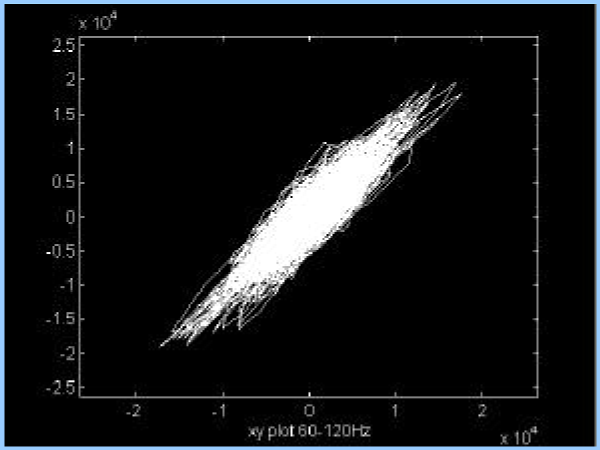
X-Y plot of Oriana front pair with a 200Hz LP filter. Red is Cardioid, Yellow is Supercardioid

The same data, filtered at 100Hz. Note that now the supercardioid is behaving like an omni.

Measure the decorrelation in the playback room



Hall sound from Slavery Documents. All speakers decorrelated.



We can make an x-y plot of the left ear and right ear signals, after boosting the L-R component ~10dB below 200Hz. These plots cover the frequency range of 20-100Hz.

The change in the ear correlation is quite audible.

Same, with a signal correlation of 30%

Final Mix

Conclusions

- Recording is a lot of fun!!!
- It is a great pleasure, and is often useful, to understand some of the science behind the microphones.
- Although simple techniques using microphone pairs or arrays can be seductive, a world-class sound usually requires many microphones, a lot of work, and artificial acoustic augmentation.
  - Time delay panning is undemocratic. Avoid it.
- Make SURE your reverberation is decorrelated, particularly at low frequencies.

THE END