DSP
Online
Conference

Refined Audiometrics Laboratory

www.dsponlineconference.com

Bending Sound to Suit the Musician

Musically Accurate Hearing Restoration

David McClain

AGENDA

1. Timbre, Barkhausen, and Critical Bands
2. Sensioneural Hearing Loss
3. Sones and How Things Ought to Sound
4. The Fundamental Equation for Hearing Corrections
5. How to Measure Bark Channel Power
6. The Crescendo DSP Algorithm

1 Timbre, Barkhausen, and Critical Bands

- Why musical instruments sound distinctive,
- Dr. Bekesey and his Position Theory of Pitch Perception,
- Frequency vs zBark,
- Critical Bands and Upward Loudness Masking

Musical Timbre

The Sustained Sonic Character of Musical Instruments is their Timbre

- Two items distinguish different sounds - Attack, and Timbre
    - A woman singing and a flute can sound remarkably similar, if you skip the attack
- Timbre is the relationship between the harmonics and its fundamental tone
    - Harmonics are developed by the excitation
    - The spectral envelope over the harmonics produce formants
- Resonator shape and size dictate the envelope over the harmonics
- We can turn an oboe into a trumpet by shaping harmonics with an equalizer

Dr. Bekesey, 1961 Nobel Prize in Physiology

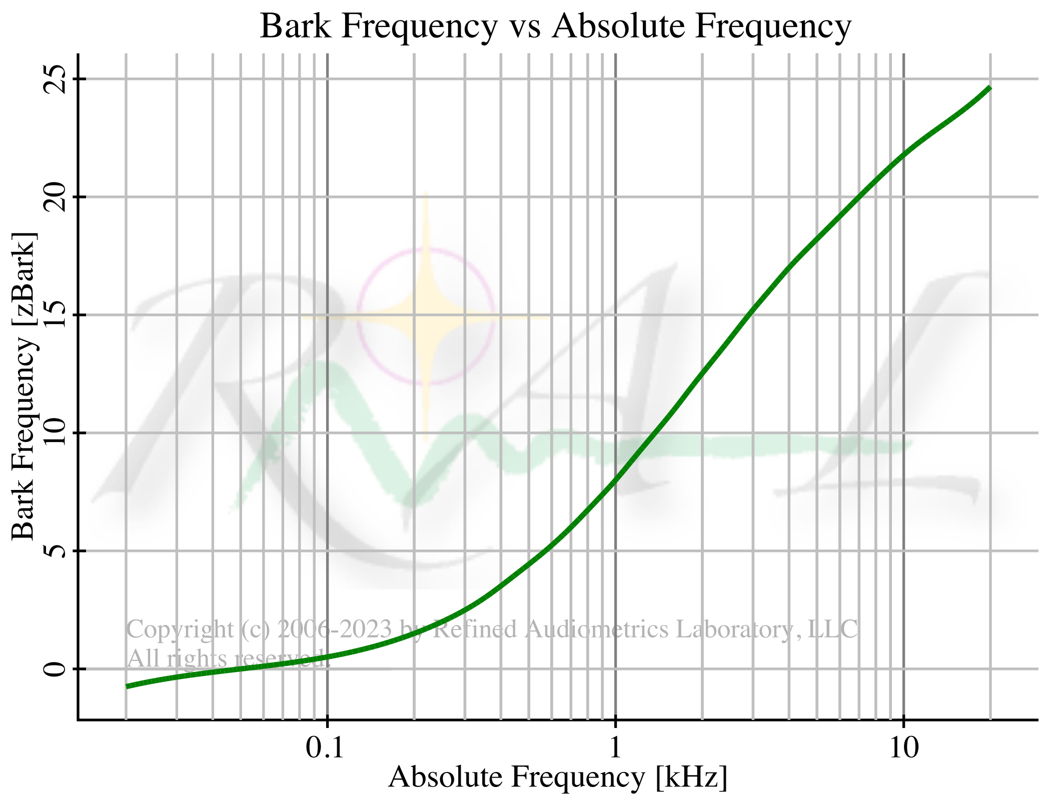
He Developed the “Position Theory for Pitch Perception”

- Tones of different pitches are sensed in different regions along the basilar membrane
- Bass sensed at the apex (far end) of the cochlea
- Treble sensed nearest the oval window where sound enters the cochlea

Barkhausen

Human Hearing Responds Nonlinearly with Frequency

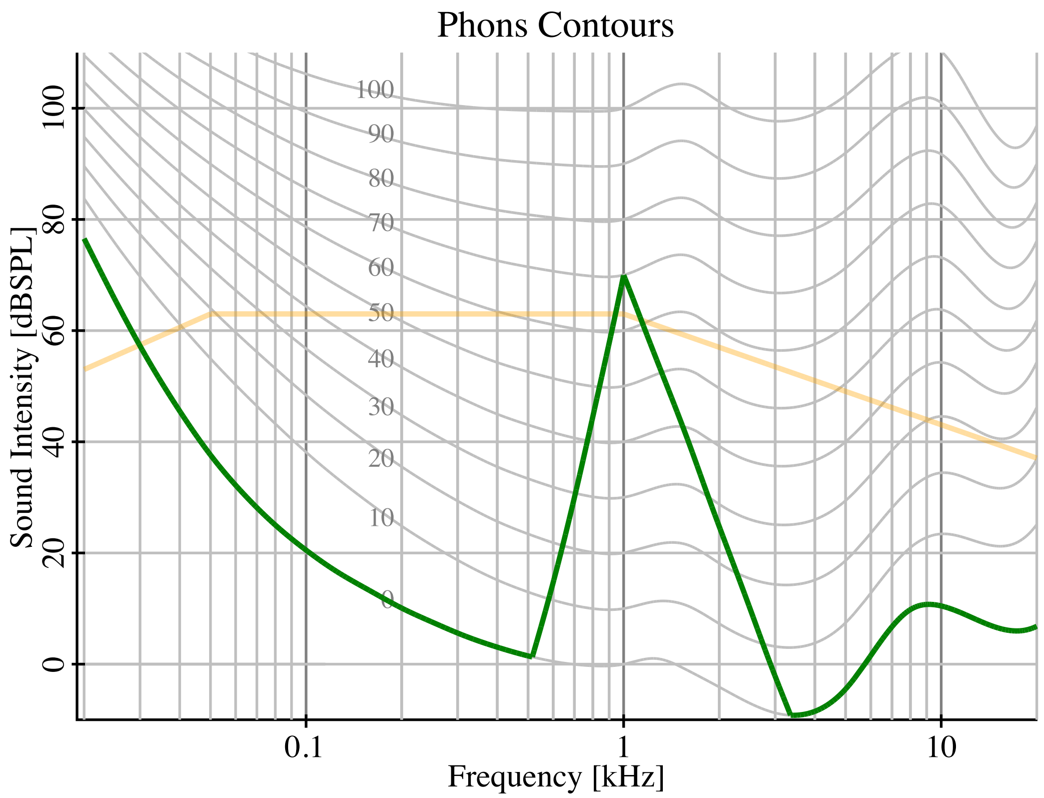
- Unlike an FFT with fixed-width frequency bins
- Nearly logarithmic above 500 Hz
- Nearly constant Q bandwidths above 500 Hz
- About 100 Hz fixed width below 500 Hz
- Human hearing is more appropriately frequency mapped using zBark instead of Hz
- 25 Bark bands across the audible range
- Bark frequency corresponds with linear distance along the basilar membrane, measured from the apex



