



www.akutek.info

PRESENTS

**The Relationship Between Audience
Engagement and Our Ability to Perceive the
Pitch, Timbre, Azimuth and Envelopment of
Multiple Sources**

David Griesinger
Consultant
Cambridge MA USA
www.DavidGriesinger.com

Overview

- This talk consists of an *introduction*, followed by sections from three fields.
 - The intro states the goal of the talk - to help build better concert halls and opera houses.
 - To do this we need to understand how acoustics affects the perception of sound.
- Part one – *Physics* – describes a physical mechanism by which human hearing may detect the pitch, timbre, azimuth and distance (near/far) of several sound sources at the same time, using frequencies the range of vocal formants (1000Hz to 4000Hz.)
 - The acuity of these perceptions is reduced in the presence of reflections and reverberation in a consistent and predictable fashion.
 - The consequence of this reduction in acuity is the perception of distance from the source, and a loss in excitement or engagement with the performance.
 - A computer model of this mechanism can be used to measure the psychological clarity of a hall from recordings of live music.
- Part two – *Psychology* – discusses the psychological importance of the perception of “near” and “far” on the ability of a sound to hold the attention of the audience.
 - And makes a plea for hall and opera designs that maximize audience engagement
- Part three – *Acoustics* – looks at the acoustic reasons certain concert halls are more engaging than others.
 - Hall shape does not scale. A shoebox shape that works for a hall with 2000 seats large hall will produce muddy sound over a wide range of seats if it is scaled to 1000 seats
 - Rectangular diffusing elements – coffers and niches – act as frequency dependent retro-reflectors, and are an essential ingredient in maintaining high clarity over a large number of seats.

Warning! Radical Concepts Ahead!

- The critical issue is the *amount*, the *time delay*, and the *frequency content* of early reflections relative to the direct sound.
 - If the direct to reverberant ratio above 700Hz is above a critical threshold, early energy and late reverberation can enhance the listening experience. But, if not...
- Reflections in the time range of 10 to 100ms *reduce* clarity, envelopment, and engagement – whether lateral or not.
 - and the earliest reflections are the most problematic.
- Reflections off the back wall of a stage or shell *decrease* clarity
 - They are typically both early and strong – and interfere with the direct sound.
- Side-wall reflections are desirable in the front of a hall, but reduce engagement in the rear seats.
 - They are earlier, and stronger relative to the direct sound in the rear.
- Reflections above 700Hz directed into audience sections close to the sound sources have the effect of *reducing* the reflected energy in other areas of the hall – with beneficial results.
 - These features *increase* the direct/reverberant ratio in the rear seats
 - And attenuate the upper frequencies from side-wall reflections in the rear.
 - *Coffers, niches, and/or open ceiling reflectors are invariably present in the best shoebox halls.*

A few that work – note the rectangular coffers and niches



Boston Symphony Hall



Amsterdam Concertgebouw



Vienna Grosse Musikverreinsaal



Tanglewood Music Shed

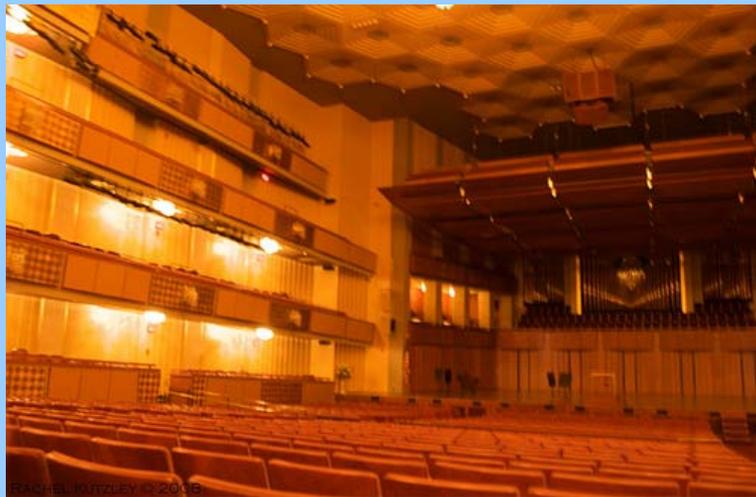
Nice try ... – and there are plenty more...



Avery Fisher Hall, New York



Alice Tully Hall, New York



Kennedy Center Washington, DC



Salle Playel, Paris

Introduction

- This talk is centered on the properties of sound that promote *engagement* – the focused attention of a listener.
 - Engagement is usually subconscious – and the study of its dependence on acoustics has been neglected in most acoustic research.
 - At some level the phenomenon is well known:
 - Drama and film directors insist that performance venues be acoustically dry, with excellent speech clarity and intelligibility.
 - As do producers and listeners of popular music, and customers of electronically reproduced music of all genres.
 - The same acoustic properties that create excitement in a play or film can increase the impact of live classical music – but many current halls and opera houses are not acoustically engaging in a wide range of seats.
 - Halls with poor engagement decrease audiences for live classical music.
- Engagement is associated with sonic clarity – but currently there is no standard method to quantify the acoustic properties that promote it.
 - Acoustic measurements such as “Clarity 80” or C80, were developed to quantify intelligibility, not engagement.
 - Venues often have adequate intelligibility – particularly for music – but poor engagement
- Acoustic engineers and architects cannot design better halls and opera houses without being able to specify and verify the properties they are looking for.
 - So we desperately need measures for the kind of clarity that leads to engagement.

The story of “near”, “far”, and harmonic coherence

- The author has been fascinated with engagement for a long time
 - particularly the perception of muddiness in a recording, and the lack of dramatic clarity in a hall or opera house.
- This fascination led to a discovery that a major determinant of engagement was the perception of “near” and “far”,
 - which humans can determine immediately on hearing a sound, even with only one ear, or with a single channel of recorded sound.
 - The perception has vital importance, as it subconsciously determines the amount of attention we will pay to a sound event.
- The importance of this perception, and the speed with which we make it, argue that determining “near” and “far” is a fundamental property of sound perception.
 - But how do we perceive it, and how can it be measured?
- In searching for the answer, the author found that engagement, “near” and “far”, pitch perception, timbre perception, and direction detection are all related to the same property of sound:
 - **the phase coherence of harmonics in the vocal formant range, ~1000Hz to 4000Hz.**

Example: The syllables one to ten with four different degrees of phase coherence. The sound power and spectrum of each group is identical



Near, far, and sound localization

- The first step to the realization of the fundamental importance of phase coherence came from author's listening experience, which suggested that the perception of "near" and "far" is closely related to the ability to accurately identify the direction of a sound source.
 - When individual musicians in a small classical music ensemble sounded engaging and close to the listener, they could be accurately localized.
 - And when they sounded distant and non-engaging, they were difficult to localize.
 - Engagement is mostly sub-conscious and difficult to quantify – but localization experiments are relatively easy to perform – so I studied localization.
- Experiments with several subjects showed that the ability to localize sounds in a reverberant environment depends on frequencies between 700Hz and 4000Hz,
 - and that poor localization occurs when the sum of early reflections in the time range from 5ms to 100ms from any direction becomes stronger than the direct sound.
 - The earlier a reflection comes, the larger is its detrimental effect.
- With the help of localization data it was possible to construct a measure for the ability to localize sound in a reverberant environment.
 - The input to the measure is a measured or calculated binaural impulse response at a particular seat, – ideally with an occupied hall and stage.

Equation for Localizability – 700 to 4000Hz

- We can use a simple model to derive an equation that gives us a decibel value for the ease of perceiving the direction of direct sound. The input $p(t)$ is the sound pressure of the source-side channel of a binaural impulse response.
 - We propose the threshold for localization is 0dB, and clear localization and engagement occur at a localizability value of +3dB.

- Where D is the window width (~ 0.1 s), and S is a scale factor:

$$S = 20 - 10 * \log \int_{.005}^{\infty} p(t)^2 dt$$

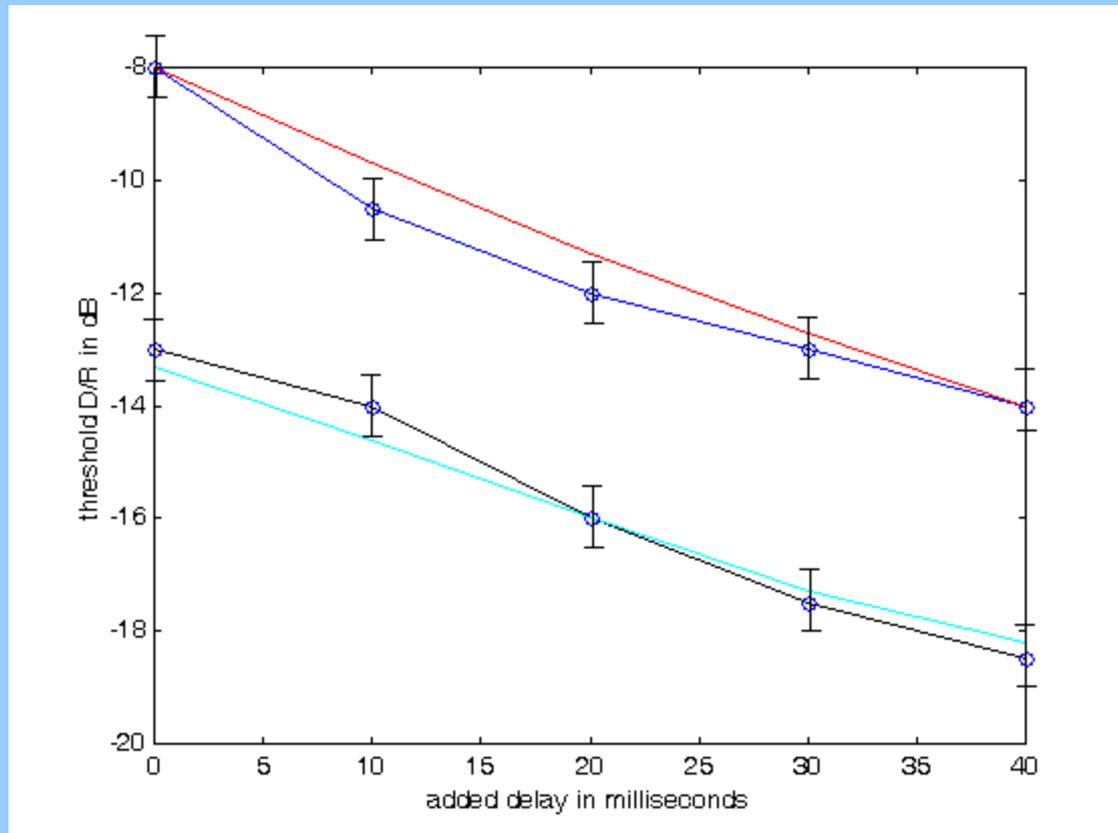
S is the zero nerve firing line. It is 20dB below the maximum loudness. POS in the equation below means ignore the negative values for the sum of S and the cumulative log pressure.

- Localizability (LOC) in dB =

$$S - 1.5 + 10 * \log \int_0^{.005} p(t)^2 dt - (1/D) * \int_0^{D-.005} POS(S + 10 * \log \int_{.005}^{\tau} p(t)^2 dt) d\tau$$

- The scale factor S and the window width D interact to set the slope of the threshold as a function of added time delay. The values I have chosen (100ms and -20dB) fit my personal data. The extra factor of +1.5dB is added to match my personal thresholds.
- Further description of this equation is beyond the scope of this talk – but it is explained on the author's web-page..

Broadband Speech Data verifies the LOC equation:



Blue – experimental thresholds for alternating speech with a 1 second reverb time. Red – the threshold predicted by the localization equation. Black – experimental thresholds for RT = 2seconds. Cyan – thresholds predicted by the localization equation.

Measures from live music

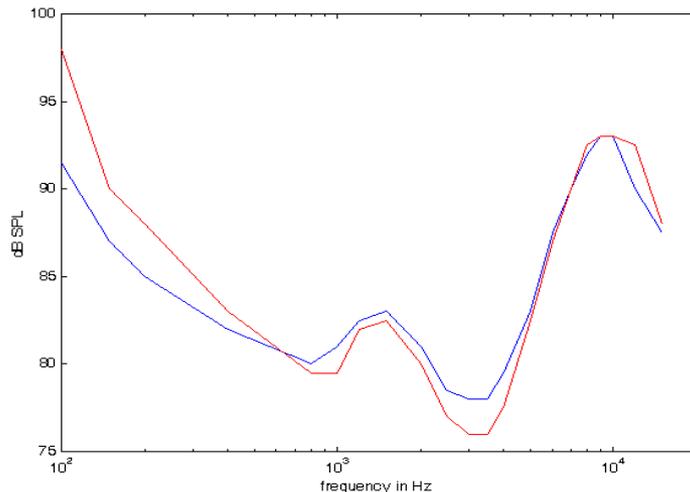
- Binaural impulse responses from occupied halls and stages are very difficult to obtain!
 - But if you can hear something, there must be a way to measure it.
- Part one of this talk describes a physiologically derived model of human hearing. The model arose from the search for a measure for “near” and “far”.
 - But the model is capable of explaining (and measuring) far more.
 - The model provides a means of separating sounds from multiple sources into independent neural streams,
 - And allows independent analysis of each stream for pitch, timbre, and azimuth.
- The model may not be neurologically correct in detail
 - But it predicts many known properties of human hearing, and shows that all of them depend on the phase coherence of the incoming sound.
 - It provides a method that this phase coherence can be measured from binaural recordings of live music.
- This model is the subject of part one of this talk. Parts two and three show why the model is needed.

Part one – Physics

- Part one describes a physical mechanism by which human hearing could detect pitch, timbre, azimuth and distance (near/far) of several sound sources at the same time, using the phase coherence of harmonics in the range of vocal formants (1000Hz to 4000Hz.)
 - The model is built from functions that are known to be present in human hearing.
- Signals from the basilar membrane are analyzed not just for their average amplitude, but for modulations produced by interference between harmonics.
 - This information derives from the phase relationships between harmonics .
- A conceptually simple mechanism is suggested that allows the information from these modulations to be separated into independent neural streams, one for each sound source.
 - This this mechanism explains our uncanny abilities to detect the pitch, timbre, azimuth and distance of several sources at the same time.
 - And it also predicts the observed decrease in these abilities in the presence of reflections.
- The model need not be entirely correct to support the main point of this talk, which is that the model's success in predicting what is and is not audible strongly supports the conclusions of part two and part three.
 - The phase coherence of harmonics in the vocal formant range give rise to our abilities to separate sounds from multiple sources, and independently perceive pitch, timbre, azimuth, and distance for each source.
 - The acuity of these perceptions is reduced in the presence of reflections and reverberation in a consistent predictable, and measureable fashion.
 - In the absence of this acuity sources become psychologically distant and non-engaging.
- The model provides a means for measuring the degree of phase coherence – and thus the engagement – in an individual seat using only live sound as an input.

Perplexing Phenomena

- The frequency selectivity of the basilar membrane is approximately 1/3 octave (~25% or 4 semitones), but musicians routinely hear pitch differences of a quarter of a semitone (~1.5%).
 - Clearly there are additional frequency selective mechanisms in the human ear.
- the fundamentals of musical instruments common in Western music lie between 60Hz and 800Hz, as do the fundamentals of human voices.
 - But the sensitivity of human hearing is greatest between 500Hz and 4000Hz, as can be seen from the IEC equal loudness curves.



Blue: 80dB SPL ISO equal loudness curve.

Red: 60dB equal loudness curve

The peak sensitivity of the ear lies at about 3kHz.

Why? Is it possible that important information lies in this frequency range?

