Sound and Audio
vs.
The Ear

James D. (jj) Johnston

Retired audio scientist

Outline – This talk is in two parts, with a break in the middle

- Part 1 – How sound gets to the brain (at normal levels)
  - Sound – how it's created and propagates
  - Acoustics of rooms
  - Acoustics of the head/outer ear
  - Middle ear behavior
  - Cochlear analysis
- Part 2 – What happens after it goes down the auditory nerve?
  - Monaural processing
    - Masking
    - Pre-echo
  - Binaural processing
    - Time correlation,
    - Frequency correlation
    - Masking effects
  - Selective post processing

What is sound in air, physically?

- Sound is variations in air pressure
  - Yes, air pressure varies with time. Yes, that’s very low frequency sound.
  - Sound propagates through air, not all that quickly, either.
    - At STP (0.0 C) sound propagates at 331 meters/second (1.086 kf/s)
    - At 20C (68 F) sound propagates at 343 meters/second (1.125 kf/s)
    - Variation in the speed of sound is nontrivial in any acoustic setting
    - Those speeds of propagation are for dry air, by the way. At altitude, 971 f/s, and in very hot, low-level dry areas, it can reach 1150 f/s
  - These speeds only apply to low levels of sound, say 100dB SPL or below
  - At exceptionally high levels, air is nonlinear, and “speed” turns into a shockwave propagation that can be much faster. You don’t want to live that.

So, 100dB SPL is a LOW level, you say?

- In terms of atmospheric pressure, yes, it is.
  - If air were linear (but it's not), 194dB SPL would be a waveform going from vacuum to 2 atmospheres (give or take a dB due to current atmospheric pressure) (194dB SPL is +/- 1 atmosphere relative to current atmospheric pressure and is commonly only experienced in the military, and is to be avoided there.)
  - Yes, that means that 0dB SPL (we're talking pressure here, remember) is 2*10^-10 atmosphere.
  - 10^(-194/20)   Figure it out yourself. Yes, an undamaged human ear can hear that kind of variation in atmospheric pressure ***within the right frequency range***.

What’s the lowest level a human can hear?

- It’s very hard to determine, but to do that, we have to define the frequency or frequency range of what we’re talking about. Some examples:
  - The atmosphere, which consists of molecules that bounce off everything, including your eardrum, has, therefore, noise level due to the physical nature of air. If we consider the level and spectrum of that noise, given the area of the eardrum, it’s white noise between 6 dB SPL and 9dB SPL in the 20-20kHz bandwidth.
  - Can we hear that? No. We can not, because most of that energy is far outside the most sensitive frequency range. More on that later when discussing the cochlea.
  - Wait, atmospheric pressure can vary by 3-4%. That’s deafening, but we’re not deaf. Again, the frequency range is the issue. Yes, it’s 176 dB SPL <unweighted>, but we simply do not hear .0001Hz frequencies, the eardrum is completely insensitive to them, except to “pop” during abrupt changes.
  - At what level does the air stop being linear? It’s hard to say, it's specific to frequency, humidity, temperature, and so on, but above 120dB SPL or so, linearity is starting to fail, and at 140dB SPL you can discard the assumption of linearity.

So, then, how about “how low does it go?”

- Simply put, it varies by person even with normal hearing (which few people have any longer), but:
  - At ear canal resonance, it’s about -5 to -15 dB SPL in the band directly around the ear canal resonance. That band ranges about ¼ octave, somewhere in the range of 2000 to 4000Hz, and varies by individual. Standard Curves run around -5dB SPL, but lower has been reported.
  - At very low frequencies (say 40Hz), it’s more like 50 dB SPL, but again depending, this time on the individual’s eardrum resonance.
  - At frequencies lower than that, it’s hard to say, because body sensation starts to muddle the question of “is it hearing or is it body sensation”.
  - At very high frequencies, again, for a person with completely undamaged hearing, 10kHz requires at least +10 dB SPL, and for most people over 20 years old or so, that’s probably very optimistic.

So, then, something to take home.

- 6dB SPL is a good estimate of the white noise due to the atmosphere.
  - Remember, you can’t hear that, quite.
  - The only way to get rid of that is to get rid of the atmosphere.
- 120dB SPL is a point at which both the atmosphere and the ear start to go nonlinear
  - You’re better to stick to average levels under 80dB SPL (flat weighted)
  - Yes, percussion is much louder on peaks.
  - Yes, percussion can make you deaf.
- So, now we have a range, of sorts, as to what levels need to be considered.

So, at what frequency?

- Well, the stock answer is 20Hz to 20kHz
  - For most of us in the modern world, that's quite extensive overkill
      - The modern world is loud. It destroys your hearing. Hearing up to 15kHz is good nowadays
      - BUT kids can hear above 20khz. So can a few very few humans who have protected their hearing.
  - But 20Hz? That's the bottom?
    - Your ability to hear low frequency sounds falls off at about 6dB/octave below somewhere around 700 Hz, give or take.
    - You can certainly sense 20Hz in the chest, skin, and abdomen
    - Maybe, at just the right levels, you even hear it from your hearing apparatus.
    - 40Hz is where there is sensation via the ear.
    - But, remember, sensation via body sensation is a thing!