Critical Bands

A Strong Tone Develops Self-Organized Bands Around the Tone

- A critical bandwidth is a frequency zone around a loud tone, where lesser tones are masked
- The threshold of hearing is raised nearby
- A critical bandwidth is approximately 1 zBark wide
- Bands below are masked ≈ -20 dB/zBark
- Bands above are masked ≈ -10 dB/zBark
- Upward Loudness Masking
- The masking profile is the shape of an auditory filter in reverse
- While our pitch perception is quite acute, our loudness perception with frequency is comparatively dull



2

Sensioneural Hearing Loss

- Audiograms, ATH, and IsoPhons (ISO-226)
- The Ski-Jump Multi-band Compressor
- DIY Solutions? Equalizers? Compressors?

Sensioneural Hearing Loss - I

Most typically exhibits a loss of high frequency hearing

- As frequencies increase the hearing grows more impaired
- Can be measured as an elevation in your threshold of hearing at each of several test frequencies
- Commonly referred to as Recruitment Hearing
- Related conditions:
    - HyperRecruitment or HyperAcusis
        - Sounds grow louder too quickly, once above your elevated threshold
        - Often hinted at by a diminished maximum comfortable level
    - Decruitment - sounds never become as loud as they ought to
    - You may have all 4 conditions in separate frequency zones
        - Normal hearing in bass - those frequencies are the most protected by physiology
        - Recruitment hearing due to all the various causes
        - HyperRecruitment nearby to profound narrow frequency notches in your hearing
        - Decruitment hearing in the highest frequencies

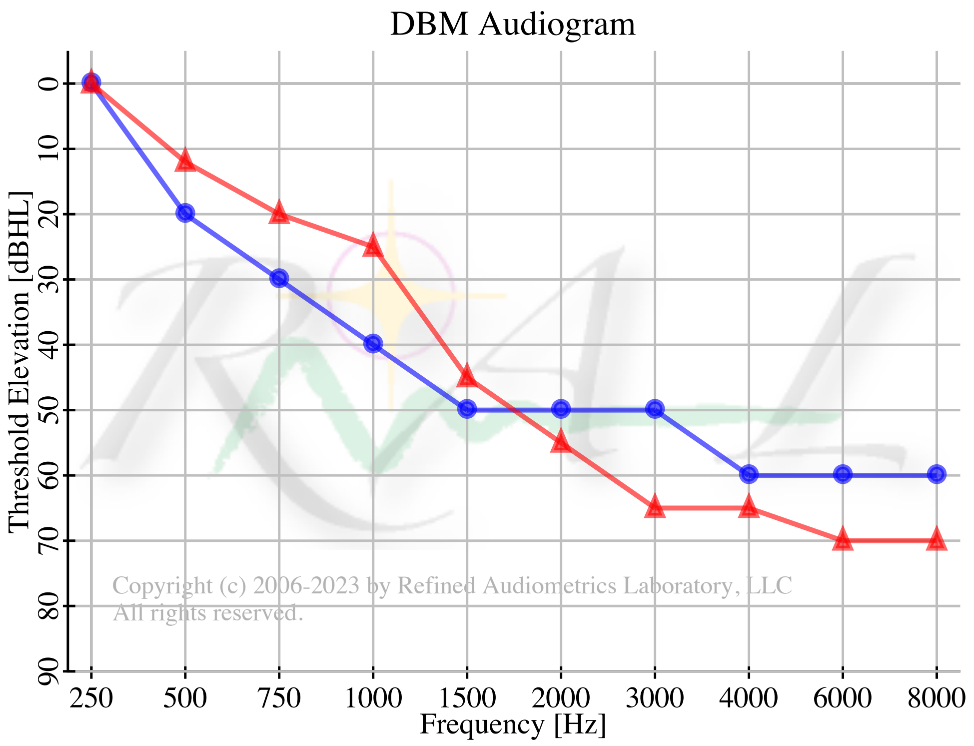
Sensioneural Hearing Loss - II

The most common form of hearing impairment

Arises from:

- Aging (presbycusis) - happens to all of us to varying degrees
- Noise exposure
  - overloud performance venues
  - playing your music too loudly through headphones
  - riding too loud motorcycles
  - industrial noise from day jobs
  - participation in war zones
- Ototoxic medications
  - Aspirin
  - Chemotherapy
- Genetic predisposition

An Audiogram Showing Impaired Hearing - I



Audiograms plot the elevation of your threshold of hearing at each of half-octave testing frequencies.

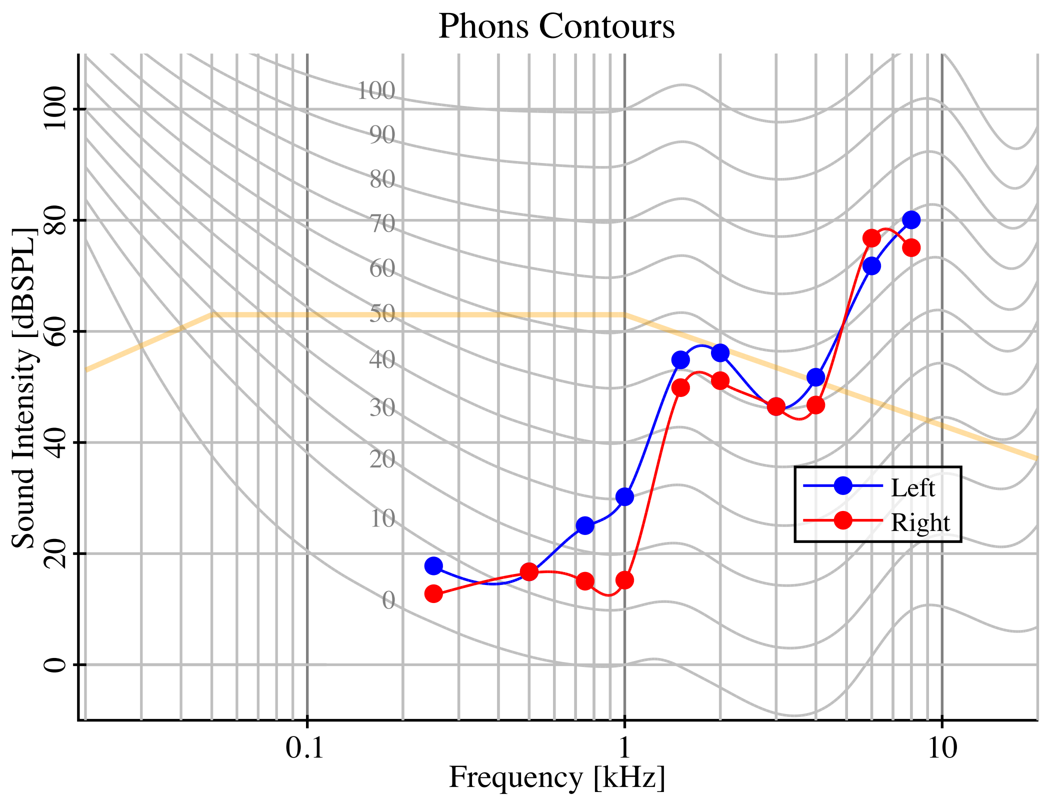
- Plotted upside down, by convention. Worse hearing is lower down. Perfect hearing along the top.
- Normal hearing is anything less than 20 dBHL threshold elevation.