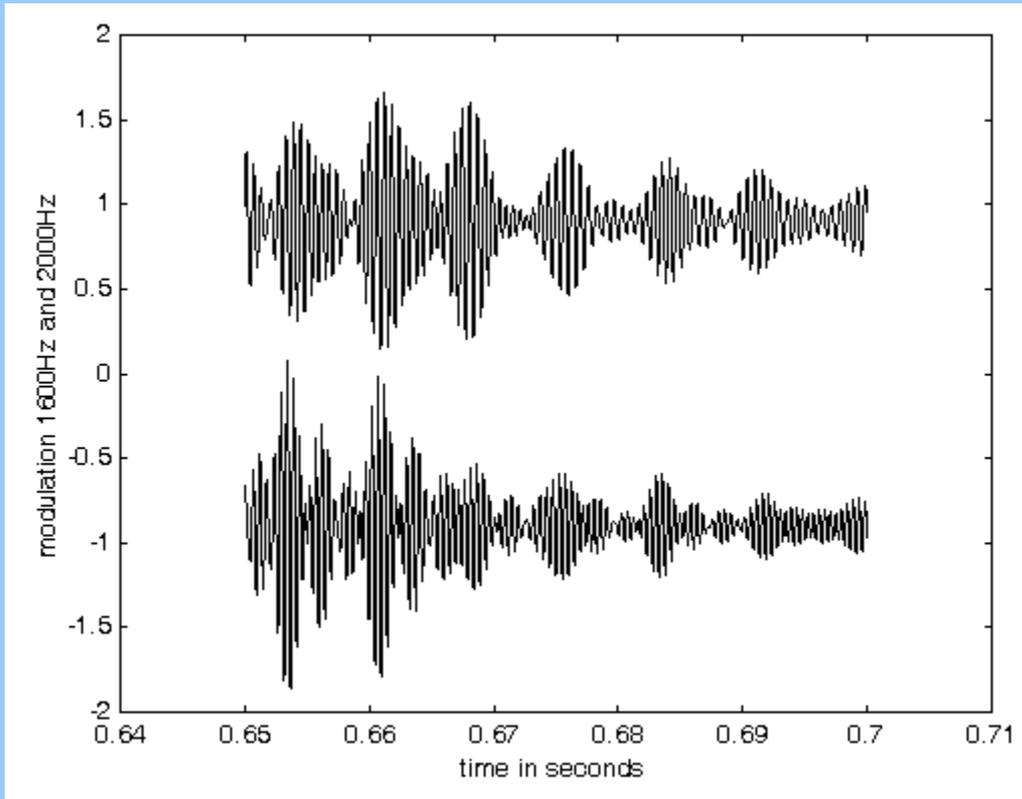
More Perplexing Phenomena

- Analysis of frequencies above 2kHz would seem to be hindered by the maximum nerve firing rate of about 1kHz.
 - Why has evolution placed such emphasis on a frequency range that is difficult to analyze directly?
- A typical basilar membrane filter above 2kHz has three or more harmonics from each instrument within its bandwidth
 - How can we possibly separate them?
- How is it possible that in a good hall we can routinely detect the azimuth, pitch, and timbre of three or more sound sources (musicians) at the same time?
 - Even in a concert where a string quartet subtends an angle of ± 5 degrees or less! (The ITDs and ILDs are miniscule...)
- Why do some concert halls prevent you from hearing several musical lines at once?
 - And what can be done about it?
- The hair cells in the basilar membrane respond mainly to negative pressure – they approximate half-wave rectifiers, which are strongly non-linear devices. How can we claim to hear distortion at levels below 0.1% ?
- Why do many creatures – certainly all mammals – communicate with sounds that have a defined pitch?
 - Is it possible that pitched sounds have special importance to the separation and analysis of sound?

Answers

- Answers become clear with two basic realizations:
 - *1. The phase relationships of harmonics from a complex tone contain more information about the sound source than the fundamentals.*
 - *2. And these phase relationships are scrambled by early reflections.*
- For example: my speaking voice has a fundamental of 125Hz.
 - The sound is created by pulses of air when the vocal chords open.
 - Which means that exactly once in a fundamental period all the harmonics are in phase.
- A typical basilar membrane filter at 2000Hz contains at least 4 of these harmonics.
 - The pressure on the membrane is a maximum when these harmonics are in phase, and reduces as they drift out of phase.
 - The result is a strong amplitude modulation in that band at the fundamental frequency of the source.
- When this strong modulation is absent, or noise-like, the sound is perceived as distant.

Basilar motion at 1600 and 2000Hz



Top trace: A segment of the motion of the basilar membrane at 1600Hz when excited by the word “two”

Bottom trace: The motion of a 2000Hz portion of the membrane with the same excitation. The modulation is different because there are more harmonics in this band.

In both bands there is strong amplitude modulation of the carrier, and the modulation is largely synchronous.

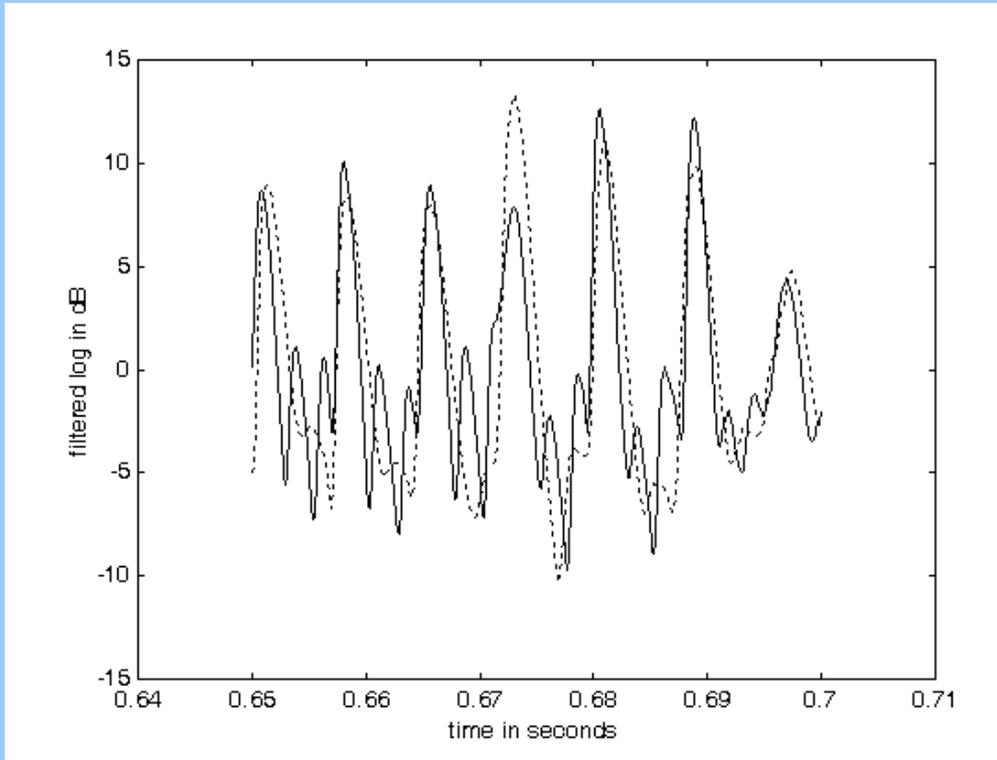
When we listen to these signals the fundamental is easily heard



In this example the phases have been garbled



Nerve firing rates



This picture shows the amplitude envelope of the previous picture, plotted in dB. It represents the rate of nerve firings from each band. The rate varies over a sound pressure range of 20dB.

Nerve cells act like an AM radio detector, which recovers the frequency and amplitude of the modulation, while filtering away the frequency of the carrier.

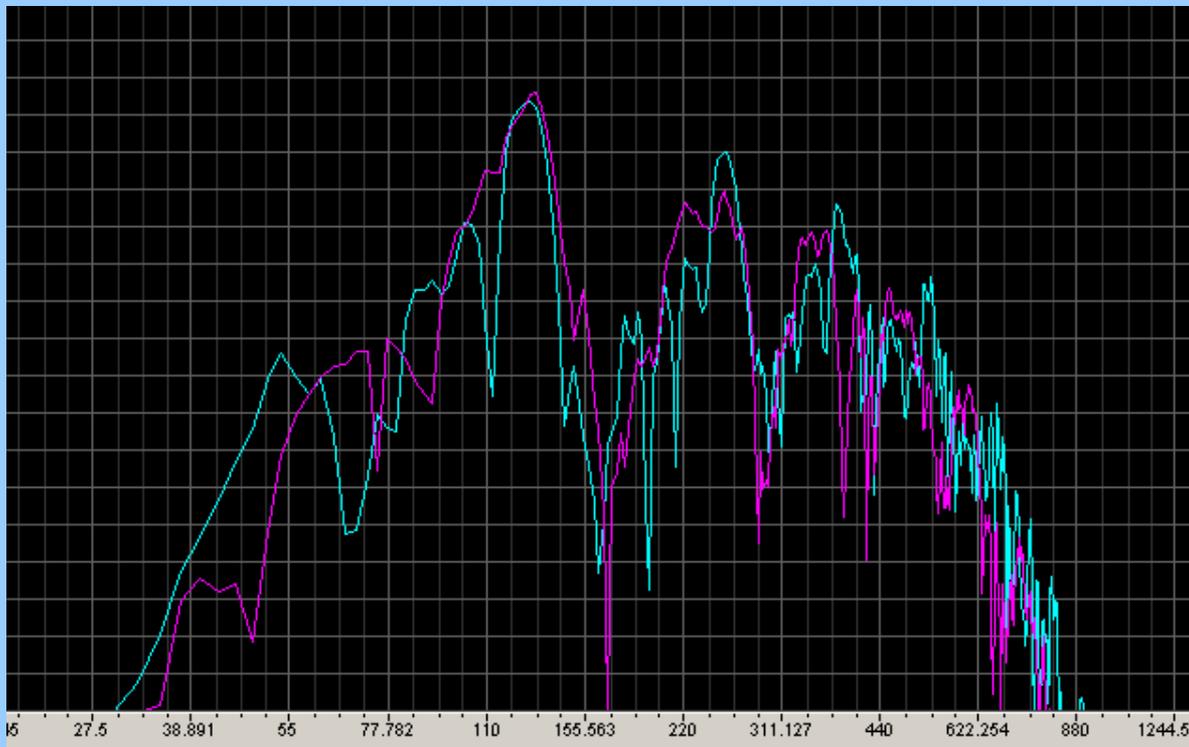
Like the detectors in AM radios, the hair cells (probably) include AGC (automatic gain control) – with about a 10ms time constant. The response over short times is linear, but appears logarithmic over longer periods.

AM Radio

- AM radio consists of a carrier at a fixed high frequency that has been linearly modulated by low frequency signals.
- An AM receiver half-wave rectifies the carrier, and filters out the high frequency components.
 - What remains is the recovered low frequency signals.
 - So an AM radio receiver uses a strongly non-linear device to recover a linear signal.
- But the rectification process can be viewed as a kind of sampling – it also produces aliases of the modulation.
 - In the case of an AM radio the aliases are at very high frequencies, and can be easily filtered away.
 - In the basilar membrane the carrier is close to the frequencies of the modulation – and the aliases can be problematic.

Amplitude modulation of a noisy carrier

- The motion of the basilar membrane when excited by phase-coherent harmonics appears to be an amplitude modulated carrier – but the carrier is not a fixed frequency, but an artifact of a filter with a finite bandwidth.
 - And the frequency of the carrier is within the audio band.
- Thus, rectification by the hair cells produces aliases that are both broad-band and highly audible. 📢



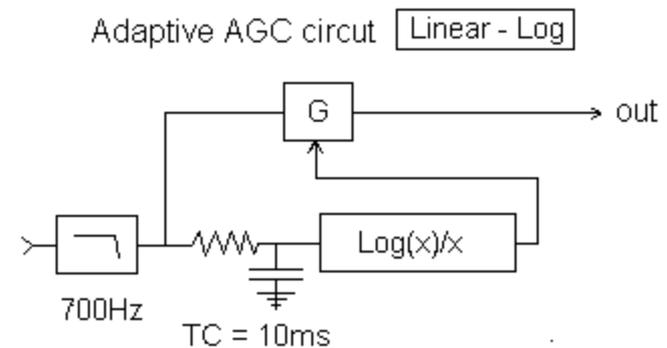
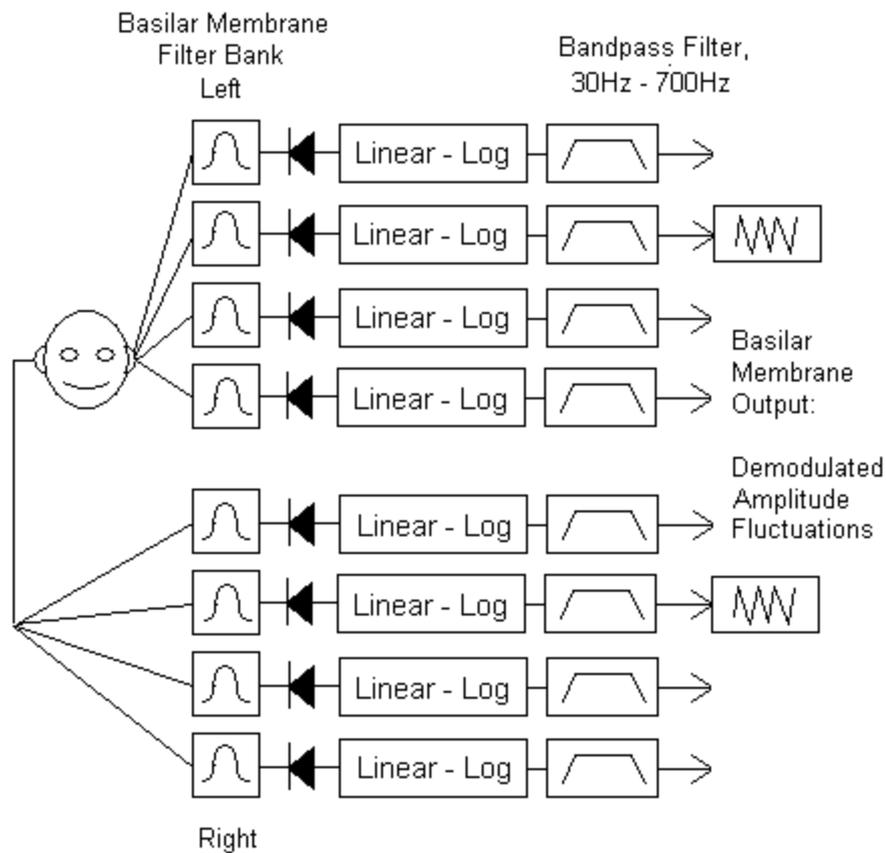
Spectrum of the syllable “three” from the rectified and filtered 2000Hz 1/3 octave band (blue) and the 2500Hz 1/3 octave band. (red)

Note the fundamental frequency and its second harmonic are the same in both bands. The garbage is different.