So, let’s talk more about Acoustics.

What does the room (or lack thereof) do?

There's the direct path

- It's a coherent signal that comes from each source to the listener without any reflections or obstructions
  - A few speakers are designed to have no direct path, yes, I know.
  - The direct path can, in the absence of other direct paths, be localized.
  - I'll say rather a bit more about direct signals later

  - The auditory system is designed to detect direct signals and notice them.
    - Even when somewhat buried in the diffuse field.
  - The direct signal always, I repeat, always, gets there first in any normal acoustical setup.
  - Direct sound power follows the inverse square law.
  - The farther a direct signal travels, the less coherent it becomes.

Less coherent? How’s that?

- The propagation of sound in air can be less than uniform
  - Thermal currents or gradients can create diffusion.
  - HVAC, turbulence, etc., can create diffusion
  - Wind can literally blow the signal around. Ask any stadium sound builder

- As the signal becomes less coherent, it starts to sound more “distant”.
  - This comes partially from the way that the signal lines up across frequency after cochlear analysis.
  - Later, that comes later.

There are early reflections

- There are specular reflections.
  - That means that the wavefront remains mostly coherent
  - They can have various effects, from good (enhance directional sensation) to bad (create false directional sensations, obscure phantom images)
  - They can lead to dramatic frequency shaping.

- There are diffuse reflections
  - A diffuse reflection does not maintain the coherency of the waveform
  - They can help with phantom imaging, sometimes
  - They are often ignored by the auditory system deep in the brain (more on that later, too)

And, of course, late reflections

- Late reflections, if specular, are effectively heard as separate sources, especially if the delay is substantial (i.e. echoes)
  - They are most often quite undesirable
    - They can wreck articulation
    - They can create difficult feedback (monitor problems) for performers
    - Best to be avoided if possible.

- Late, diffuse reflections are the start of the diffuse field.
  - They will be reflected back and forth in the venue, incoherently
  - They can help with front/back disambiguation