Left channel in Blue, Right channel in Red.

Audiometry claims accuracies of ± 5 dBHL, and probes in increments of 5 dB increases of presentation level.

A zero dBHL threshold is the same as the Absolute Threshold of Hearing (ATH).

An Audiogram Showing Impaired Hearing - II



The same audiogram, but now presented in normal scientific terms:

- Log frequency horizontal axis, and
- dBSPL vertical axis.

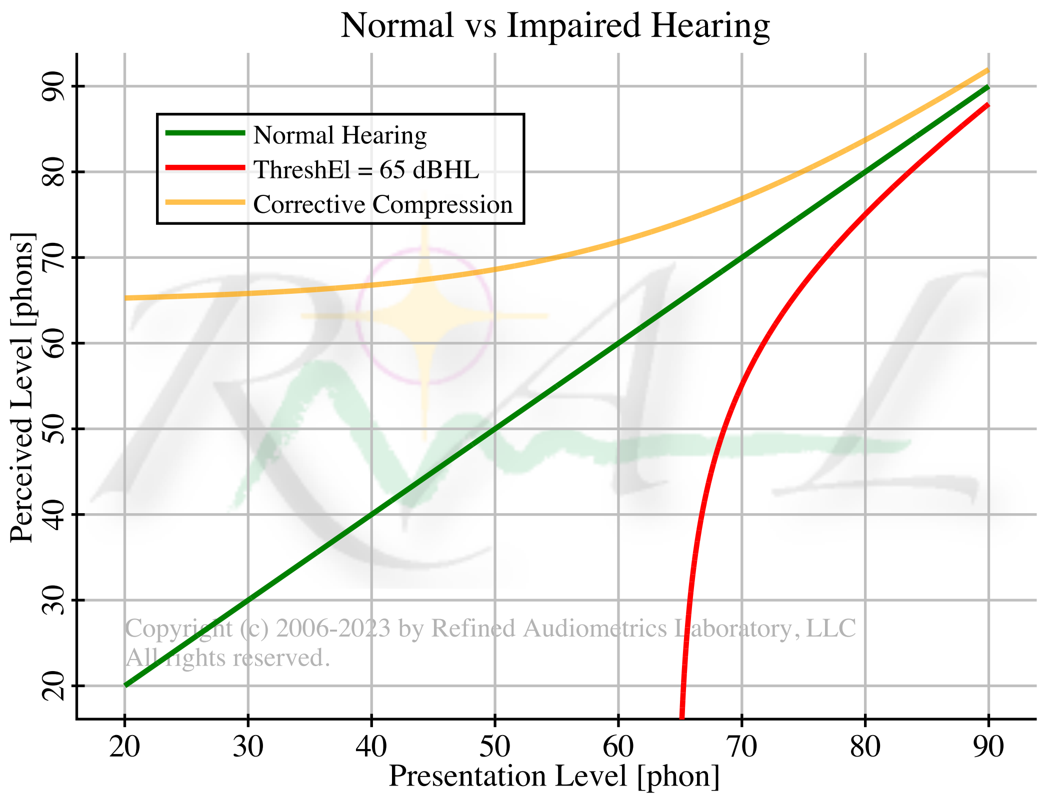
Here you see successive contours representing levels of constant loudness - a perceptual measure.

They vary with frequency, as does the ATH.

The levels represent isophons, labeled in phons, which agree with dBSPL only at 1 kHz, in both zero point and scaling.

dBSPL is what our meters read, what our microphones respond to, and what our loudspeakers produce.

Recruitment Hearing



A Perceptual Diagram Showing Recruitment Hearing
- and How to Correct It

- Horizontal axis is sound presentation level in phon

- Vertical axis is sound perception in equivalent phon

- Normal hearing correctly perceives the level presented

- Bright red curve = gain expansion. This is recruitment hearing.

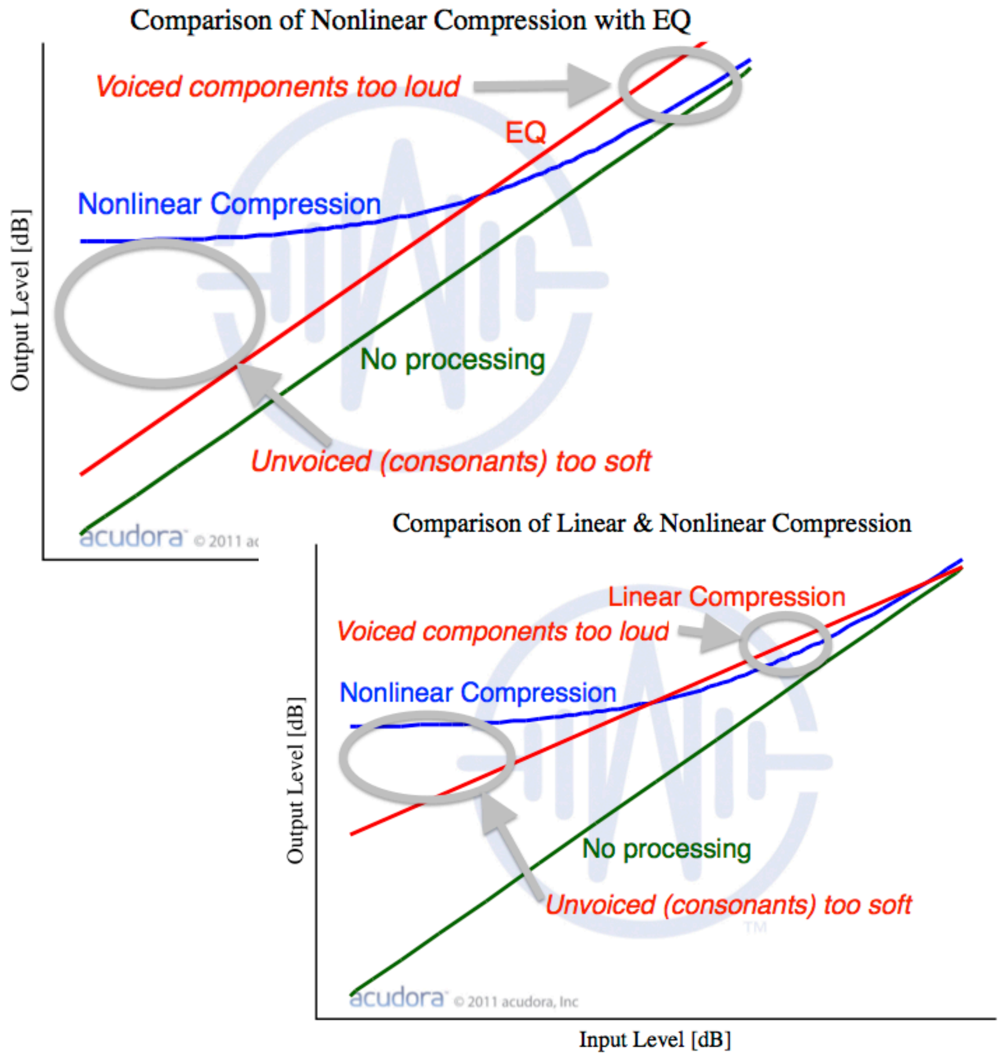
- Elevated threshold is 65 phon at this frequency