Recovering a linear signal

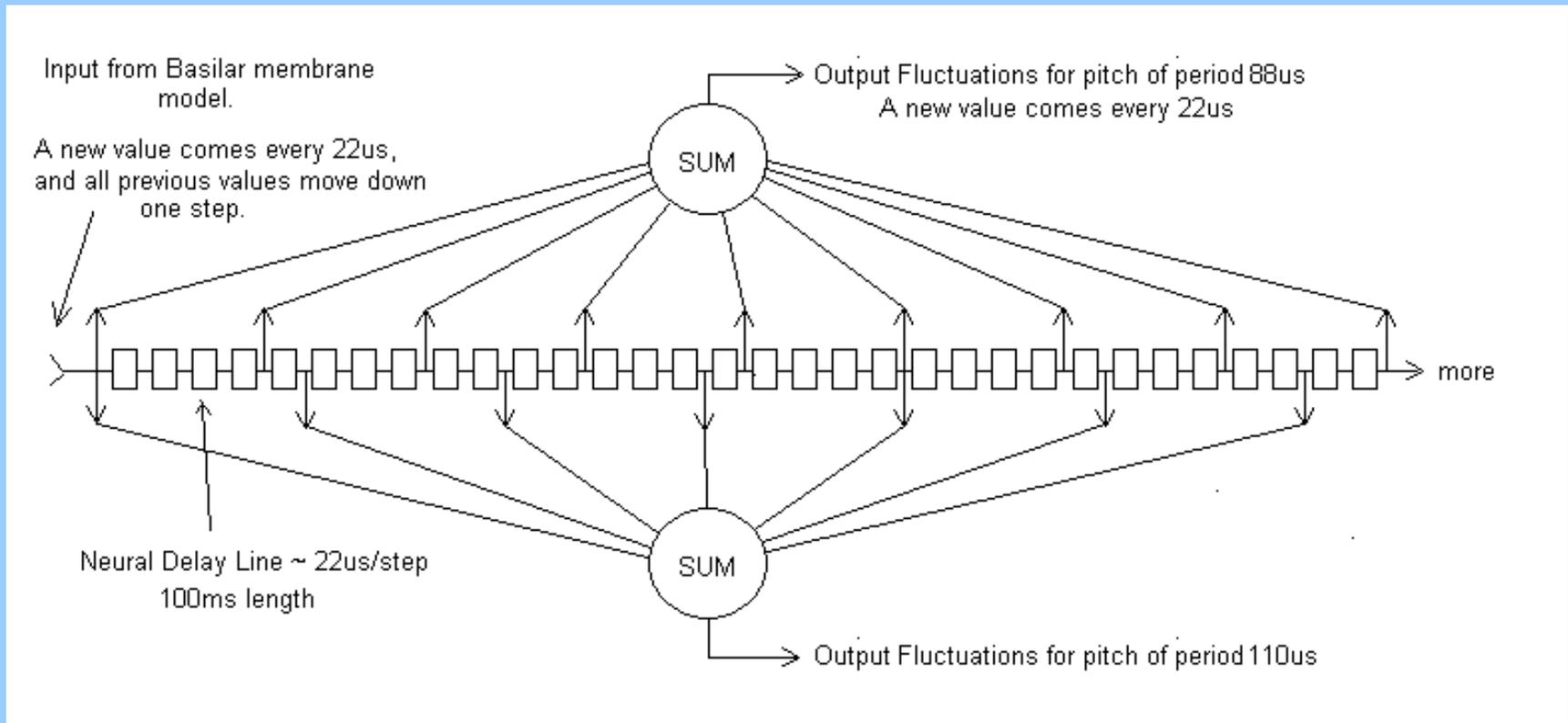
- To recover a linear signal from these hair cells we need to have to combine and compare the outputs from many overlapping critical bands.
 - The aliases in each band are different because the carriers have different frequencies – but the modulations we wish to hear are nearly the same.
- Since for most signals the artifacts are not constant in time – we must also average the hair-cell firings over a period of time.
 - My data suggests an averaging time of 100ms.
- Because the carrier is broad-band, the aliases are also broad-band.
 - The signals are generally narrow band – so broad band signals may be ignored.
- Our hearing mechanism does all of these things.

An amplitude-modulation based basilar membrane model



The gain stage G is driven by the amplitude envelope of the rectified basilar membrane motion. The output of the circuit is linear for short-time variations in amplitude, but logarithmic for the average amplitude.

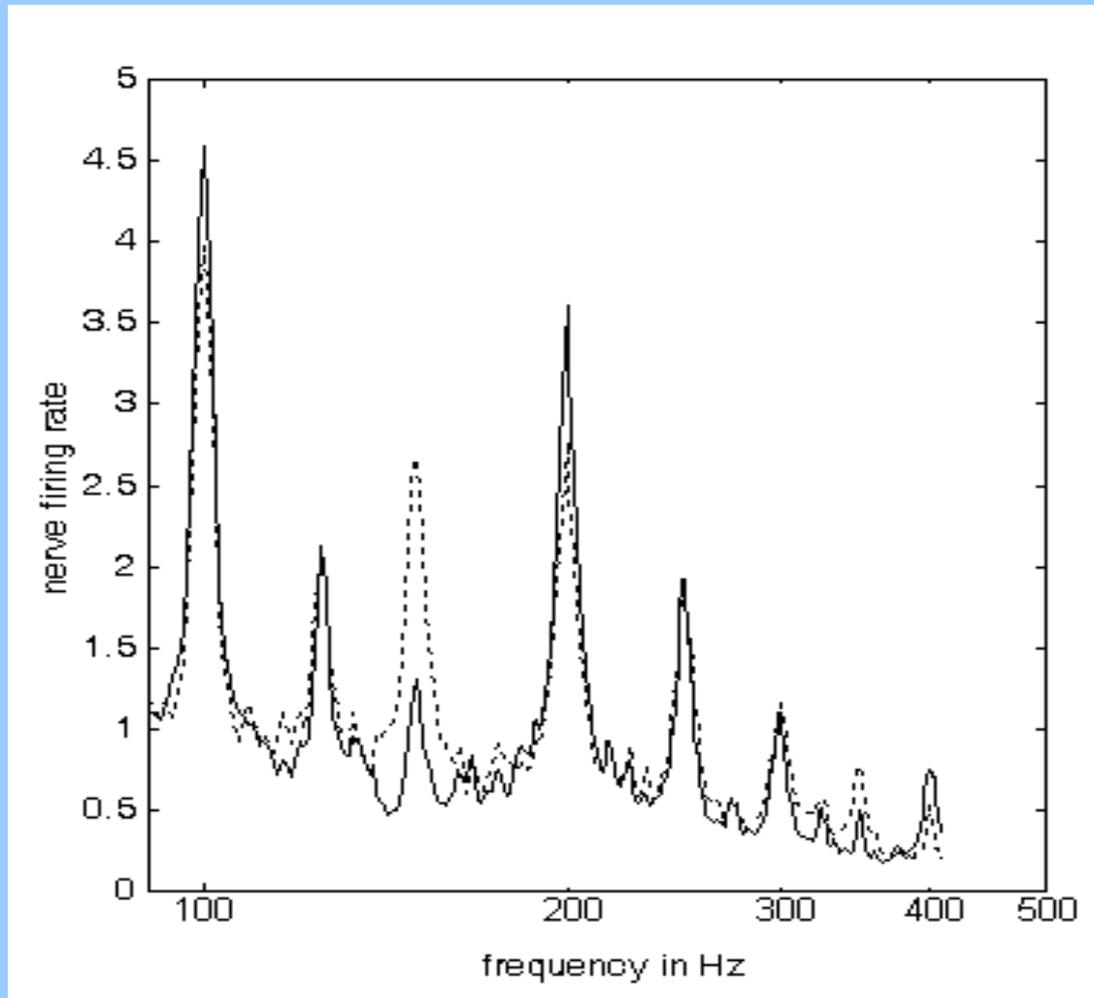
A Pitch Detection Model



A neural daisy-chain delays the output of the basilar membrane model by 22us for each step. Dendrites from summing neurons tap into the line at regular intervals, with one summing neuron for each fundamental frequency of interest.

Two of these sums are shown – one for a period of 88us, and one for a period of 110us. Each sum produces an independent stream of nerve fluctuations, each identified by the fundamental pitch of the source.

Pitch acuity – A major triad in two inversions



Solid line - Pitch detector output for a major triad – 200Hz, 250Hz, 300Hz

Dotted line – Pitch detector output for the same major triad with the fifth lowered by an octave: 200Hz, 250Hz and 150Hz.

Note the high degree of similarity, the strong signal at the root frequency, and the sub-harmonic at 100Hz

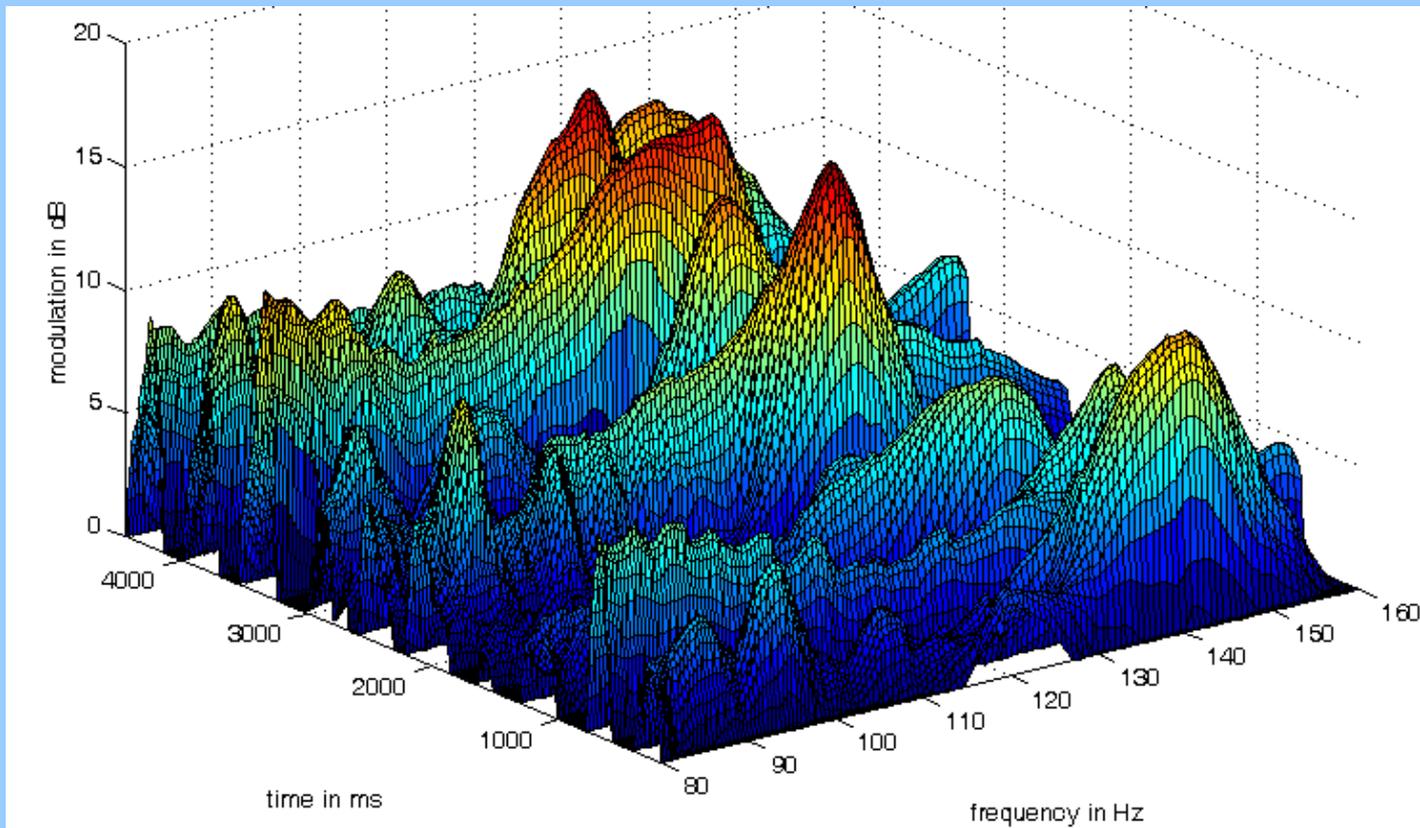
Summary of model

- We have used a physiological model of the basilar membrane to convert sound pressure into demodulated fluctuations in nerve firing rates for a large number of overlapping (critical) bands.
- Our physiological model of the frequency separation mechanism is capable of analyzing the modulations in each band into perhaps hundreds of frequency bins.
 - Strong, narrow-band signals at particular frequencies are selected for further processing
- The result: we have separated signals from a number of sources into separate neural streams, each containing the modulations received from that source.
 - These modulations can then be compared across bands to detect timbre, and IADs and ILDs can be found for each source to determine azimuth.

Advantages

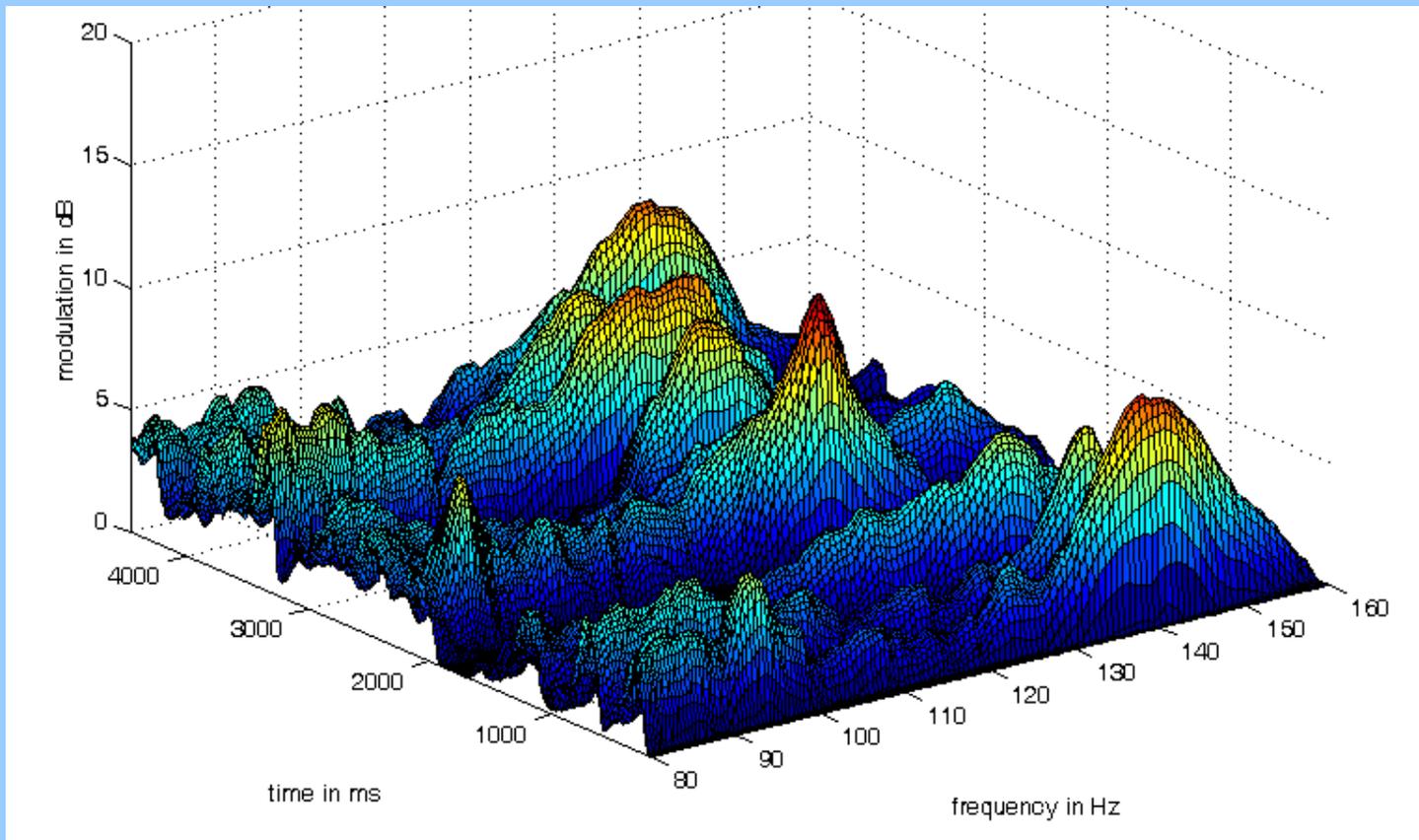
- The separated streams from each source can be easily analyzed for timbre, ITD and ILD with known neural circuits.
- The model is conceptually simple – it is built out of a (large) number of building blocks that are known to exist in human neurology.
 - It is easy to see how it could have evolved.
- The circuit is fast. Useful data on timbre, ITD, and ILD is available within 20ms of the first input.
 - As the sound is held pitch and azimuth acuity increases.
- Because the ILD is created by high frequency harmonics, small differences in azimuth can create large differences in level
 - Thus azimuth acuity is high enough to explain our ability to localize musicians.