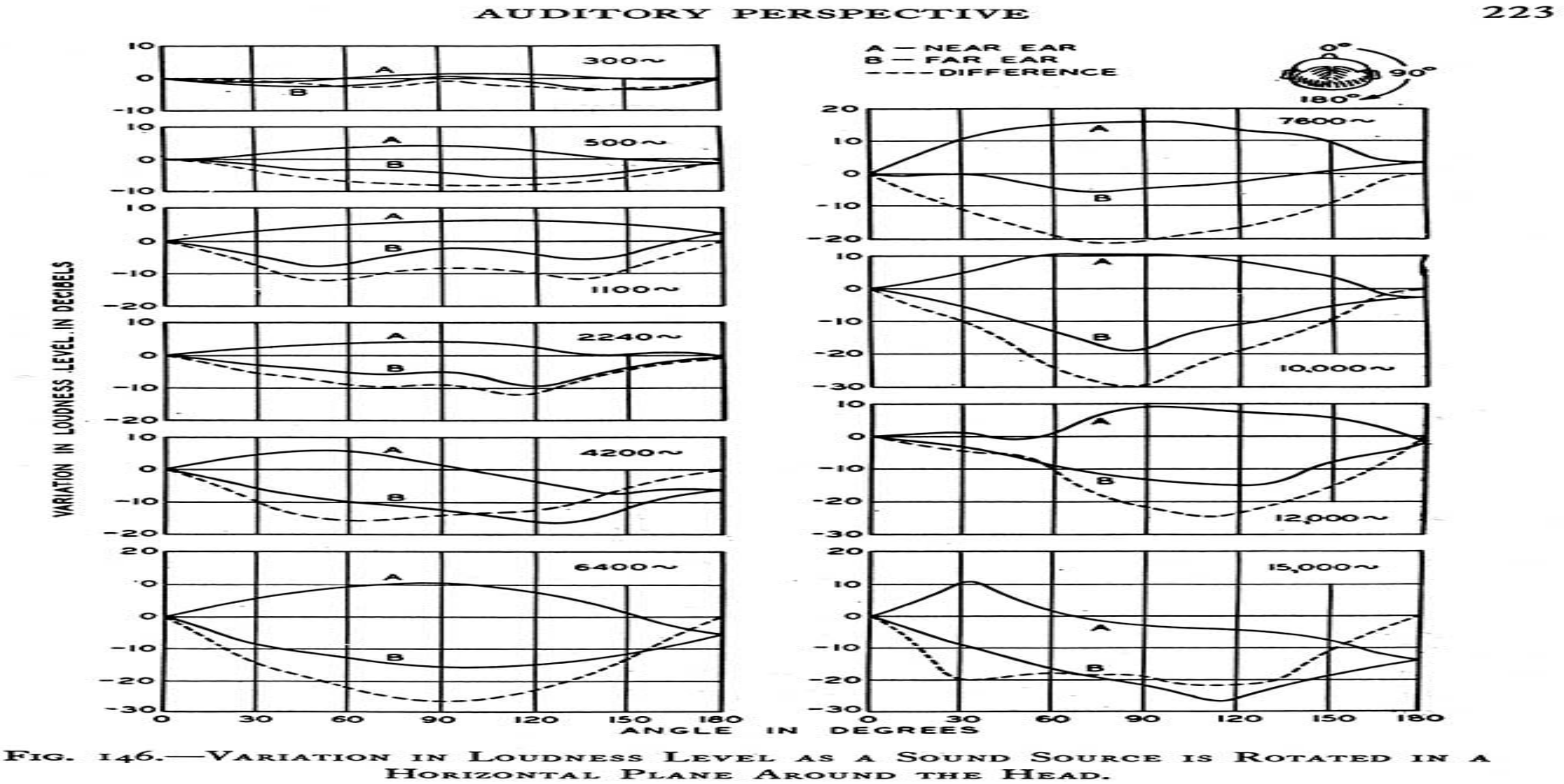
And, the diffuse field

- That’s the sound bouncing around over and over through the listening space.
  - It is often regarded as “more or less uniform”
    - But, it’s not really uniform in many venues
    - The onset of wavefronts is almost completely obscured, except at very low frequencies
    - It is immersive (it sounds “around” you)
    - The signal envelope lags the direct sound, sometimes by seconds.
    - It’s that “hall sound” that you either love or hate depending on the hall
  - Something to remember for later, our auditory system is rather effective at defending us from the diffuse field.

Sound vs. the Human Head

- Well, at this point everyone has heard of the HRTF or Head Related Transfer Function, alternatively HRIR (Head Related Impulse Response).
  - It is a function of distance, horizontal angle (theta) and vertical angle (phi)
  - Delay is INCLUDED to each ear for this discussion.
  - This provides the usual interaural delay we all expect (having learned our own HRTFs' by using our ears).
  - An HRTF does not only have different gain across frequency, it also has different time delay (to some extent) across frequency.
  - Dispersion, reverberation, and noise in a room can create ambiguity in the Interaural Time Difference at a given frequency.

Some very old examples



AUDITORY PERSPECTIVE 223

FIG. 146.—VARIATION IN LOUDNESS LEVEL AS A SOUND SOURCE IS ROTATED IN A HORIZONTAL PLANE AROUND THE HEAD.

Ear Canal Resonance

- Your ear canal response is in part a factor of externalization
  - It is personal to each individual
  - To some extent, you can learn small differences
  - Unlike HRTFs, which vary with
    - Hair
    - Collar
    - Head angle to shoulders
    - (and other stuff)

    It is quite constant.

  - You know and learn your personal ear canal resonances, which are not generally exactly the same in both ears.
  - Ear canal resonance, along with dispersion across frequency, is a primary cue of "externalization", which is to say the sensation that the sound is outside your head
    - This may in fact fail in a very quiet or non-reverberant room

Eardrum and transmission path to the cochlea

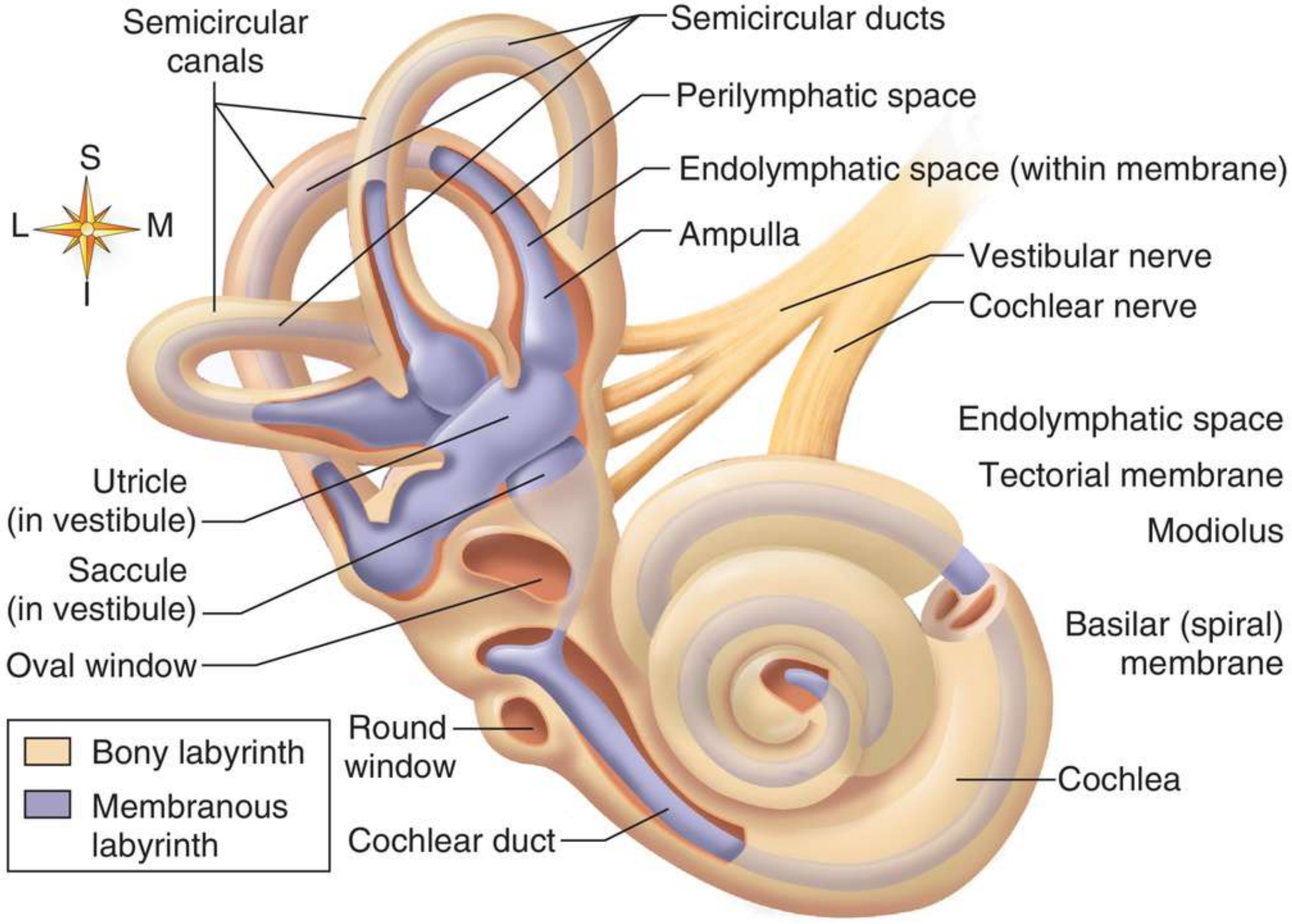
- There are basically two effects from the middle ear system.
  - A highpass filter, approximately 6dB/octave with the 3dB point at about 700 Hz.
    - This is part of the low-frequency sensation limit of the ear
    - It also protects from changes in barometric pressure at single digit Hz and below
  - A way to limit the level into the cochlea (the stapes reflex) that is a purely protective reaction.
    - It also helps to protect your ears from your own voice.