- A rapid rise in loudness once above threshold levels

- Behaves like normal hearing at loud enough levels.

- Complementary nonlinear compression looks like a ski-jump. This would rectify the recruitment curve to produce normal hearing at all levels.

DIY Solutions?



Simple EQ and Linear Compression Are Not Enough

- Simple Equalization can only be correct, for recruitment hearing, at one loudness level. Above that it becomes too loud, and below it grows rapidly too soft.

- Linear Compression is better, but can only be correct at two loudness levels. Inside those levels it is too much, and outside it is too little. There is no single compression ratio slope that corrects every loudness level.

3 Sones and How Things Ought to Sound

- A Target for our Hearing Corrections
- What are Sones, really?
- Cube Root Compression in Daily Experience
- Equal Loudness Contours

Early Hints from the Lab

You can measure human hearing in the same way that you measure radio receivers - 3rd Order IMD Intercept

We have long known that hearing is nonlinear…

Therefore:

- It develops IMD products at loud enough sound levels
- The 3rd Order IMD Intercept moves with increasing loudness (!!)
- Testing via Julius Goldstein’s probe tones
- Yields the exponent of nonlinearity, dominated by 3rd order IMD
- Hence the EarSpring Equation

Simplest Possible Model for the Whole of Human Hearing

EarSpring:

$$\left[\frac{d^2}{dt^2} + \beta\frac{d}{dt} + k\left(1 + \gamma\left\langle y^2\right\rangle\right)\right] y(t) = F(t)$$

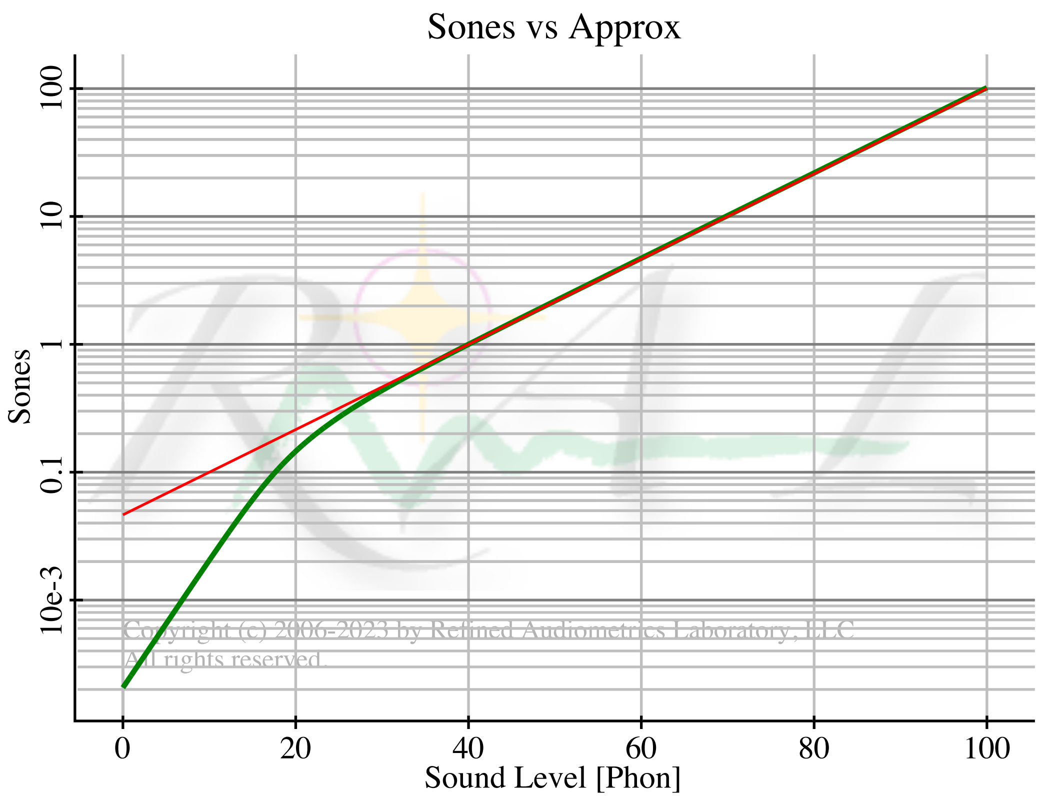
- A damped harmonic oscillator whose stiffness grows with increasing average power of vibration

- Captures salient characteristics of human hearing:

    - Linear behavior near threshold levels (sound isolation booth levels)
    - Cube-root behavior above 40 phon (everyday loudness levels)
    - Pitch flattening at loud levels
    - Develops subharmonics and IMD products at loud levels
    - Expressly *not* merely a cochlear model
    - Encompasses the whole of human hearing:
      cochlea, afferent 8th nerve, brain, efferent 8th nerve - a complex feedback and control system