Speech without reverberation: 1.6kHz-5kHz



Note that the voiced pitches of each syllable are clearly seen. Since the frequencies are not constant, the peaks are broadened – but the frequency grid is 0.5%, so you can see that the discrimination is not shabby.

Speech with reverberation: RT=2s, D/R -10dB

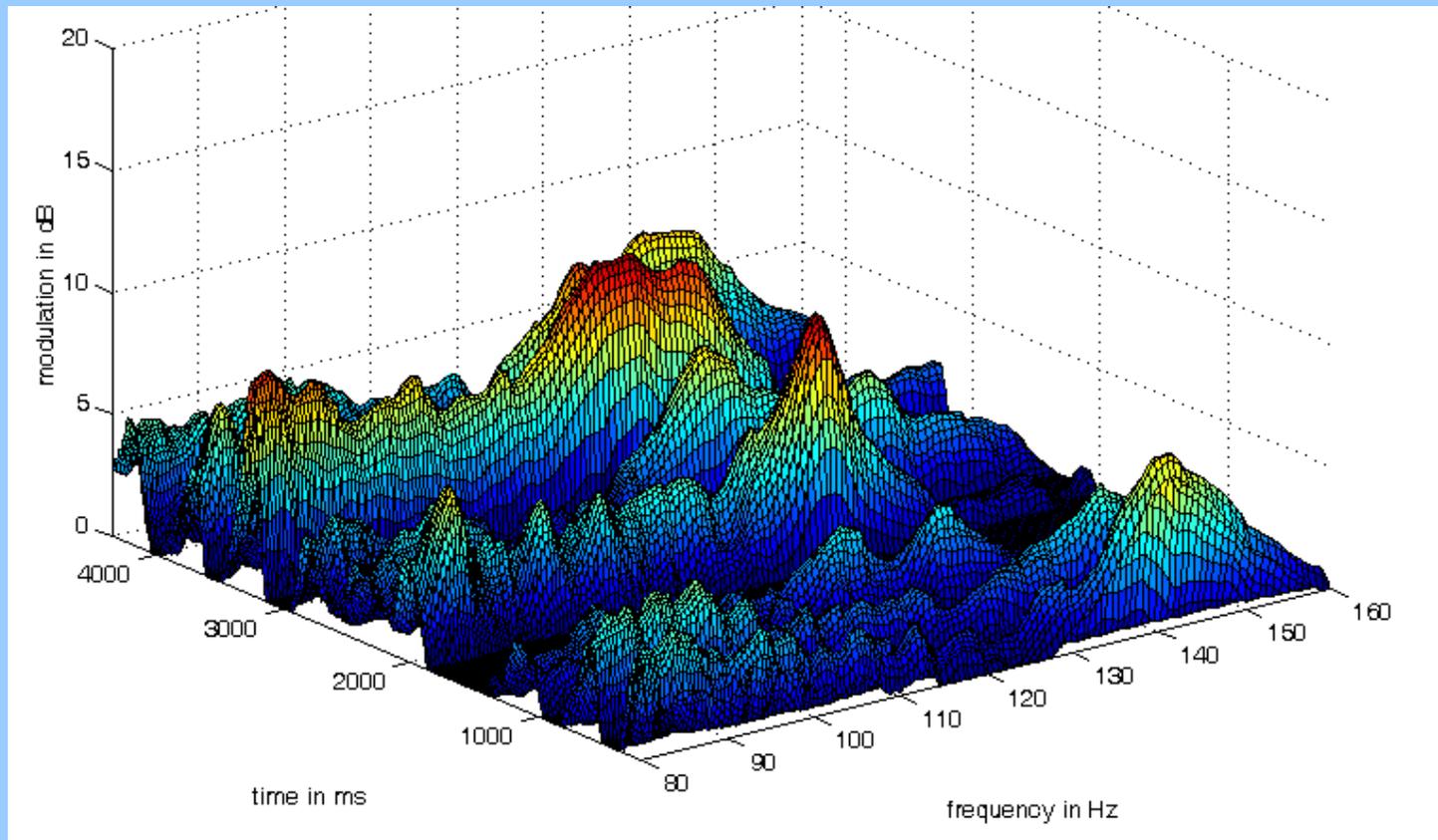


The binaural audio sounds clear and close.



If we convolve speech with a binaural reverberation of 2 seconds RT, and a direct/reverberant ratio of -10dB the pitch discrimination is reduced – but still pretty good!

Speech with reverberation: RT=1s, D/R -10dB



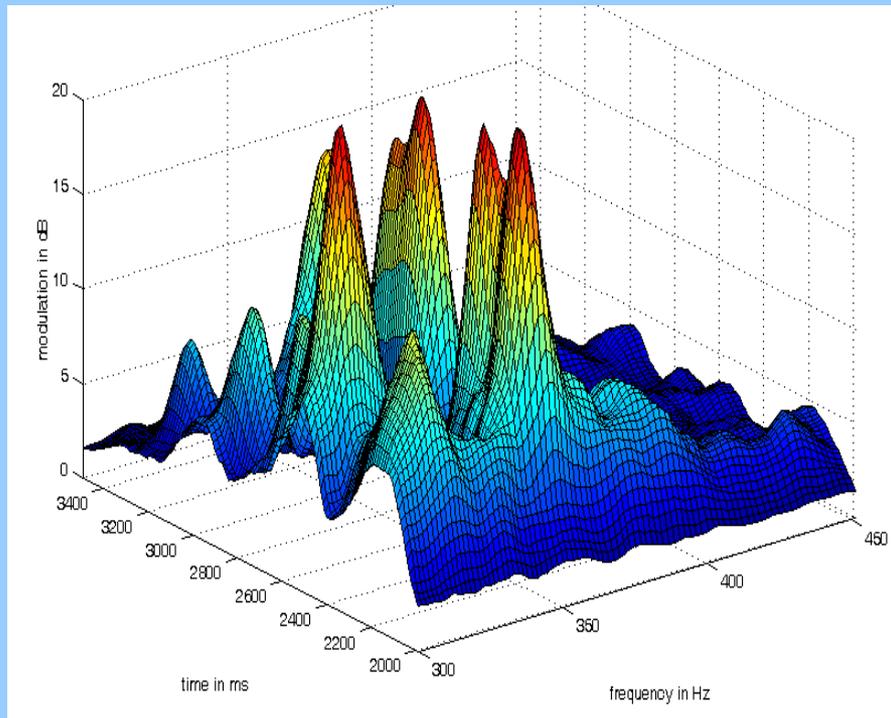
The binaural audio sounds distant and muddy.



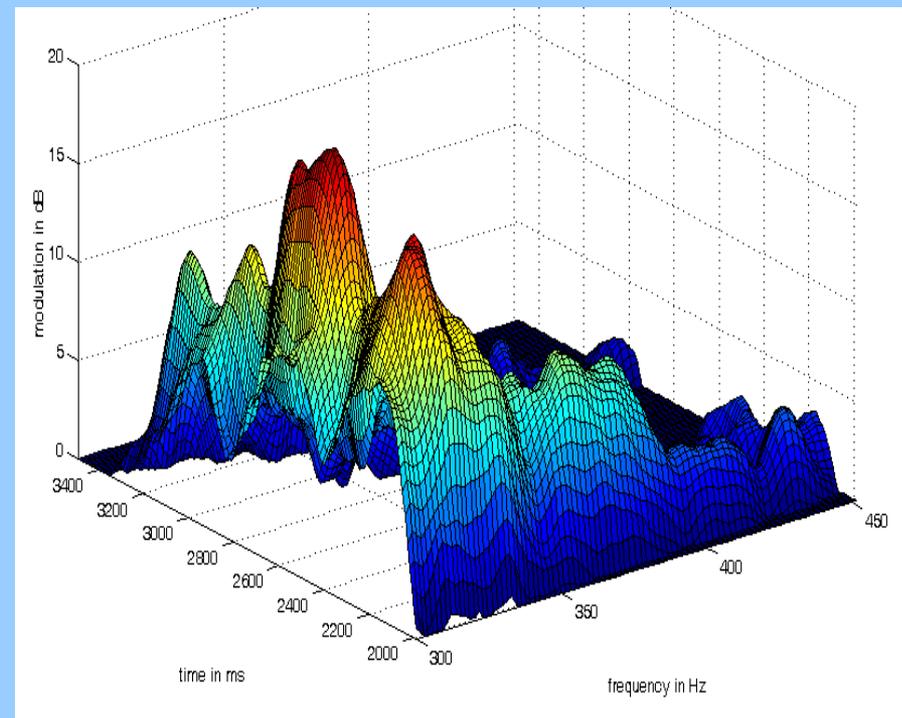
When we convolve with a reverberation of 1 seconds RT, and a D/R of -10dB the brief slides in pitch are no longer audible – although most of the pitches are still discernable, roughly half the pitch information is lost.

This type of picture could be used as a measure for distance or engagement.

Two violins recorded binaurally, ± 15 degrees azimuth



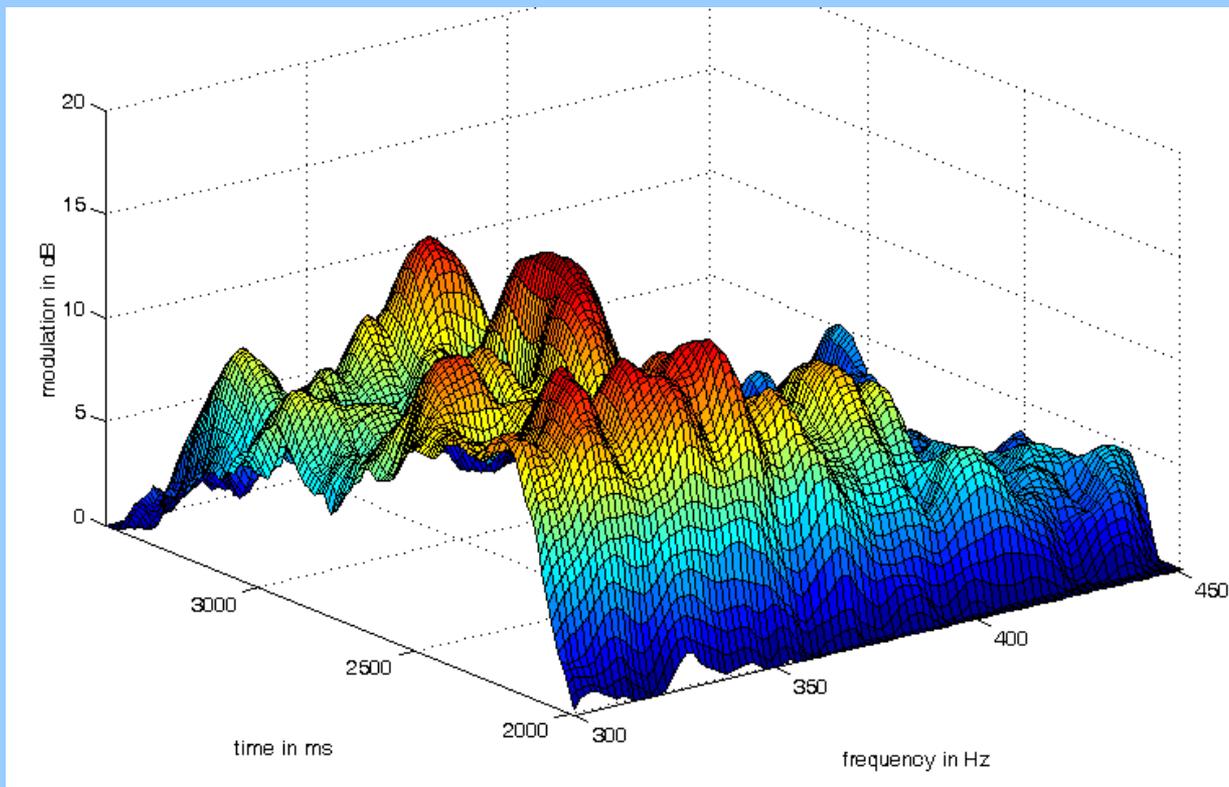
Left ear - middle phrase



Right ear - middle phrase

Note the huge difference in the ILD of the two violins. Clearly the lower pitched violin is on the right, the higher on the left. Note also the very clear discrimination of pitch. The frequency grid is 0.5%

The violins in the left ear – 1s RT D/R -10dB



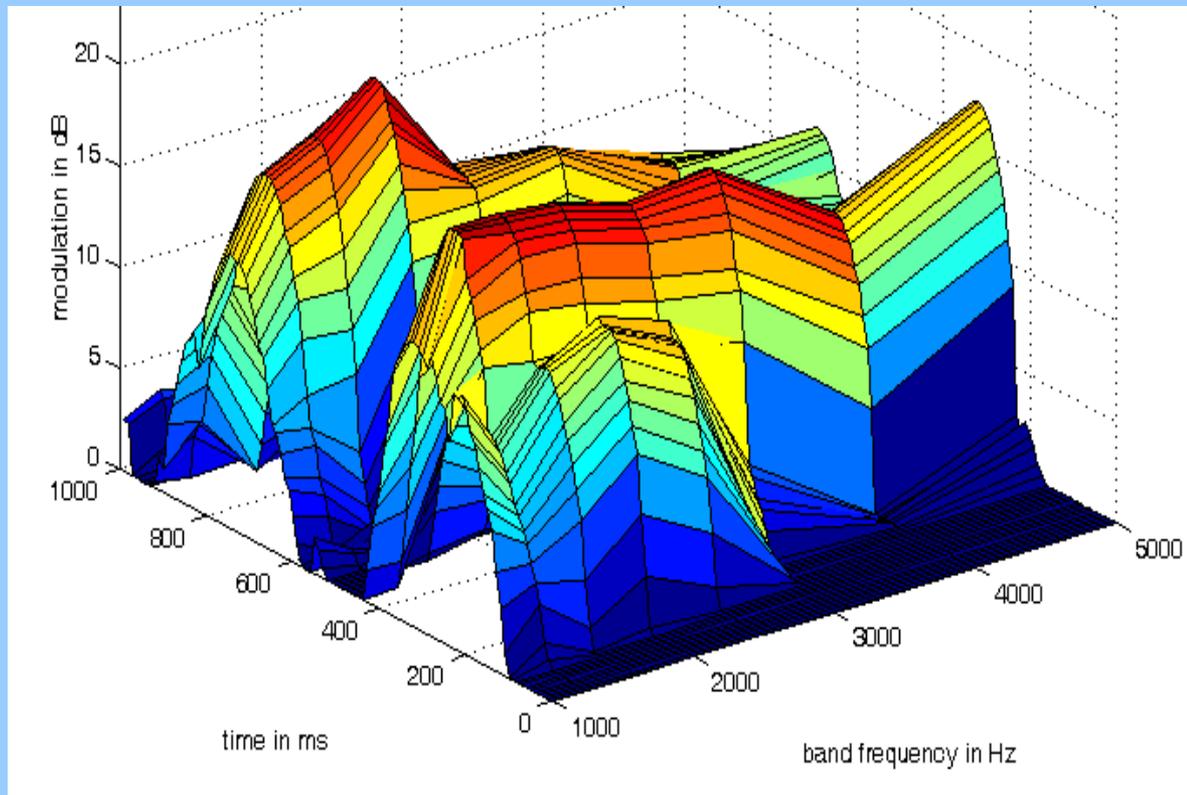
When we add reverberation typical of a small hall the pitch acuity is reduced – and the pitches of the lower-pitched violin on the right are nearly gone. But there is still some discrimination for the higher-pitched violin on the left.