Then, the stapes (stirrup) connects to the oval window of the cochlea, and it gets complicated.

Cochlear Analysis

- NOTE! I am going to describe a model that seems to accurately describe the results of the cochlear processing. This is not, and can not be, complete or precise, but it does describe the measured performance of the ear after it's also processed through the brain. So from here on out, I have confidence in the phenomena, but not the entire mechanism.

- In short, the cochlea is a time/frequency analyzer that turns the pressure applied to the cochlear input into a pulse-position-plus-rate signal that travels down the auditory nerve to the brain.

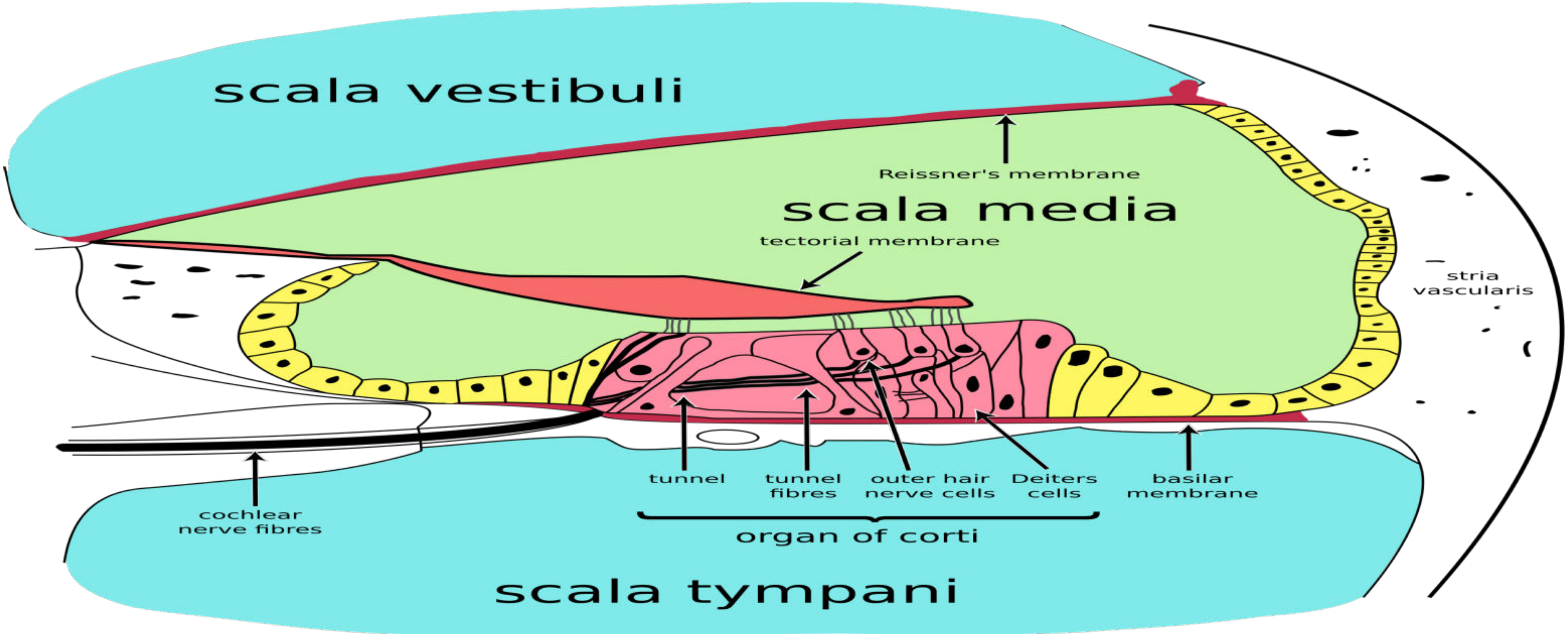

Semicircular canals
Semicircular ducts
Perilymphatic space
Endolymphatic space (within membrane)
Ampulla
S
L
M
I
Vestibular nerve
Cochlear nerve
Endolymphatic space
Tectorial membrane
Modiolus
Utricle (in vestibule)
Saccule (in vestibule)
Oval window
Basilar (spiral) membrane
Round window
Cochlea
Cochlear duct
Bony labyrinth
Membranous labyrinth