- Provides a target for how things ought to sound

Cube Root Compression in Daily Experience



The Solution to EarSpring

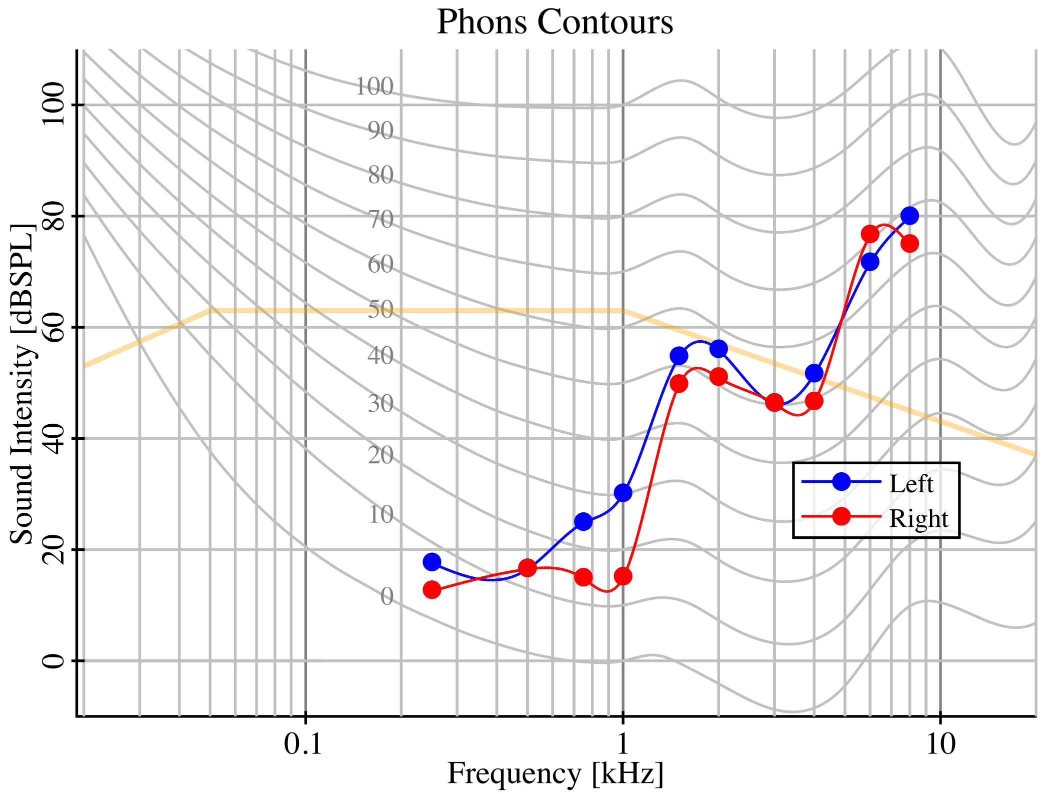
- Subjected to laboratory measured boundary conditions

- Cube-root compression above 40 phon

The red line is a convenient approximation for daily experience, but fails near threshold levels where hearing becomes linear.

$$S(P) \approx 10^{\frac{P-40}{30}}$$

Equal Loudness Contours (Phons)



ISO-226, formerly Fletcher-Munson Equal-Loudness Contours

The dip between 3-4 kHz in the ATH (bottom contour) is due largely to the quarter-wave resonance of the outer ear canal. So too for the dip around 12 kHz.

We may have evolved sensitivity there to prevent our being eaten alive by prowling saber toothed tigers, and to have heightened awareness of screams for help from infants and other people.

Notice how deafened we all are below 100 Hz. And yet, once audible, the bass only needs a slight increase to sound much louder to us.

What are Phons and Sones?

1 Sone = 40 Phon = 40 dBSPL = 40 dBHL at 1 kHz

At 1 kHz, Phon = dBHL = how many dBSPL above ATH (Absolute Threshold of Hearing)

At every other frequency they depart in scaling and zero point from SPL.

Sones and Phons are perceptual measures.
At every frequency, 0 phon = ATH = 0 dBHL, and 1 sone = 40 phon

You can't read them from any meters - you have to ask the audience for a concurrence of opinion. Phons are often used to describe single frequency carriers, while Sones are used for more complex sounds, i.e., music.

Microphones, loudspeakers, and voltage meters measure SPL and work on the dBSPL scale. So too, for all of our audio mixing and equalization. dBHL use the same scaling as dBSPL, but have a frequency dependent zero point = elevation in dB above ATH.

Using Phons, instead of dB from SPL-space, allows us to work independently of frequency. Phons intensity is that which reaches the cochlear sensors, after being filtered by the outer and middle ear, and after taking into account a frequency dependent coupling strength.

What Are Sones, Really?

Sones of Phons: $S(P) \approx 10^{\frac{P-40}{30}}$

Phons of Sones: $P(S) \approx 40 + 30\log_{10}(S)$

Good to < 1.7% above 36 phons (i.e., everyday loudness levels)

You may have heard that a 10 dB increase in sound power is equivalent to doubling its loudness. It's actually, 9.03 dB, but could you really tell the difference? (cube root compression above 40 phon) And just how, exactly, do you judge "twice as loud" at any particular frequency?

Considering EarSpring in a frequency normalized form, ***we can identify Sones as the power ratio***:

$$S_P = \frac{\langle y \rangle_P^2}{\langle y \rangle_{40}^2} = 10^{\frac{P-40}{10}} \cdot \frac{4\hat{\beta}^2 + \Gamma_{40}^2}{4\hat{\beta}^2 + \Gamma_{40}^2 S_P^2}$$

a cubic equation in S_P. Factor $\hat{\beta}$ is a frequency normalized damping factor, and Γ_{40} is a measure of the γ nonlinearity in the EarSpring equation, at the 40 phon level. $\hat{\beta}$ can be determined from the amount of pitch flattening discerned for a constant pitch tone between 40 phon and 90 phon presentation levels.

We can write the closed form solution for S_P, but for most cases, the approximation above is good enough.

4 The Fundamental Equation for Hearing Corrections

- A Surprising Relation!
- Perception vs Physics
- Dead Hair Cells?
- Effects of Errors in Audiometry

The Fundamental Equation of Hearing Corrections

$$\langle y^2 \rangle_{P+dP} - \langle y^2 \rangle_{P_{thr}} = \langle y^2 \rangle_P - \langle y^2 \rangle_0$$

In order to allow a person with impaired hearing, with threshold elevation P_{thr}, to have the same experience as a person with normal hearing, you need to present a sound level as far above that threshold, as happens for a person with normal hearing above their threshold, P_0.

This may seem surprising since so much of our engineering and physics depends on SNR (signal to noise ratios). But SNR belongs to physics. What, in our physiology, could possibly perform that kind of ratio? Instead, this is a perceptual equation, not physics. This is hour our mind works. Attempting to model corrections based on SNR leads to absurd conclusions.

The human mind perceives constancy with dulling attention, and responds only to changes above a threshold. Just pinch yourself and hold, and after a while you stop noticing it.

So this means:

$$S(P + dP) \approx S(P) + S(P_{thr})$$

since, $S_0 = 0.002$ sone, and normal sound levels are around 1 sone and higher (40 phon and above).