Both violins sound muddy, and the timbre is poor!

Timbre – plotting modulations across critical bands

- Once sources have been separated by pitch, we can compare the modulation amplitudes at a particular frequency across each 1/3 octave band, from (perhaps) 500Hz to 5000Hz.
- The result is a map of the timbre of that particular note – that is, which groups of harmonics or formant bands are most prominent.
- This allows us to distinguish a violin from a viola, or an oboe from a clarinet.
- I modified my model to select the most prominent frequency in each 10ms time-slice, and map the amplitude in each 1/3 octave band for that frequency.
- The result is a timbre map as a function of time.
 - The mapping works well if there is only one sound source.

Timbre map of the syllables “one two”

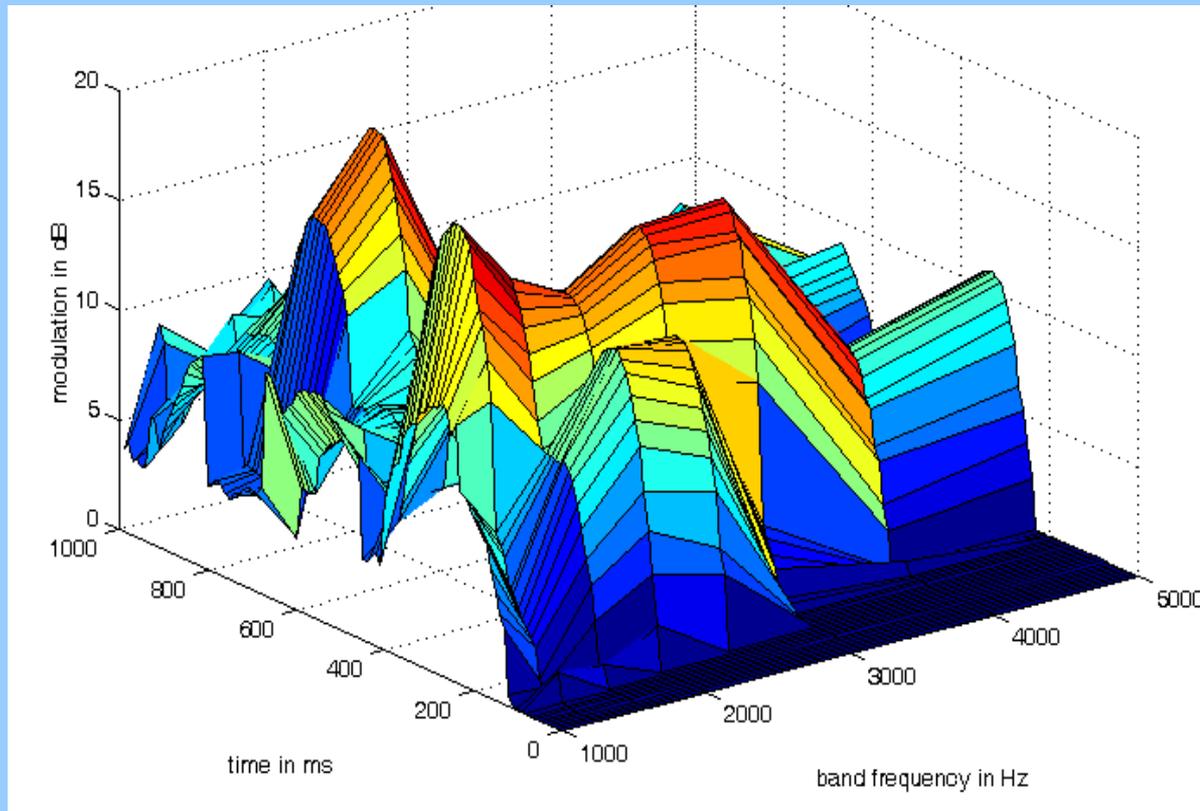


All bands show moderate to high modulation, and the differences in the modulation as a function of frequency identify the vowel.

Note the difference between the “o” sound and the “u” sound.



Timbre map of the syllables “one two” with reverberation 2s RT -10dB D/R

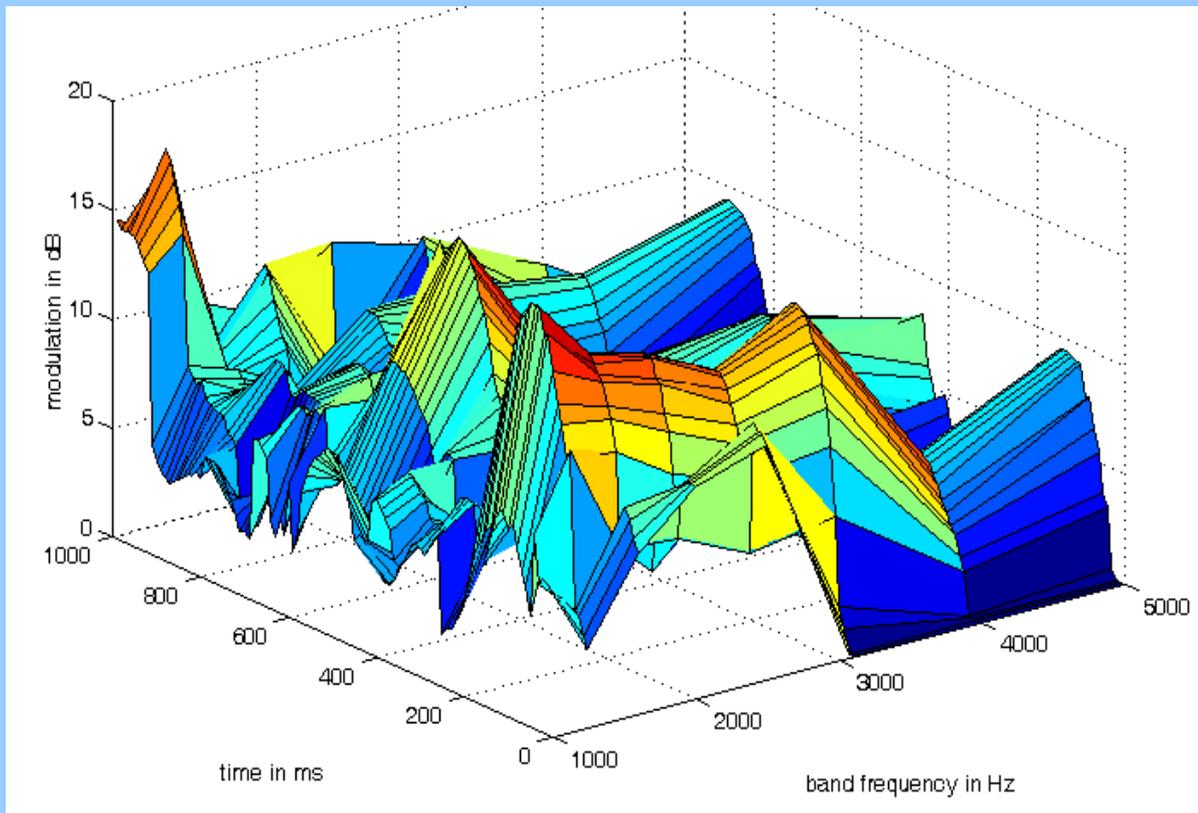


All bands still show moderate to high modulation, and the differences in the modulation still identify the vowel.

The difference between the “o” sound and the “u” sound is less clear, but still distinguishable.



Timbre map of the syllables “one two” with reverberation 1s RT -10dB D/R



The clarity of timbre is nearly gone.

The reverberation has scrambled enough bands that it is becoming difficult (although still possible) to distinguish the vowels.



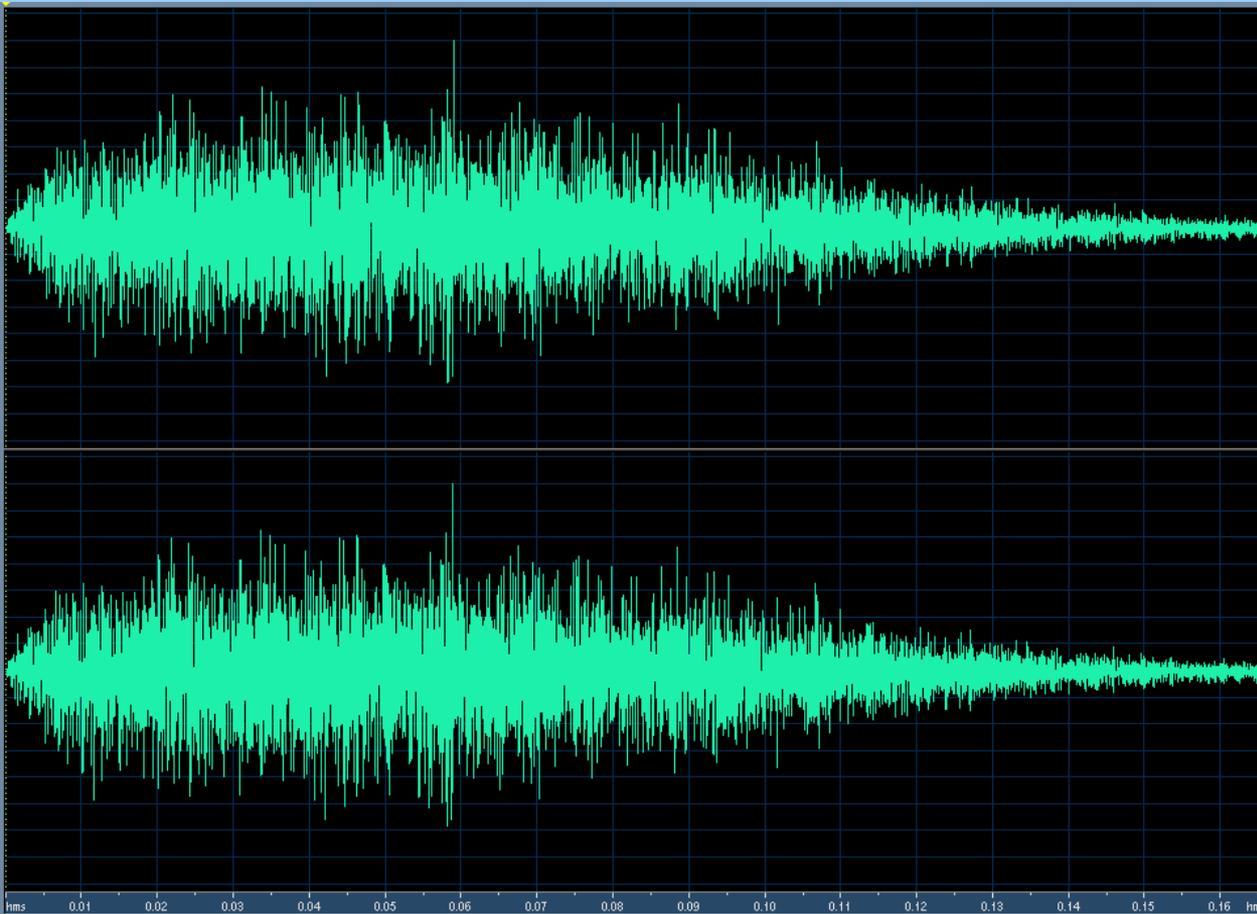
A one-second reverberation time creates a greater sense of distance than a two second reverberation because more of the reflected energy falls inside the 100ms frequency detection window.

Non-coherent sources

- So far I have been considering only sources that emit complex tones with a distinct pitch.
 - What about sources that are not coherent, like a modern string section with lots of vibrato, or pink noise?
- Nearly any sound source – when band-limited – creates noise-like modulations in the filtered output.
 - Pink noise is no exception. Narrow-band filter it, and the amplitude fluctuates like crazy.
- Sources of this type cannot be separated by frequency into separate streams – but they can be sharply localized, both by ITD and ILD.
- This explains why in a good hall we can easily distinguish the average azimuth of a string section.
- If the strings play without vibrato they are perceived as a single instrument, with no apparent source width!

Example – Pink noise bursts with identical ILDs

- I created a signal that consists of a series of pink noise bursts, one of which is shown below. The noise is sharply high pass filtered at 2kHz.



During the 10ms rise-time the noise is identical in the left and right channels.

After 10ms, the noise in the right channel is delayed by 100us.

The next burst in the series is identical, but the left and right channels are swapped.

When you listen to this on headphones (or speakers) the sound localizes strongly left and right.



Azimuth is determined by the ITDs of the modulations – not the onset

Summary of part 1:

- We have shown that the human ear has evolved to analyze fluctuations or modulations in the amplitude of the basilar membrane motion at frequencies above 1000Hz.
 - And not necessarily the average amplitude of the motion.
 - So long as the phases of the harmonics that create the modulations are not altered by reflections, the modulations from each source can be separated by frequency, and separately analyzed for pitch, timbre, azimuth, and distance.
- The modulations – especially when separated – carry more information about the sound sources than the fundamental frequencies.
 - And allow precise determination of pitch, timbre, and azimuth.
- All of these perceptions depend on the ear's ability to perceive the direct sound from the source!!! (And there is currently no standard measure...)
- Reflections from any direction – particularly early reflections – scramble these modulations and create a sense of distance and disengagement.
 - But they are only detrimental to music if they are too early, and too strong.
- The C language model presented above makes it possible to visualize the degree to which timbre and pitch can be discerned in a recording of live music.
 - With calibration a single-number measure for engagement should be possible.

Direct sound and Envelopment

- Recent work by the author in both experiments with several subjects, and in live lecture demonstrations with loudspeakers, has shown that the sense of both reverberance and envelopment increases when the direct sound is separately perceived.
 - Where there is no perceivable direct sound the sound can be reverberant, but comes from the front.
 - When the direct sound is above the threshold of localization the reverberation becomes louder and more spacious.
 - Envelopment and reverberance are created by late energy – at least 100ms after the direct sound.
 - When the direct sound is inaudible the brain cannot perceive when a sound has started.
 - So effectively the time between the onset of the direct sound and the reverberation is reduced, and less reverberation is heard.
 - In the absence of direct sound syllabic sound sources (speech, woodwinds, brass, solo instruments of all kinds) are perceived as in front of the listener, even if reflections come from all around.
 - The brain will not allow the perception of a singer (for example) to be perceived as all around the listener.
 - In addition, Barron has shown that reverberation is always stronger in front of a hall than in the rear – so in most seats sound decays are perceived as frontal.
 - But when direct sound is separately perceived, the brain can create two separate sound streams, one for the direct sound (the foreground) and one for the reverberation (the background).
 - A background sound stream is perceived as both louder and more enveloping than the reverberation in a single combined sound stream.