That’s kind of busy:

- So, I’ll go back to that, but focus on
  - The basilar membrane
  - The oval window connects to the stapes, that’s where “sound goes in” via vibrations. These vibrations are minute.
  - The round window lets the energy in the endolymph back out of the cochlea

- Something you can’t tell from that drawing:
  - The oval window and round window are on opposite sides of the organ of Corti
    - The tectoral and basilar membranes are part of the organ of Corti.
    - The basilar membrane and Reissner’s membrane separate the scala timpani (space connected to the ear drum) from the scala vestibuli (space connected to the air behind the ear drum).
    - The tectoral membrane extends from one side to the middle of the basilar membrane, inside the Scala Media (middle space)
- The three spaces are filled by fluid, each slightly different
  - The three spaces also have small electrical voltage differences.
  - That probably matters, but don’t ask me how, today.

The organ of Corti



Something to remember

- This is a tiny, tiny, thin, delicate organ

- It is now established that the inner hair cells do almost all of the detection
- The outer hair cells provide “compression”.

- Exactly how the compression works is the subject of endless debate at the present.

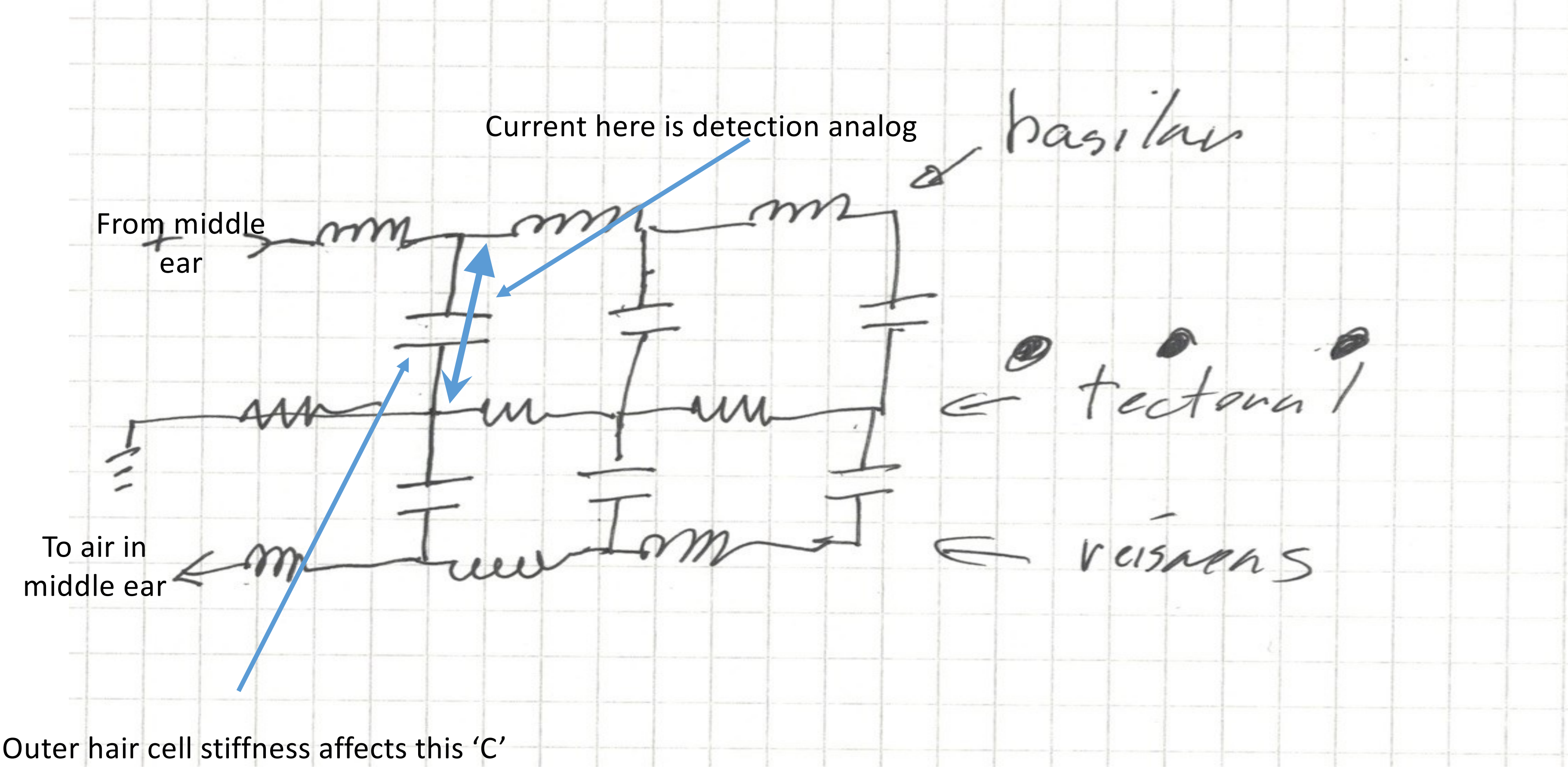
So how does this all make us hear anything?