Solve for dP to find the needed corrective gain, in phon:

$$dP \approx 30 \log_{10} \left(1 + 10^{-\frac{P - P_{thr}}{30}} \right)$$

So if the threshold elevation is P_{thr}, then in order to perceive that level as itself, you need a gain of 9.03 phon. (Twice as loud!)

If your threshold elevation at some frequency is 60 dB, it ***DOESN'T*** mean that you need 60 dB of gain!! That's only for discerning threshold level sounds. Not for everyday sound levels. At the loudest levels, impaired hearing becomes nearly normal.

Dead Hair Cells?

People often talk about hearing damage as arising from dead hair cells in the cochlea.

So if you have profound hearing loss, $P_{thr} > 90$ dBHL, then you must have lost most of your hair cells, right?

Well, consider that some fraction of hair cells are “dead”, in the sense that they are screaming out to the brain with a signal level equivalent to the threshold of pain, 120 phon. But they are surrounded by some fraction, f_{live}, of live hair cells which are operating normally. Then to have an elevated threshold of P_{thr} we need that:

$$f_{live}\, S(P_{thr}) = (1 - f_{live})\, S(P_{dam})$$

Plug in the numbers and you find out that for 90 dBHL profound hearing loss, only around 9% of the hair cells must be damaged. People with profound hearing loss can still hear loud sounds. It’s just that in order to help them recover some hearing, it may take so much corrective gain that it endangers the remaining live hair cells. So profound hearing loss is often considered a lost cause.

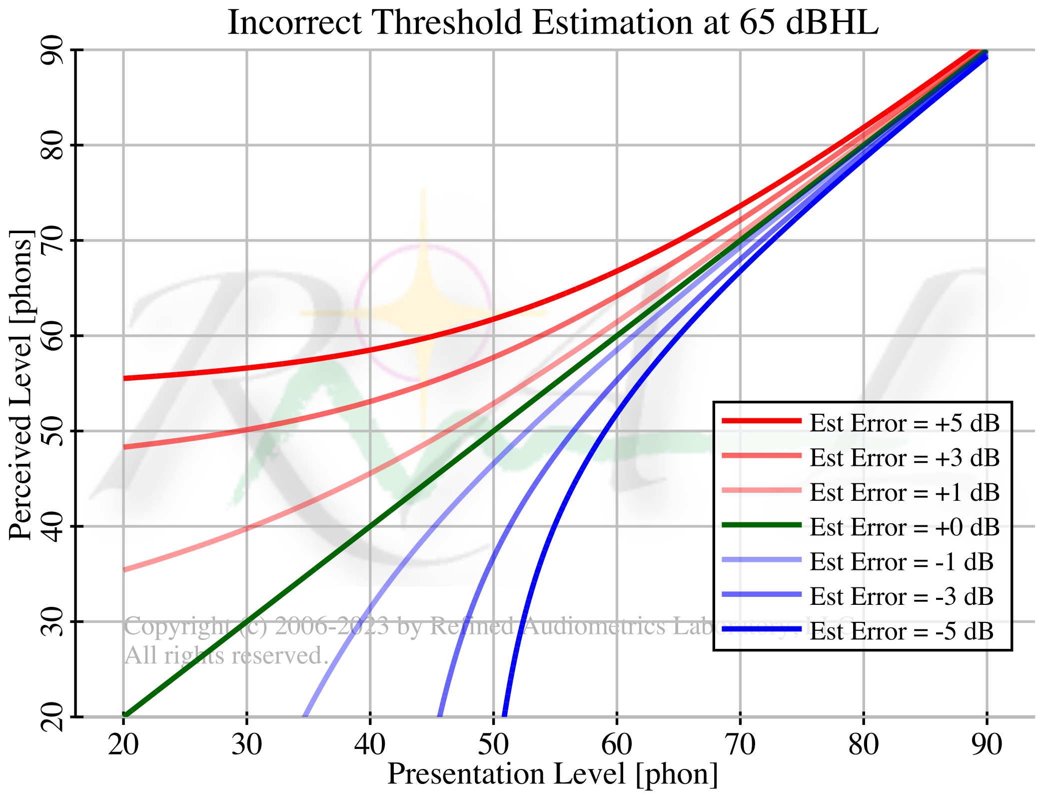
And we can augment our equations to accounting for remaining live cells, as: $f_{live}\left(S(P + dP) - S(P_{thr})\right) \approx S(P)$

$$f_{live} = \frac{1}{1 + S(P_{thr})/S(P_{dam})}$$

and

$$dP \approx 30 \log_{10}(1/f_{live} + 10^{-(P-P_{thr})/30})$$

Effects of Errors in Audiometry



Audiometry Claims ±5 dBHL Accuracy

We use elevated threshold measures to inform our hearing corrections. Any mis-measure leads, most likely, to obvious overcorrection in the most afflicted frequencies.

Perfect restoration is represented by the green diagonal line which shows that what you perceive is that which is presented.

You are less likely to have an under-reported threshold elevation, unless the audiometer is out of calibration.

Even as little as 1 dBHL error gives rise to substantial deviations from perfect restoration at low loudness levels.

5 How to Measure Bark Channel Power

- Power Delivered to a Critical Band
- Designing a Bark Channel Filter
- Filter Equivalent Rectangular Bandwidth (ERB)
- Relating Bark Channels to FFT Frequency Bins

Power Delivered to a Critical Band

The Equations, so far, Have Applied to Only One Frequency

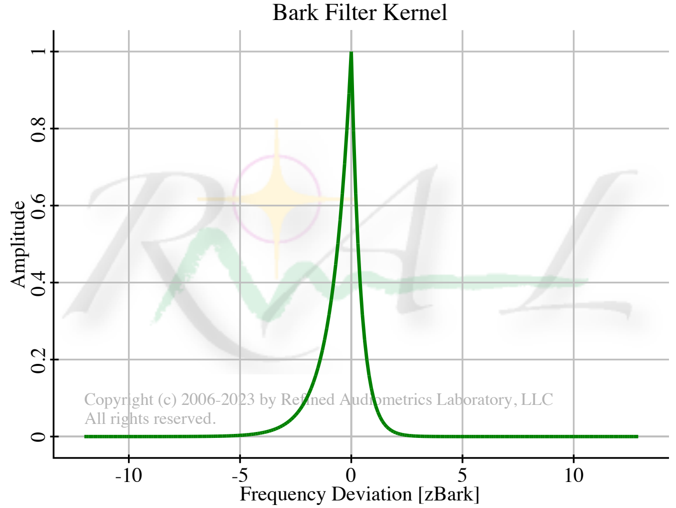
- We know we have different degrees of impairment at each different test frequency
- So we need to divvy up the incoming sound and measure the power level in each critical band ≈ each zBark band.
- But our test frequencies were spaced apart by ≈ 2 zBark
- You can use any number of channels to analyze the sound
    - I began with 100 quarter-zBark bands in stereo (a lot of CPU needed)
    - I once did a 5-band system and it worked well enough for Cell Phones
    - But we want a high-fidelity system for musical listening
        - How many bands do you really need?
        - Our hearing is a “tin-ear” when it comes to loudness discernment
        - Maybe 12 bands from 20 Hz to 20 kHz would work?
        - There seems to be a sweet spot around this many bands