Time for Demos

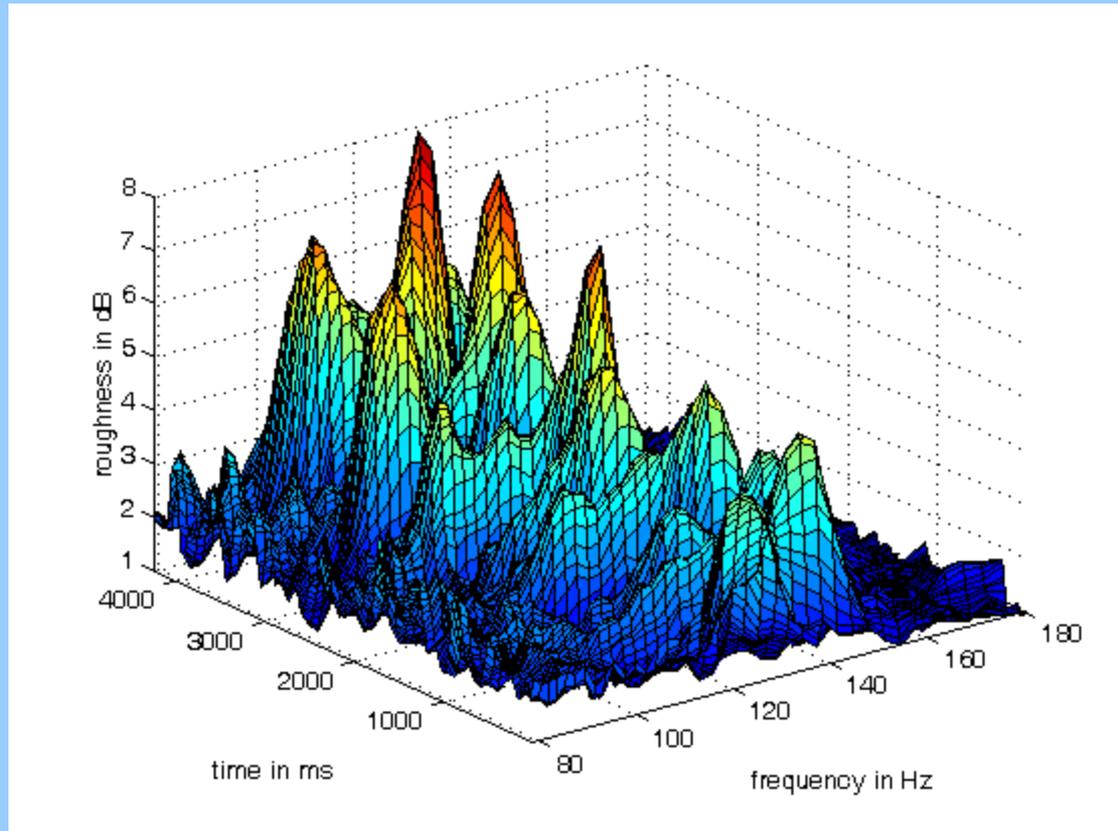
- I will attempt to demonstrate the effects of the direct sound on the perception of distance, muddiness, and envelopment.
- The four speakers around the audience will play reverberation as it might exist in a hall.
- The center channel will produce the direct sound.
- The D/R will vary depending on where you are sitting.
 - And I will vary it to give everyone a chance to hear the effects near the threshold of audibility for the direct sound.



Part 2: Near, Far, and Engagement

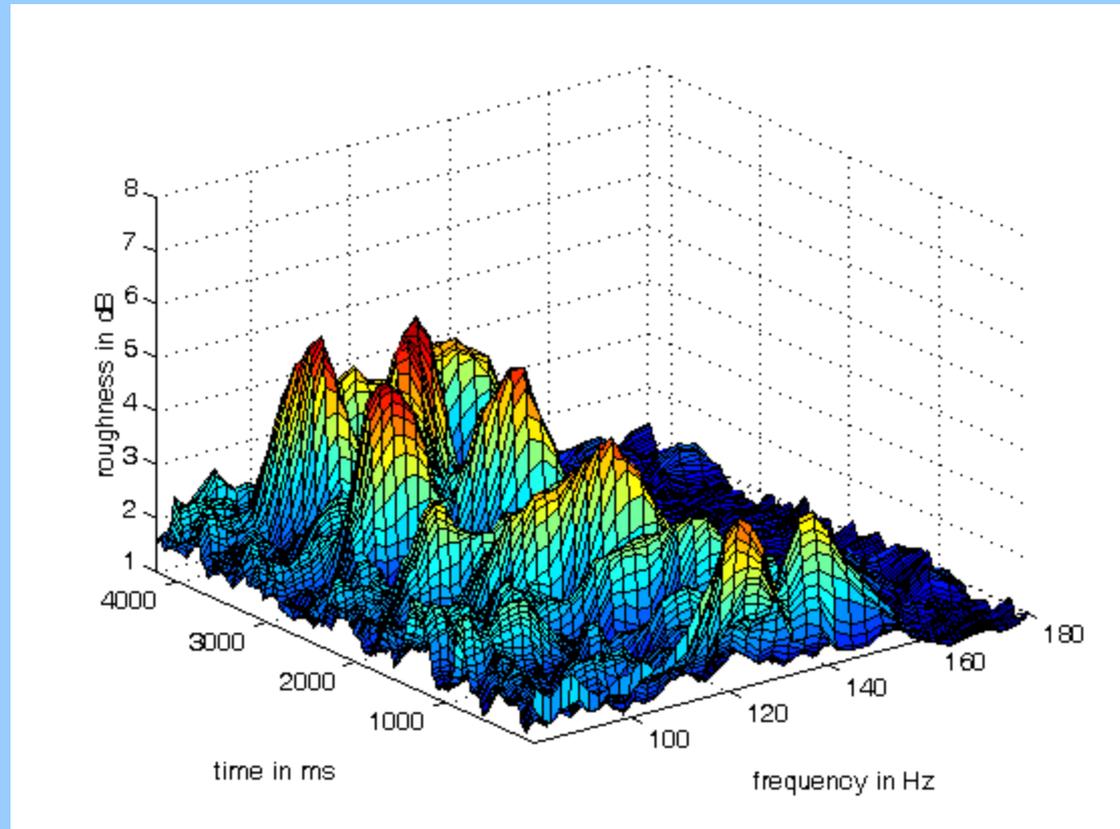
- The apparent closeness of a sound source is a fundamental perception for all of us.
 - We can tell instantly if a person talking is within a few feet of us, or further away – and this perception has survival value.
 - The perception of “Near” depends critically on our ability to perceive the direct sound – the sound that travels to the listener without reflecting.
 - Surprisingly, in a theater or hall it is possible to perceive the performers as both acoustically close to the listener *and* enveloped by the hall.
 - The best halls (Boston Symphony Hall, Concertgebouw, the front half of the Musikverein) provide both, but many, perhaps most, provide only reverberation.
- Harmonic coherence of speech and music is a principle cue for perceiving near and far.
 - The audio examples in the click box above show the decrease in apparent distance caused by increasing amounts of harmonic coherence.
 - Note that all of the examples have high intelligibility – but their emotional effect is quite different.
- The perception of “near” and “far” correlates with ability to localize sound sources – and the ability to separately hear several musical lines at the same time.

Amplitude modulation analysis – direct sound



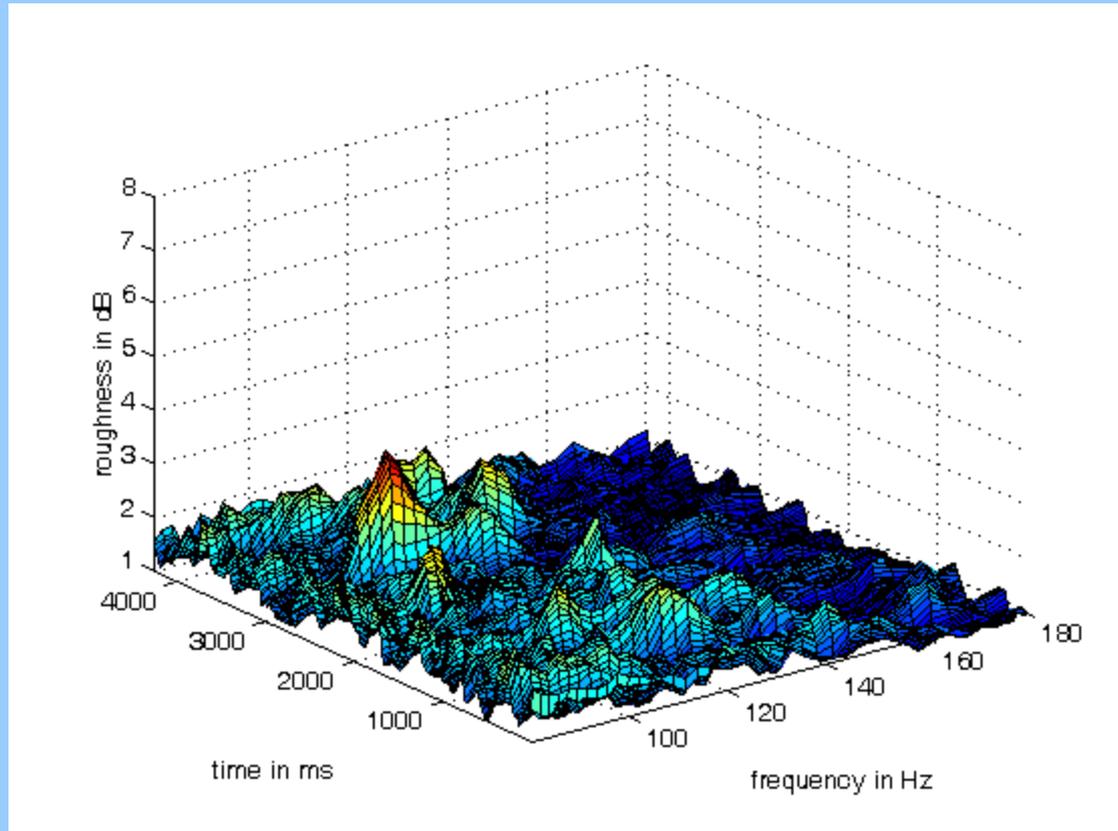
Spoken syllables “one” to “ten” analyzed by the neural model presented in Part one. Note the clear detection of the pitch of each vowel

Amplitude modulation analysis “one” to “ten” with 88ms all-pass reflections



Note the pitch acuity has been reduced, along with the signal to noise ratio.

Amplitude modulation analysis “one” to “ten” with 133ms reflections



The pitch discrimination with this analysis model (an early one) is poor with these reflections.

Experiences – Staatsoper Berlin

Barenboim gave Albrecht Krieger and me 20 minutes to adjust the LARES system in the Staatsoper.

My initial setting was much too strong for Barenboim. He wanted the singers to be absolutely clear, with the orchestra rich and full – a seemingly impossible task.

Adding a filter to reduce the reverberant level above 500Hz by 6dB made the sound ideal for him.

The house continues with this setting today for every opera.

Ballet uses more of a concert hall setting – which sounds amazingly good.



In this example the singers have high clarity and presence. The orchestra is rich.



Experiences – Bolshoi – a famously good hall for opera

The Bolshoi is a large space with Lots of velvet.

RT is under 1.2 seconds at 1000Hz, and the sound is very dry.

Opera here has enormous dramatic intensity – the singers seem to be right in front of you – even in the back of the balconies. It is easy for them to overpower the orchestra



This mono clip was recorded in the back of the second balcony.



In this clip the orchestra plays the reverberation. The sound is rich and enveloping



New Bolshoi before modification



This theater suffers greatly from having the old Bolshoi next door!
(The theater has received poor reviews from musicians and the press.)

The Semperoper was the primary model for the design of the new Bolshoi. As in Dresden the sound on the singers is distant and muddy, and the orchestra is too loud.

RT ~1.3 seconds at 1000Hz.

New Bolshoi 

Dresden 

What is it about the **SOUND** of this theater that makes the singers seem so far away?

Experiences – Amsterdam Muziektheater

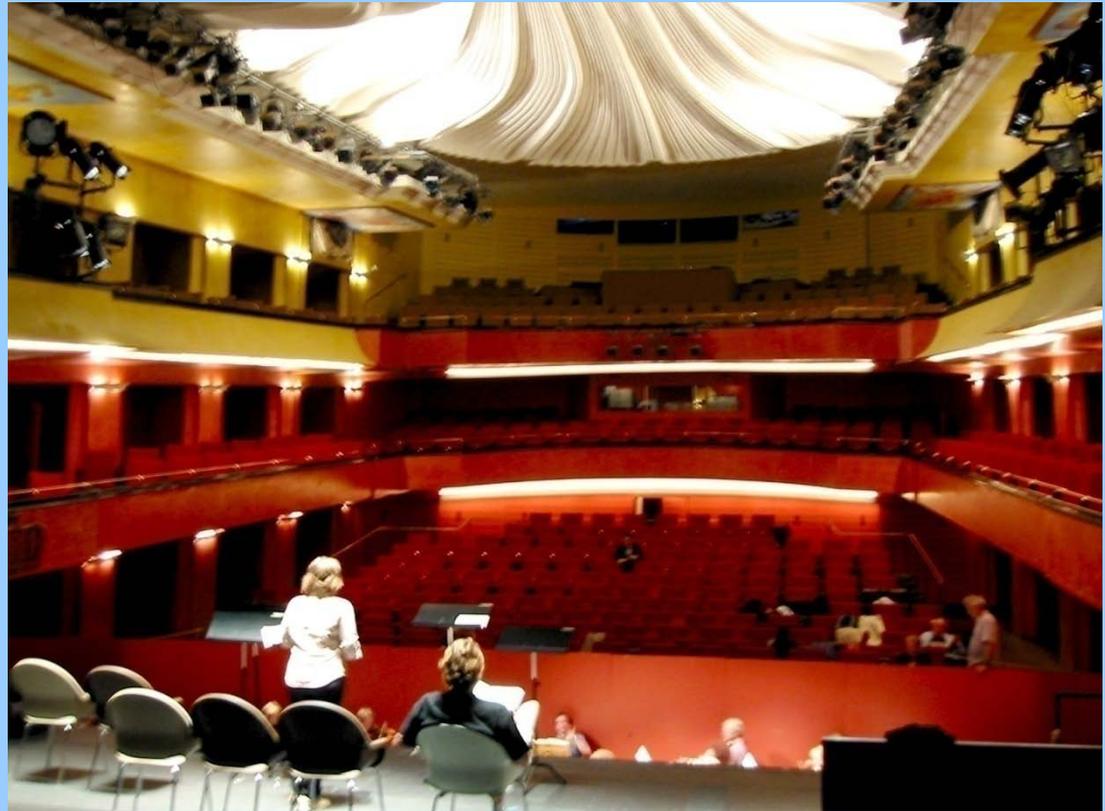
- Peter Lockwood and I spent hours adjusting the reverberant level using a remote in the hall.
 - He taught me to hear the point where the direct sound becomes no longer perceptible, and the sonic distance dramatically increases.
 - With a 1/2 dB increase in reverberant level, the singer moved back 3-4 meters.
 - In Copenhagen, I once decreased the D/R by one dB while Michael Schonwandt was conducting a rehearsal. He immediately waved to me from the pit, and told me to put it back.
- Given a chance to listen A/B, these conductors choose dramatic intensity over reverberance.
 - When they do not have this chance, reverberation is seductive, and the singers be damned!

Experiences, Copenhagen New Stage

We were asked to improve loudness and intelligibility of the actors in this venue.

64 Genelec 1029s surround the audience, driven by two line array microphones, and the LAREAS early delay system. A gate was used to remove reverberation from the inputs.

5 drama directors listened to a live performance of Chekhov with the system on/off every 10 minutes.



The result was unanimous – “it works, we don’t like it.” “The system increases the distance between the actors and the audience. I would rather the audience did not hear the words than that this connection is compromised.”

A slide from Asbjørn Krokstad - IoA, NAS Oslo 2008

[With permission]

To succeed: [in bringing new audience into concert halls...]

ENGAGING



“Interesting” “Nice”

[We need to make the sonic impression of a concert engage the audience – not just the visual and social perceptions. Especially since audiences are increasingly accustomed to recordings!]