- The setup creates a filterbank
  - High frequencies at the input end
  - Low frequencies at the far end

- The system is a kind of biologically created, nonlinear, transmission line filterbank.

- The outer hair cells change stiffness, and by doing so, changing membrane tunings.

For the electrical engineer among us:



Current here is detection analog

From middle ear

To air in middle ear

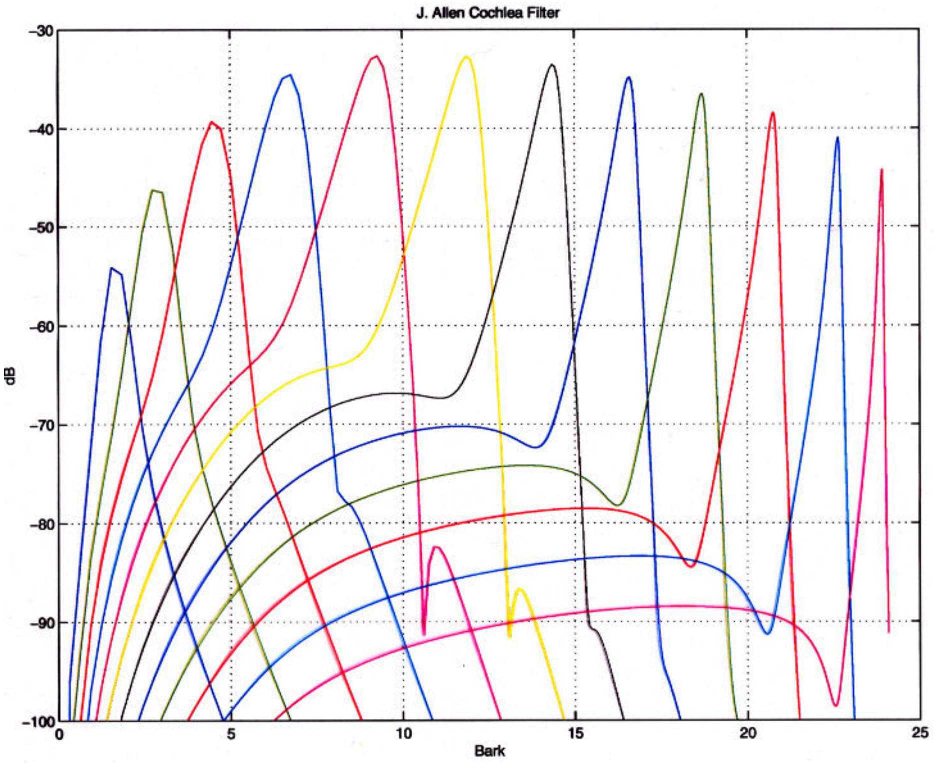
Outer hair cell stiffness affects this ‘C’

No, that’s not exact. It’s much more messy than all that.

- Everything is nonlinear
- The detection is mixed between “all or nothing” and “level”. In addition, leading edges are favored.
- Its delicate, and easy to overload.

- None the less, it creates a filter bank.

An example of some filters (no, not human)



How many filters are there?

- There are about 2500 inner hair cell groups.
- Each group, effectively, has a filter characteristic
- Filters overlap, and overlap a lot

- The “ERBs” we use are an estimate of the filter bandwidths
- The fact we conventionally use about 90 or so 1/3 ERB filters, each with a width of an ERB, is a computational issue.
- Yes, there really are 2500 bands, overlapping enormously.