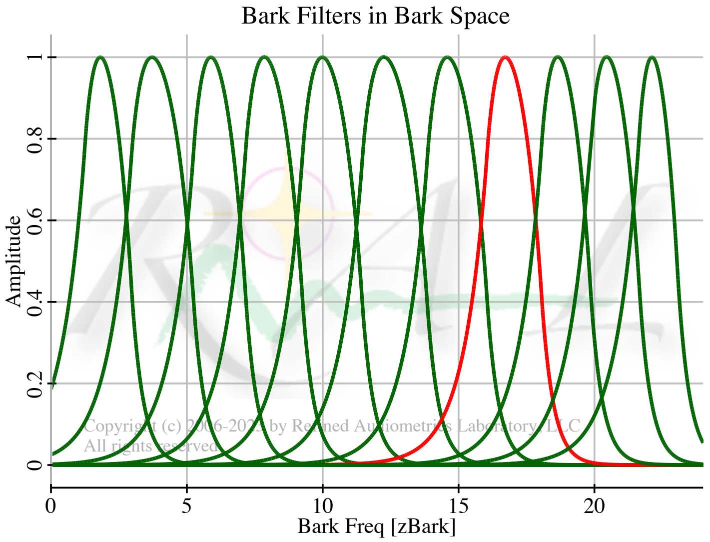
Developing a Bark Channel Filter

Use the Masking Profile to Design a Filter Kernel

- The Masking Profile is the Filter Kernel in Reverse
- Convolve the Filter Kernel across a Rectangular Bandwidth
- Do this for bands centered on the testing frequencies
- We are working in zBark, not Hz

(17 zBark = 4 kHz, in Red)





How to Convert to an FFT Filter?

Simple Mapping from zBark to Hz is Incorrect

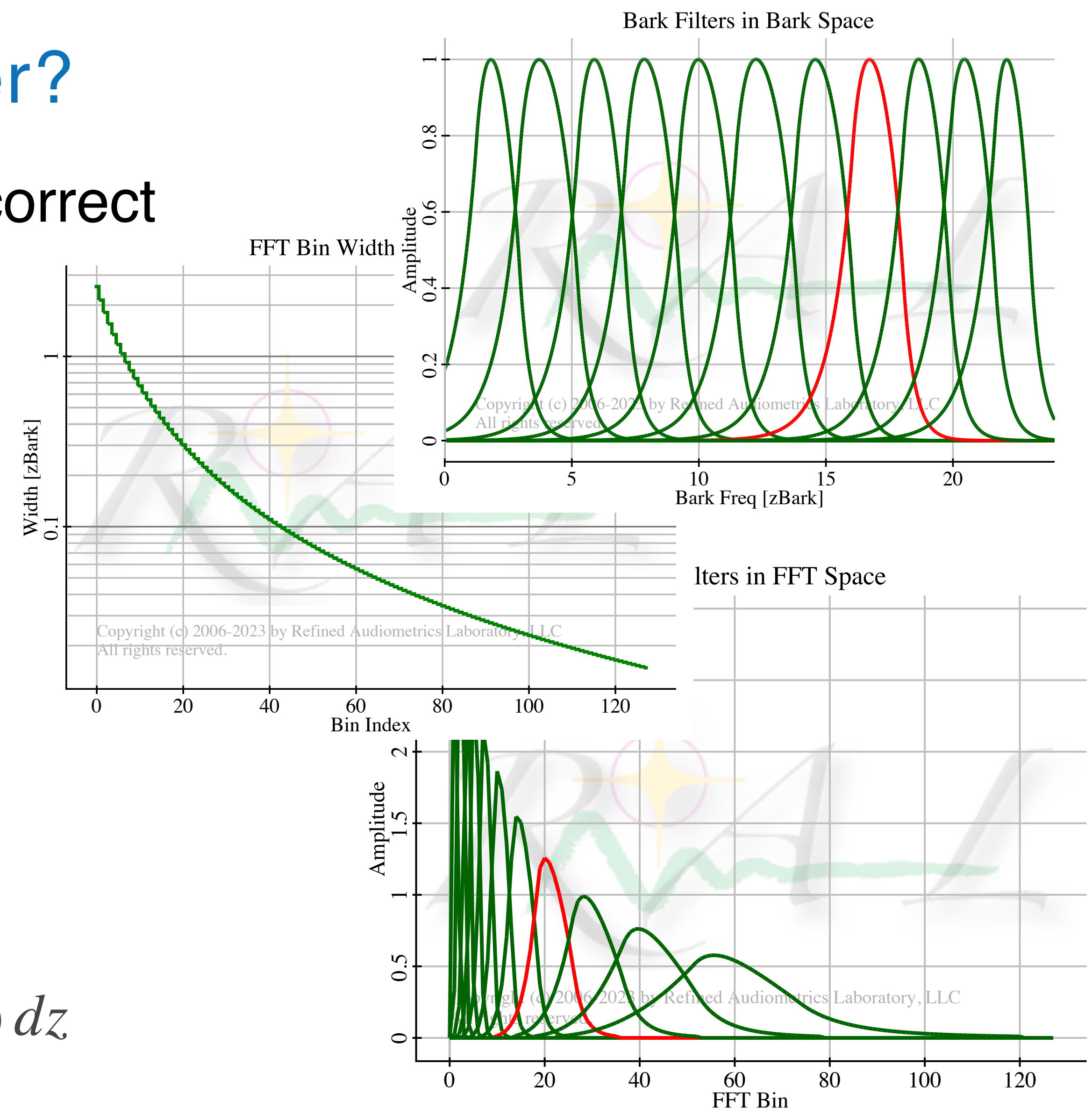
- What are we trying to measure?
- The Average Power in Each Channel:

$$< P >_c = \frac{\int F_c^2(z)\, P(z)\, dz}{\int F_c^2(z)\, dz} = \frac{\int \hat{F}_c^2(f)\, P(f)\, df}{\int \hat{F}_c^2(f)\, df}$$