ENGAGEMENT, not *NICE*

- At the IoA conference in Oslo, Asbjørn Krokstad (a musician, conductor, and Norway's best-known acoustician) gave a lecture where he insisted that acousticians needed to provide *engagement*, not just pleasant music.
 - And not just for drama and opera, but for chamber music and symphony too.
 - At the end of the lecture he showed a picture of the Teatro Colón in Buenos Aires, Argentina. “Is this the concert hall of the future” he asked?
 - This hall is not a shoebox, but a large semicircular theater with a high ceiling. It ranks at the top in Beranek's surveys, and the reverberation time is 1.6 seconds occupied.
 - Krokstad may have conducted there.
- Engagement requires the independent perception of the direct sound
- We must learn how to provide this essential element in halls.
- I have been fortunate to hear several of the live broadcasts of the Metropolitan Opera in a good theater. For example, the performance of Salome:
 - The sound was harsh and dry – radio mikes coupled to directional loudspeakers. But you could hear every syllable of Mattila's impeccable German. The performance was totally gripping!
 - This is the dramatic and sonic experience audiences increasingly demand.

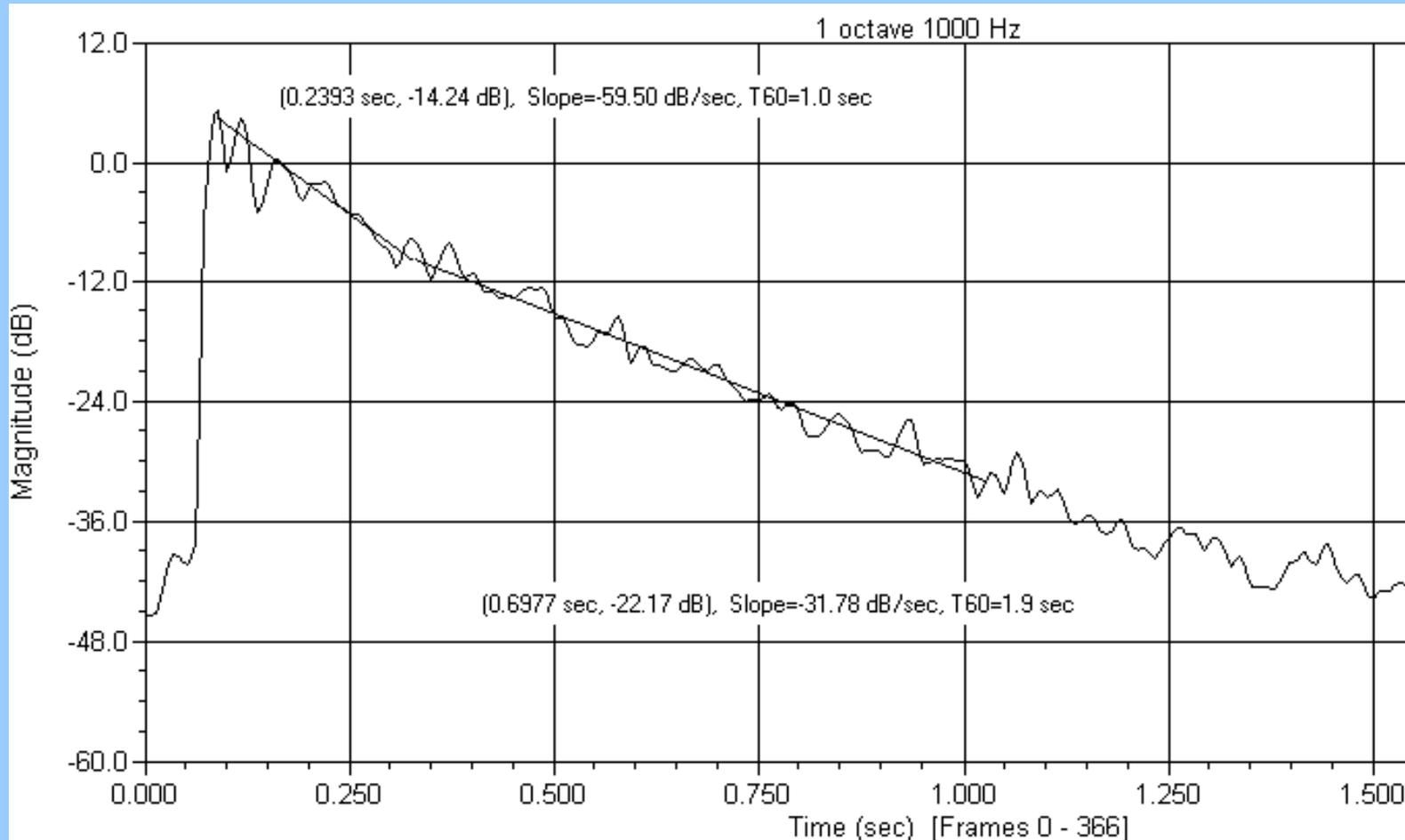
Part 3 - Main Points

- The ability to hear the *Direct Sound* – as measured by LOC or through binaural recording analysis – **is a vital component of the sound quality in a great hall.**
 - The ability to separately perceive the direct sound when the D/R is less than 0dB requires *time*. When the d/r ratio is low there must be sufficient time between the arrival of the direct sound and the build-up of the reverberation if engagement is to be perceived.
- Hall shape does not scale
 - Our ability to perceive the direct sound – and thus localization, engagement, and envelopment - depends on the direct to reverberant ratio (d/r), and on the rate that reverberation builds up with time.
 - Both the direct to reverberant ratio (d/r) and the rate of build-up change as the hall size scales – but human hearing (and the properties of music) do not change.
 - **A hall shape that provides good localization in a high percentage of 2000 seats may produce a much lower percentage of great seats if it is scaled to 1000 seats.**
 - **And a miniscule number of great seats if it is scaled to 500 seats.**

Frequency-dependent diffusing elements are necessary, and they do not scale.

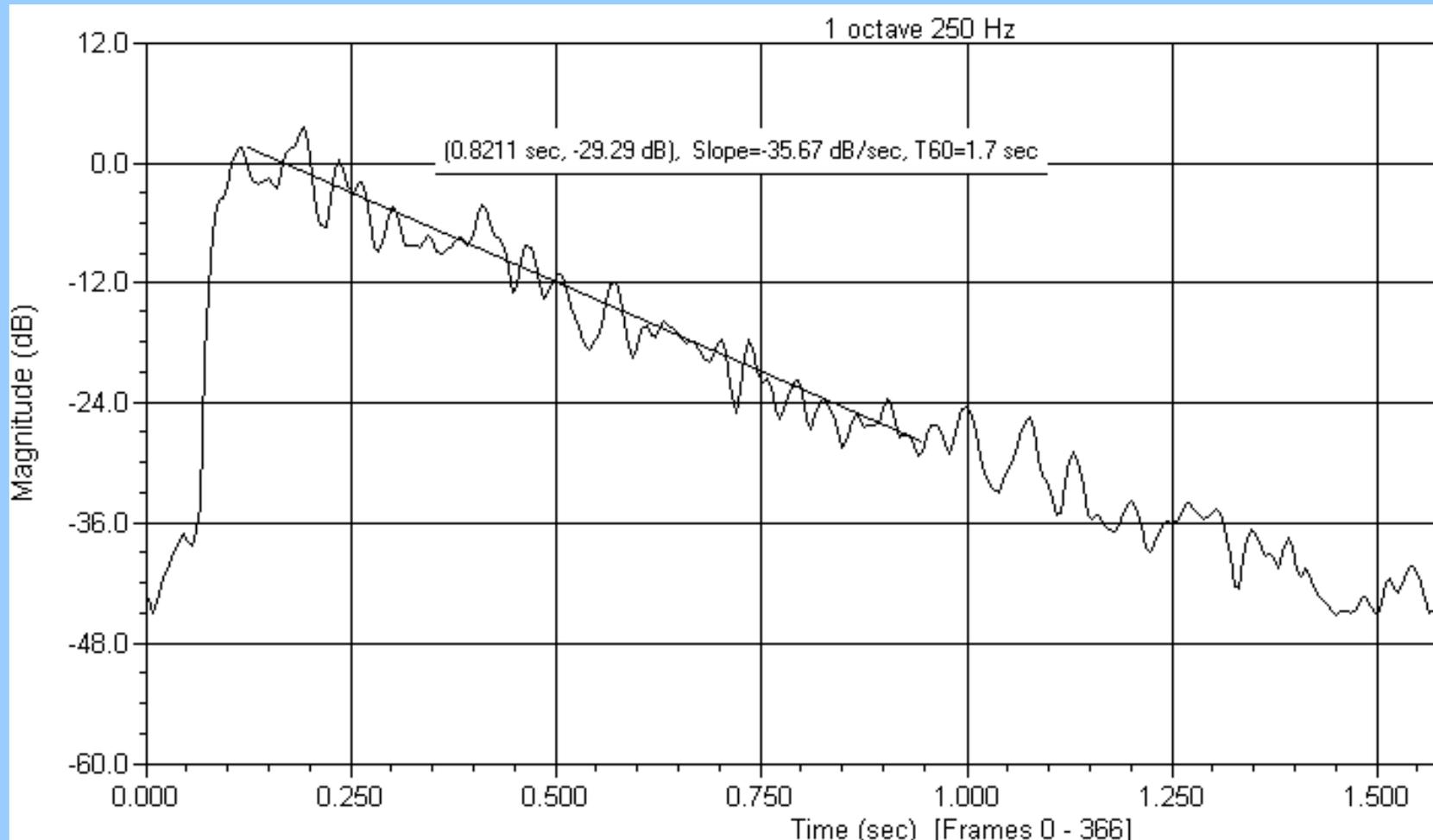
- The audibility of direct sound, and thus the perceptions of both localization and engagement, is frequency dependent. Frequencies above 700Hz are particularly important.
 - Frequency dependent diffusing elements can cause the D/R to vary with frequency in ways that improve direct sound audibility.
 - The best halls (Boston, Amsterdam, Vienna) all have ceiling and side wall elements with box shape and a depth of ~0.4m.
 - These elements tend to send frequencies above 700Hz back toward the orchestra and the floor, where they are absorbed. (The absorption only occurs in occupied halls – so the effect will not show up in unoccupied measurements!)
 - The result is a lower early and late reverberant level above 700Hz in the rear of the hall.
 - This increases the D/R for the rear seats, and improves engagement.
 - The LOC equation is sensitive to all reflections in a 100ms window, which will include many second-order reflections, especially in small halls.
- Replacing these elements with smooth curves or with smaller size features does not achieve the same result.
 - Some evidence of this effect can be seen in RT and IACC80 measurements when the hall and stage are occupied.
- Measurements in Boston Symphony Hall (BSH) above 1000Hz show a clear double slope that is not visible at 500Hz.
 - The hall has high engagement in at least 70% of the seats.

Boston Symphony Hall, occupied hall and stage, stage middle to front of first balcony, 1000Hz



Note the clear double-slope decay, with the first 12dB decaying at RT = 1s
The direct sound is clearly dominant at this frequency in this seat. The sound is very good – Leo Beranek’s favorite seat!

Boston Symphony Hall, occupied, stage to front of balcony, 250Hz



But at 250Hz, there is no evidence of the direct sound. It has been swamped by reverberation.

We need better measures

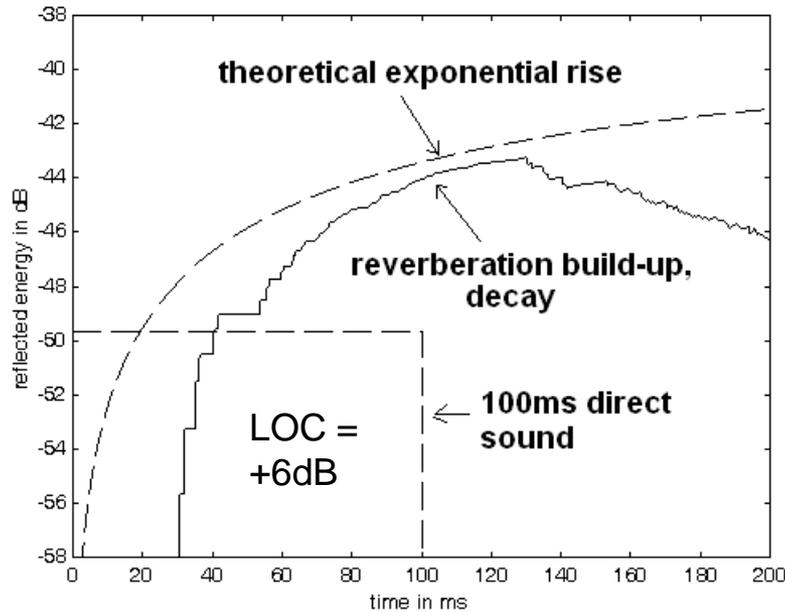
- Current acoustic measures ignore *both* the D/R and the time gap between the direct (the first wavefront) and the reverberation.
 - RT, C80, and EDT all ignore the strength of the direct sound and the effects of musical style on the audibility of the D/R.
 - IACC comes close, but measures only lateral reflections.
- LOC and my hearing model attempt to supply a simple measure for a basic human perception which depends on direct sound.
 - Impulse response measurements under occupied conditions are notoriously difficult to obtain.
 - But this problem can be circumvented through binaural recordings of live sound.
- The hearing model presented in part one promises to supply measures that use binaural recordings of actual performances as inputs.
 - And the ability to listen to these recordings to test the validity of these measures against the true experience.

Why do large halls sound different?

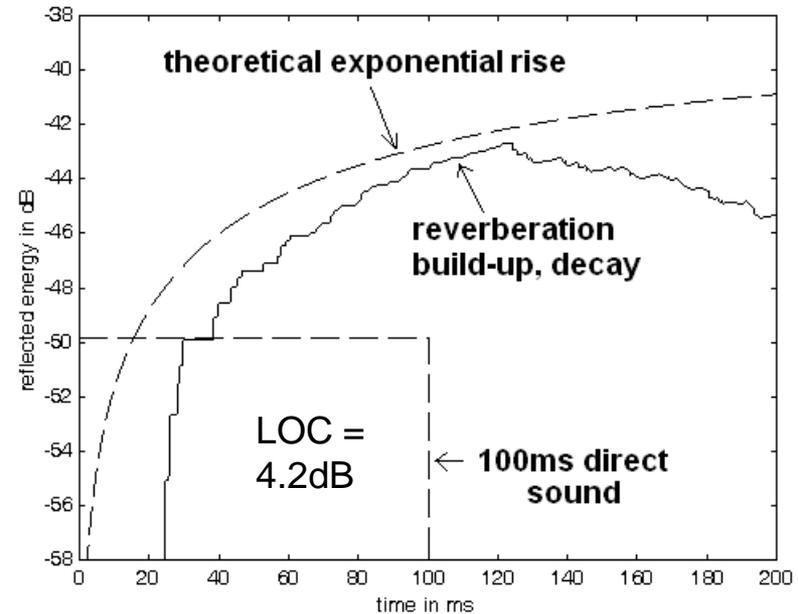
- In Boston Symphony Hall (BSH), and the Amsterdam Concertgebouw (CG) the reverberation decay is nearly identical, but the halls sound different.
 - The difference can be explained using the same model that was used to develop LOC.
 - Lacking good data with an occupied hall and stage I used a binaural image-source model with HRTFs measured from my own eardrums.

Reverberation build-up and decay – from models

Amsterdam



Boston



The seat position in the model has been chosen so that the D/R is -10dB for a continuous note.

The upward dashed curve shows the exponential rise of reverberant energy from a source with rapid onset and at least 200ms length. The reverberation for the dotted line is exponentially decaying noise with no time gap. The solid line shows the image-source build up and decay from a short note of 100ms duration. Note the actual D/R for the short note is only about -6dB.