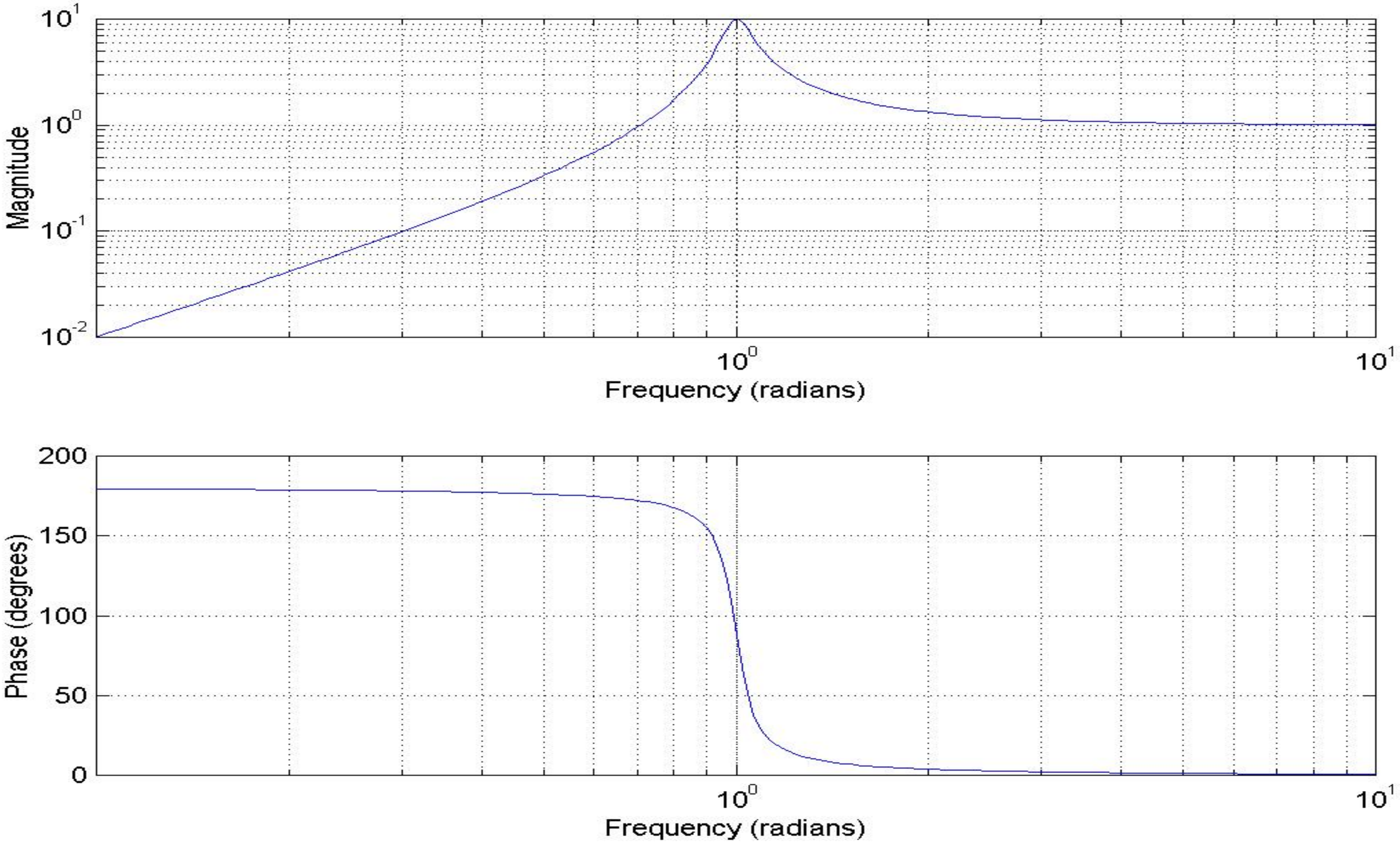
So, to take home:

- The ear is a frequency analyzer.
- The filters mean that there is actual mechanical separation of frequencies before the nerves are caused to fire.
- The shape of these filters is reasonably well determined.
  - The modern term for one filter bandwidth is “Equivalent Rectangular Bandwidth”, or “ERB”. You will hear a lot about ERB this and that later.
- So, the first two rules of audio processing are:
  - Don’t add new frequencies.
  - Especially, don’t add new signal at far-removed frequencies from the original.

How does that work?

- Well, both the basilar and tectoral membranes can be modelled as a peaky highpass filter.
  - The exact response varies with level
  - Nobody can quite assert exactly what that shape is, either.
  - Furthermore, it has to be more complicated than that, but this model suffices.

- The detectors (inner hair cells) fire on the difference between the two membranes.
  - This leads to small changes in the membrane tunings making large changes in sensitivity.
  - One example is shown in the next page. This is “for demonstration only”, the exact details are not established.


10^1
10^0
10^-1
10^-2
Magnitude
10^0
10^1
Frequency (radians)
200
150
100
50
0
Phase (degrees)
10^0
10^1
Frequency (radians)

So? What happens, now, when we link two such resonances?

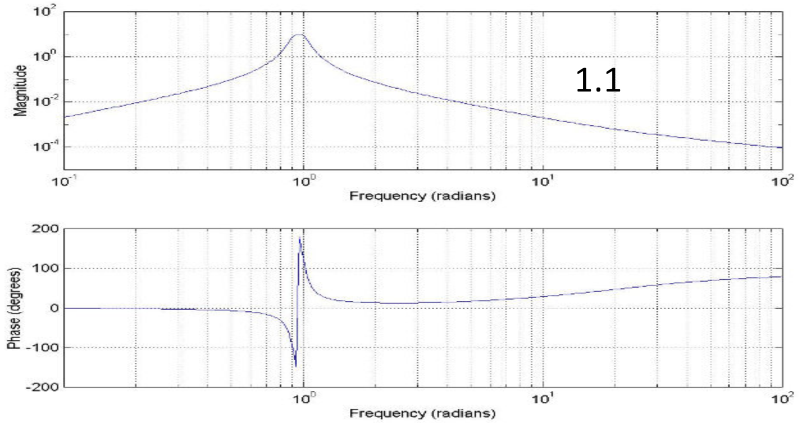
- As you may (or may not) recall, linking two resonances splits the resonance into two modes, separated by a small amount of frequency.