- Conservation of Energy

$$\hat{F}_c^2(f)\, df = F_c^2(z)\, dz$$

- Filter Amplitudes are Irrelevant: $ERB = \int F_c^2(z)\, dz$



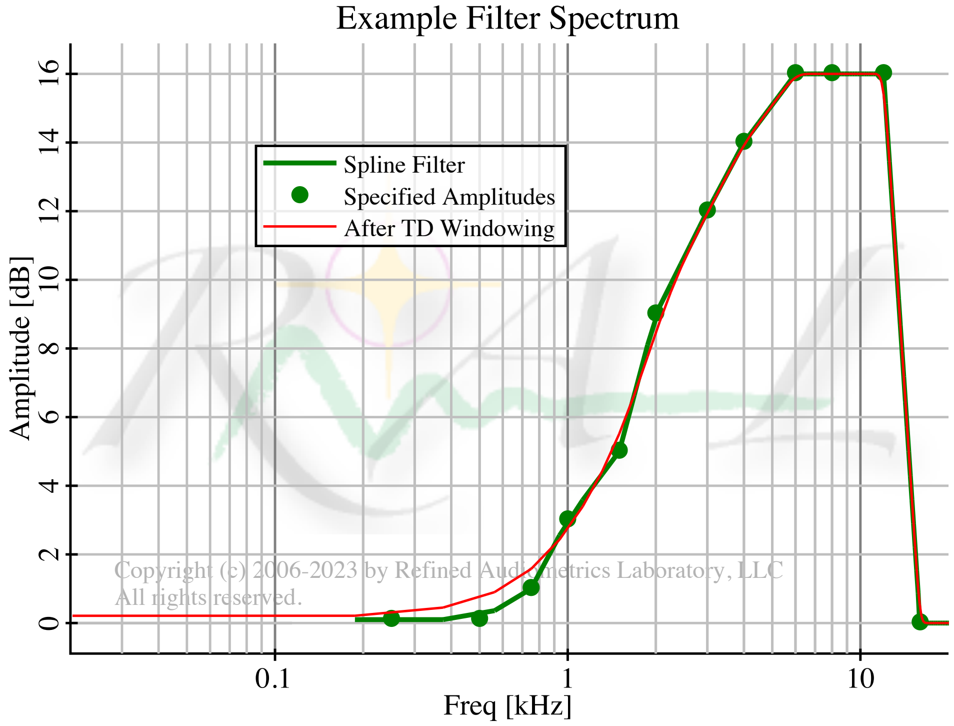
6

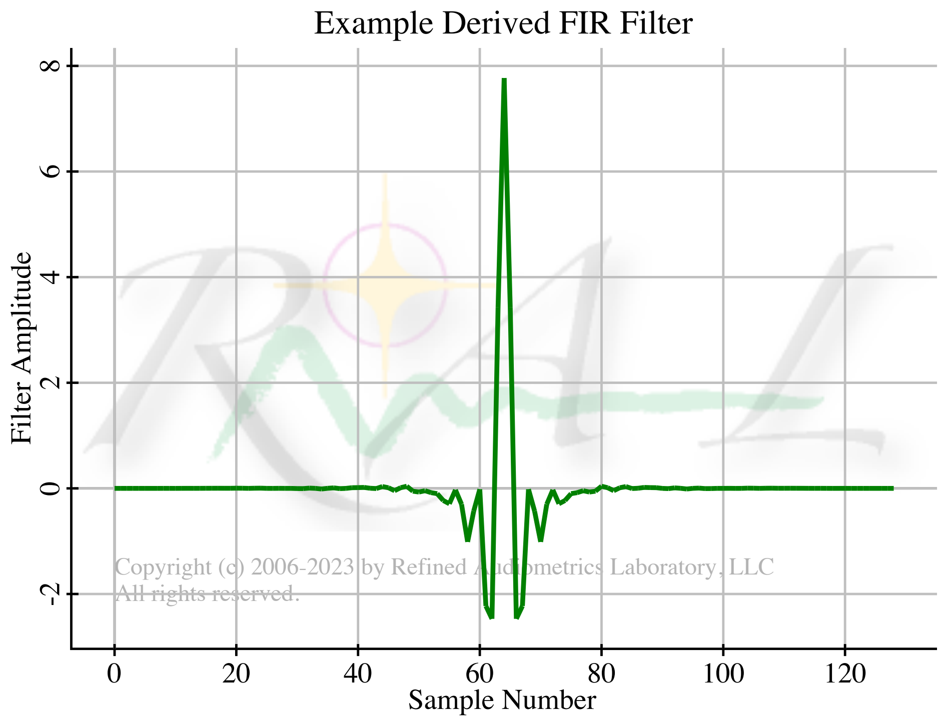
The Crescendo DSP Algorithm

- Dynamic Convolutional FIR Filters as Compressors
- Bucket Brigade Overlap-Append Processing
- Dynamics of Gain Management
- Changing Spaces (dB, Phons, Sones)
- Brief Demo

DSP? Anyone?

A MutiBand Nonlinear Compressor = A Dynamic Convolutional FIR Filter





Use Overlap Append FFT Processing
Block Size = 256 for Fsamp < 50 kHz, 512 if > 50 kHz
4 ms Throughput Latency = Good for Live Performance
375 Hz Update Rate at 48 kHz

Process 2 half-blocks at a time
Pick middle half-block as output
Stagger Half Blocks for Power Est vs Audio Proc
2.7 ms Lookahead

Dynamics of Gain Management

At Least 50% of Sound Quality Comes from Compressor Dynamics

- Compressor manages Gain Increases for Fading Sound Levels, not Gain Reduction
- Use Immediate Attack on Gain Reduction
    - Ear Safety
- Use 2-stages of Release + 20 ms Hold
    - On strong impulsive signals with rise > 6 dB go to Fast Release
    - Once signal returns to < 3 dB from mean level, go to Slow Release
    - Release = Corrective Gain Increases
    - Best Found: Fast = 34 ms, Slow = 155 ms, no audible pumping
- Limit Max Gain for ear safety

Changing Spaces

All Compressor Gains are Computed in Phon Space
and Applied in SPL Space

ATH, SPL-to-Phon, and Phon-to-SPL use (2,1) Rational MiniMax Approximations developed for each channel from cubic-spline interpolation of ISO-226

Errors within 0.1 dB from 0 to 100 dBSPL
Errors are dominated by experimental measurement errors, aka Audiology

Left/Right Stereo Channels Processed in Parallel
Each of the 24 Bark Channels are Processed in Parallel
Covers DC to 16 kHz, with 12 kHz band using the 8 kHz audiology (??)
16 kHz and up are ignored

THE SPEAKER

David McClain

Musician, Astrophysicist, Hearing Researcher

Focus: Musically Accurate Hearing Corrections, Cryptographic Protocols, and Astronomical Signal and Image Processing - DSP Centric

Owner of Refined Audiometrics Laboratory, in Tucson, AZ, USA, which does all kinds of research related things. David has had a long and varied career over the past 50+ years, with more than 40 years of DSP development for signal and image processing. Formerly an Astronomer, a Spook, a Rocket Scientist, and a Cryptographer. Now he just wants to pass along the most significant findings from his lifetime here among people on Earth, travel the World, and enjoy Lisp development and “Muscle Cars”.