The initial time gap is less in Boston than Amsterdam, but after about 50ms the curves are nearly identical. (Without the direct sound they *sound* identical.) Both halls show a high value of LOC, but the value in Amsterdam is significantly higher – and the sound is clearer.

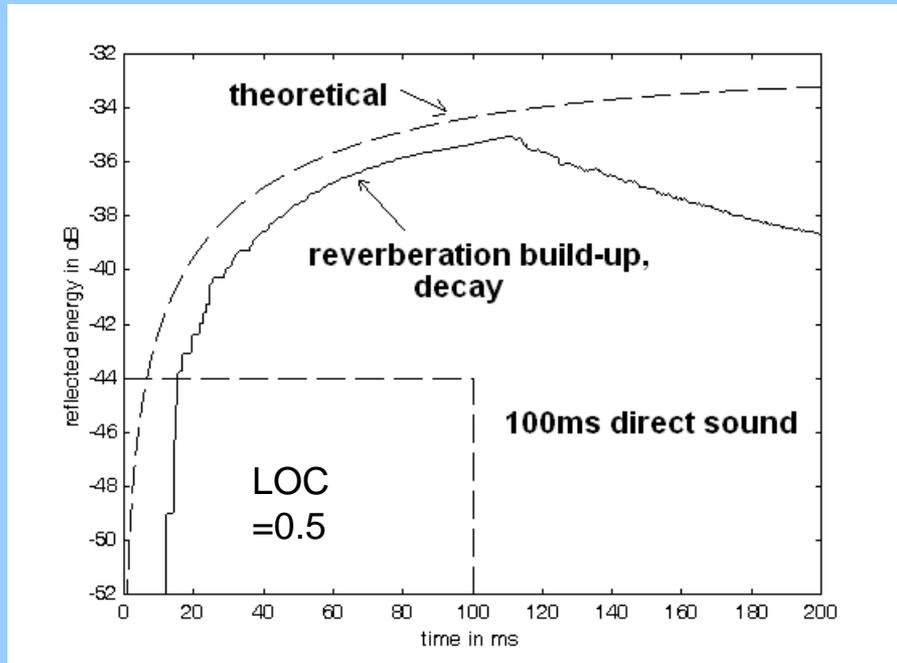
Comparisons of C80, C50, IACC80, and LOC

- Conventional measures for the models of Amsterdam Concertgebouw and Boston Symphony Hall give the following results:
- Amsterdam: C80 = .43dB, C50 = -2.8dB, IACC80 = .38, LOC = +6dB
- BSH: C80 = .65dB, C50 = -2.1dB, IACC80 = .22, LOC = +4.2dB
- Half-Size BSH: C80 = 3.7, C50 = 1.7, IACC80 = .15, LOC = 0.5dB
- Only the IACC80 shows that Amsterdam might have more direct sound than Boston. The standard Clarity measures predict the opposite – and predict that the small hall would have high clarity, and it does not.
- But IACC80 is sensitive only to *lateral* reflections. Strong reflections from the front, overhead, or rear do not affect IACC.
- An IACC of 0.22 would usually be considered too low for good sound. In spite of this BSH has both clarity and good localization in this seat.

Smaller halls

- What if we build a hall with the shape of BSH, but half the size?
 - The new hall will hold about 600 seats.
 - The RT will be half, or about 1 second.
 - We would expect the average D/R to be the same. Is it? How does the new hall sound?
 - If the client specifies a 1.7s RT will this make the new hall better, or worse?

Half-Size Boston



The gap between the direct and the reverberation and the RT have become half as long.

Additionally, in spite of the shorter RT, the D/R has decreased from about -6 in the large BSH model, to about -8.5 in the half-size model.

This is because the reverberation builds-up quicker and stronger in the smaller hall.

The direct sound, which was distinct in more than 50% of the seats in the large hall will be audible in fewer than 30% of the seats in the small hall.

If the client insists on increasing the RT by reducing absorption, the D/R will be further reduced, unless the hall shape is changed to increase the cubic volume.

The client and the architects expect the new hall to sound like BSH – but they, and the audience, will be disappointed. As Leo Beranek said about the Berlin Philharmonie: “They can always sell the bad seats to tourists.”

Great Small Halls Exist!



Jordan Hall at New England Conservatory has 1200 seats, an RT of 1.3s fully occupied. The shape is half-octagonal, with a high ceiling.

The audience surrounds the stage, with a single high balcony. The average seating distance is much shorter than a shoebox hall, increasing the direct sound.

The high internal volume allows a longer RT with low reverberant level.

The sound in nearly every seat is clear and direct, with a marvelous surrounding reverberation. But the stage house is deep and reverberant. Small groups always play far forward.

Although the hall is renowned as a chamber music hall, it is also good for small orchestras and choral performances. It was built around 1905.

The hall is in constant use – with concerts nearly every night, (and many afternoons.)

Williams Hall, NEC

- Williams hall, in the same building, has ~350 seats in a square plan with a high ceiling.
- The sound from a piano sound is clear and reverberant in most, if not all, seats.



(The audience usually sits where the orchestra is rehearsing in this picture.)

The square plan keeps the average seating distance low.

The high ceiling and high single balcony provides a long RT without a high reverberant level.

The absorbent stage eliminates strong reflections from the back wall. By absorbing at least half the backward energy from the musicians, the stage increases the d/r.

Note the coffered ceiling – similar to BSH.



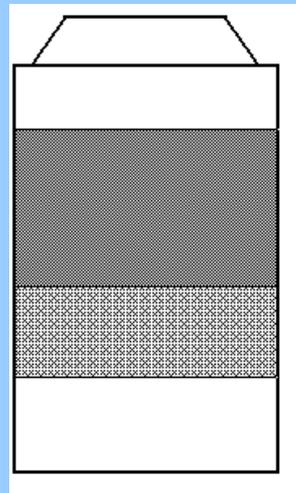
Hard learned lessons

- Where clarity is a problem in small halls, acousticians usually recommend adding early reflections – through a stage shell, side reflectors, etc.
 - We tried this in a small hall by placing plywood panels behind the piano. The sound became louder and less clear. Just the opposite of what was needed.
- These measures *reduce* the gap between the direct sound and the reflected energy and decrease LOC.
 - They increase loudness – which is usually already too high, while increasing the sense of distance to the performers.
 - A better solution is to add absorption, or perhaps some means of deflecting the earliest reflections to the ceiling, or into the front of the audience where they can be absorbed.
 - Re-direction tricks of this nature do not work well in small halls, as the second and third order reflections they create will arrive within the 100ms window that determines LOC.
 - Small halls have strong direct sound and too many early reflections. The early reflections also come too quickly. Adding more reflections is exactly the wrong thing to do.
 - Adding absorption will improve clarity but reduce the late reverberant level and the RT. Electronics, or more cubic volume, can restore the longer RT without decreasing the D/R.
- In practice, not everyone is aware of, or appreciates, engagement. It is mostly a subconscious perception. Reverberation or resonance is immediately apparent to everyone – which is why it has become so over-emphasized in hall design.
 - Adding absorption may not be appreciated by everyone unless the decrease in late reverberation can be compensated.
 - Such compensation can be surprisingly easy. Adding a few tenths of a second to the late reverberation time of a small hall can be accomplished electronically with very few loudspeakers. The result can be beautiful and completely transparent.

In the best halls the reverberant level is lower than would be expected from classical acoustics

- D/R is frequency dependent in halls, and frequencies above 700Hz are particularly important for engagement.
 - Surface features can be used to decrease the reflected energy level in the rear of the hall at higher frequencies.
- In addition, the distribution of absorption in a hall significantly alters the distribution of the reflected energy.
 - In a good hall absorption is highly non-uniform. A high ceiling with a lot of reflecting surfaces above the audience can increase RT without increasing the reflected energy level near the audience. The reverberation created tends to stay up near the ceiling.
 - This helps to keep the D/R above ~700Hz constant over a large number of seats.
 - Current modeling techniques may not properly calculate these effects.
 - Old fashioned light models might work better...

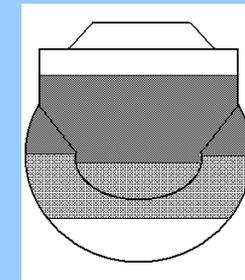
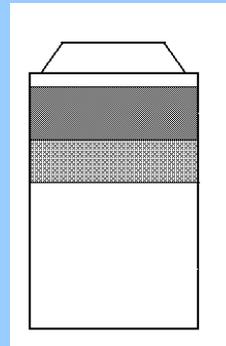
Hall Shapes and direct-sound perception threshold as a function of size



Above
threshold

Near
threshold

Below
threshold



A large hall like Boston has many seats above threshold, and many that are near threshold

If this hall is reduced in size while preserving the shape, many seats are below threshold

It is better to use a design that reduces the average seating distance, using a high ceiling to increase volume.

Boston is blessed with two 1200 seat halls with the third shape, Jordan Hall at New England Conservatory, and Sanders Theater at Harvard. The sound for chamber music and small orchestras is fantastic. RT ~ 1.4 to 1.5 seconds.

Clarity is very high – you can hear every note – and envelopment is good.

Retro reflectors above 1000Hz



Boston, Amsterdam, and Vienna all have side-wall and ceiling elements that reflect frequencies above 1000Hz back to the stage and to the audience close to the stage.

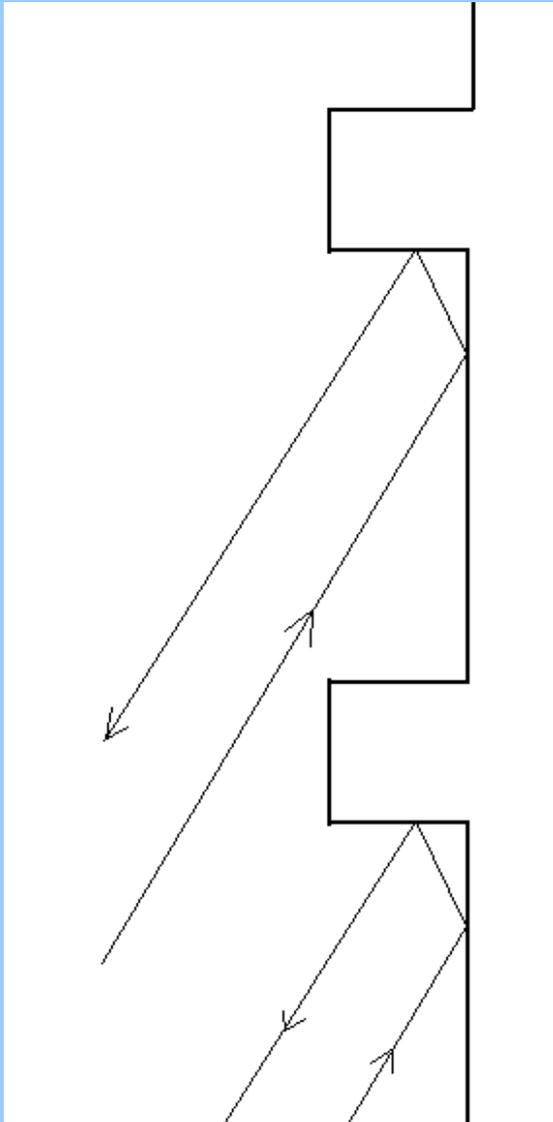
This sound is absorbed – reducing the reverberant level in the rear of the hall without changing the RT.



Another classic example is the orchestra shell at the Tanglewood Music Festival Shed, designed by Russell Johnson and Leo Beranek.

Many modern halls lack these useful features!!!

High frequency retro reflectors



Rectangular wall features scatter in three dimensions – visualize these with the underside of the first and second balconies.

High frequencies are reflected back to the stage and to the audience in the front of the hall.

The direct sound is strong there. These reflections are not easily audible, but they contribute to orchestral blend.

But this energy is absorbed, and thus REMOVED from the late reverberation – which improves clarity for seats in the back of the hall.

Examples: Amsterdam, Boston, Vienna

High frequency overhead filters



A canopy made of partly open surfaces becomes a high frequency filter.

Low frequencies pass through, exciting the full volume of the hall.

High frequencies are reflected down into the orchestra and the audience, where they are absorbed.

Examples: Tanglewood Music Shed, Davies Hall San Francisco

In my experience (and Beranek's) these panels improve Tanglewood enormously. They reduce the HF reverberant level in the back of the hall, improving clarity. The sound is amazingly good, in spite of $RT \sim 3s$.

In Davies Hall the panels make the sound in the dress circle and balcony both clear and reverberant at the same time. Very fine...

(But the sound in the stalls can be loud and harsh.)

The necessity of *occupied* measurements

- The effects of frequency dependent reflecting elements depends on the presence of absorption on the stage and the front of the audience.
- Measuring the halls without absorption in these areas will not detect these vital effects.
- In addition, engagement is highly dependent on the D/R ratio – and this is also not correctly measured in an unoccupied hall.
- Thus measurement of localization and engagement requires that both hall and stage be occupied!

Binaural Measures



The author has been recording performances binaurally for years.

Current technology uses probe microphones at the eardrums.

We can use these recordings to make objective measurements of halls and operas.

The hearing model described in part one can be used to measure the phase coherence in these recordings.

Some demos of eardrum recordings

- These recordings have been equalized for loudspeaker reproduction. You may be able to judge clarity and intelligibility over near-field loudspeakers.
 - Accurate headphone reproduction requires headphone equalization
 - A method of equalizing headphones through equal loudness measurements is described in another paper on the author's web-site.
 - In general large circumaural headphones do not work well even when equalized. On-ear phones, such as the Sennheiser 250 or 100, can work better.
- opera balcony 2, seat 11
 - Moderate intelligibility, reverberant sound.
 - OK for non-Italian speakers with subtitles
- opera balcony 3, seat 12
 - Poor intelligibility, very reverberant
- opera standing room
 - Deep under balcony 2 – good intelligibility
 - This was preferred by Italian speakers
- A concert hall – row 8 (quite close)
 - Very good sound. Not so good further back.



Conclusions

- Performance venues should maximize *engagement* over a wide range of seats, while at the same time providing adequate late reverberation. To achieve this goal the direct sound must be perceived by the brain as distinct from the reflected energy – and this includes early reflections from *all* directions.
 - Engagement is a spatial property that is sensitive to medial reflections. It can be heard with only one ear (and measured with one microphone).
 - But it is essential to measure with both hall and stage *OCCUPIED!*
- The perception of reverberance and envelopment also depends on the audible presence of direct sound.
 - In the presence of adequate late reverberation direct sound increases envelopment and reverberation loudness.
- The audibility of direct sound depends on the D/R ratio above 700Hz, and the time delay of reflections in the first 100ms.
 - Hall sound can often be improved by frequency dependent reflecting elements, or by adding absorption to the stage rear wall.
- The optimum value for the D/R ratio depends on the hall size –
 - The D/R ratio must increase as hall size is reduced if clarity, localization, and the sense of envelopment is to be maintained.
 - D/R and engagement can be increased by decreasing the average seating distance, decreasing the reverberation time, increasing the hall volume, or by careful use of rectangular diffusing elements.
 - This is particularly true in opera houses and halls designed for chamber music.
 - A 1.8 second reverberation time is NOT necessarily ideal in a 1000 seat hall!!! Remember that changes in reverberant *LEVEL* (D/R) and initial time delay are more audible than changes in RT.
- To maintain clarity, low sonic distance, azimuth detection and envelopment in a small hall (and many large halls) it is desirable to reduce the average seating distance, and widely diffuse or absorb the earliest reflections, whether lateral or not.
 - The best small halls do this already.
- Current hall measurements ignore both the D/R and the time delay between direct sound and reverberation. This talk introduces methods to overcome this lack.

Appendix (follow link)

- Slides cut from the original presentation. They may – or may not, be helpful to understanding some of the concepts.