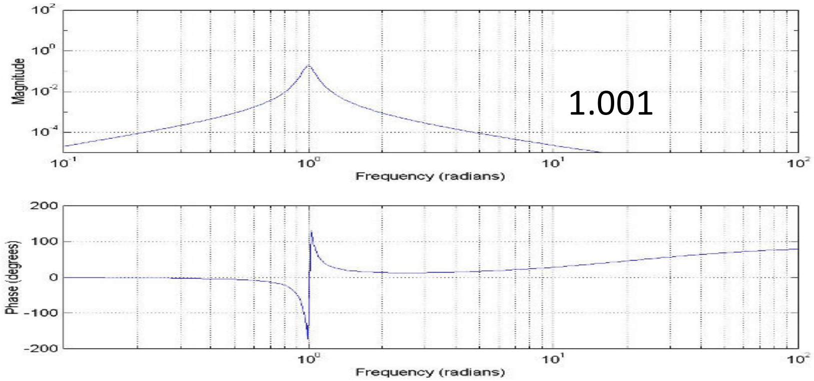
- Increasing the coupling makes the modes farther apart.

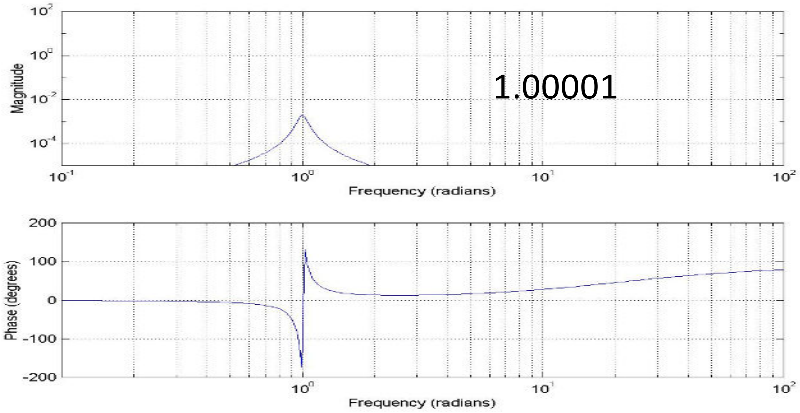
- Depolarized outer hair cells are much less stiff, i.e. provide much less coupling.

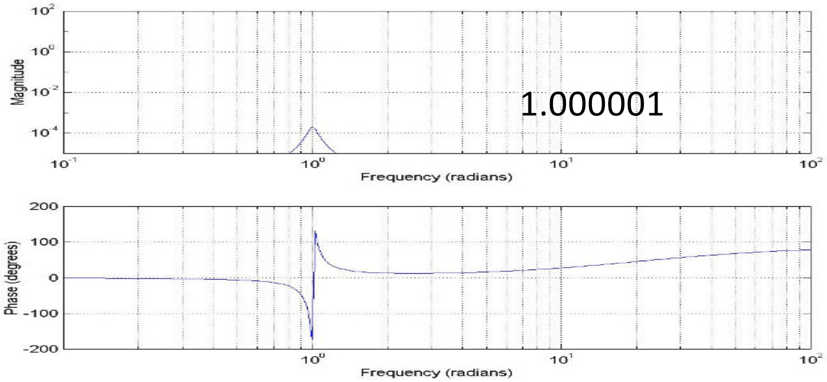
- So, something of this sort goes on. Exactly what is undoubtedly much more complex.

So, let’s look at the difference as the resonances split apart now. (numbers are relative offset)









So, notice that a small variation in tuning has several effects:

- The gain of the system as far as detection can change by many dB
- The phase of the signal about the center frequency changes extremely rapidly

- Both of these effects appear to be important on down the stream inside the brain. More on that later.

Loudness, to the hearing researcher

- Loudness refers to the PERCEIVED sensation level
- It varies with signal level
- It varies with signal duration
- It varies with signal envelope
- It varies with signal bandwidth

- As you widen a signal's bandwidth, with the same energy, the signal sounds LOUDER as long as the signal remains above absolute thresholds.

  Among other things that's the compression method used in **LOUD ADVERTISEMENTS.** Yes, it's annoying.

Two things to note about the compression effect

- First, it’s delayed by about 1 neuron firing, so compression starts about 1 millisecond after the signal onset.
  - This emphasizes leading edges of signals
  - It changes the tuning of the system quickly over time
  - This comes into play enormously in imaging, cocktail effect, and a lot of other spatial sensations, as well as for things like pre-echo.

- Second, it maps the 30dB (give or take) SNR of the inner hair cell to the 90+dB input to the cochlea, as a kind of AGC (a very fortunate one, too, especially in the modern world.

One final, and very important thing:

- As you may recall from my talk on ‘what is bandwidth’, the narrower the bandwidth, the longer the impulse response, integration time, detector window, what-have-you is.
- Conversely, the wider the bandwidth, the shorter the detection time, etc., are.

- This means that at any time, the low frequency parts of the hearing apparatus are listening to a 40 to 50 times longer time window than the high frequency parts. The actual partial loudnesses (loudness vs. frequency) respond to very different time scales.
- This is very important, moving forward.

Time for the break. During the break, I believe we’ll do the door prize drawing, and somewhere adjacent will be some refreshments.

Your speaker will retire for a few minutes to stock up on coffee, as well.