Abstract:
In this lecture, the basic concepts of sound and acoustics are introduced as they relate to human hearing. Basic perceptual hearing effects such as loudness, masking, and binaural localization are described in simple terms. The anatomy of the ear and the hearing function are explained. The causes and consequences of hearing loss are shown along with strategies for compensation and suggestions for prevention. Examples are given throughout illustrating the relation between the signal processing performed by the ear and relevant consumer audio technology, as well as its perceptual consequences.

Christopher J. Struck
CJS Labs
San Francisco, CA
19 November 2014
Frequency is perceived as pitch or tone. An **octave** is a doubling of frequency. A **decade** is a ten times increase in frequency. Like amplitude, frequency perception is also relative. Therefore, a **logarithmic** frequency scale is commonly used in acoustics. The normal range for healthy young humans is 20 – 20 kHz. The frequency range of fundamental tones on the piano is 27 to 4186 Hz, as shown. The ranges of other instruments and voices are shown.
The **Threshold of Hearing** at the bottom of the graph is the lowest level pure tone that can be heard at any given frequency. The **Threshold of Discomfort**, at the top of the graph represents the loudest tolerable level at any frequency. The shaded centre area is the range of level and frequency for speech. The larger central area is the range for music. Level Range is also commonly referred to as **Dynamic Range**.
The subjective or perceived loudness of sound is determined by several complex factors, primarily Masking and Hearing Acuity. The human ear is not equally sensitive at all frequencies. It is most sensitive to sounds between 2 kHz and 5 kHz and less sensitive at higher and lower frequencies. This difference is more pronounced at low levels than at higher levels. Due to this effect, most Hi-Fi systems incorporate a Loudness switch to boost low-frequency response at low levels. Equal Loudness Contours, or Phon Curves, indicate the sound pressure level required at any frequency in order to give the same apparent loudness as a 1 kHz tone for continuous sounds. A-Weighting is based upon the inverse of the 40 Phon curve. This data was originally measured by Fletcher & Munson in 1933. Alternative data were developed by Churcher & King (1937), Pollack (1952), and Robinson & Dadson (1956), which is shown here. This data also appears in the ISO 226 standard. Note that 40 Phons = 1 Sone, 50 Phons = 2 Sones, 60 Phons = 4 Sones, etc. The Sone scale (not shown) measures perceived loudness, with loudness doubling every 10 dB.
The presence of a signal raises the threshold of hearing at nearby frequencies. **Masking** occurs when the threshold of audibility of one sound event (the target) is raised by the presence of another sound event (the masker). The dotted line represents the hearing threshold in quiet (i.e., without the presence of a masker). The altered threshold or **masking pattern** for a continuous pure tone at 1 kHz is shown. Simultaneous sounds beneath the large dashed line cannot be heard when this tone is present. For example, lossy audio coders (e.g., MP3, AC-3, etc.) make use of masking to perform data reduction.
Masking patterns for a 160 Hz narrow-band noise centered at 1 kHz (after Zwicker and Feldkeller). The audibility of a test tone is plotted in the presence of various masker levels. At 1 kHz, the masked thresholds are only 4 dB below the level of the noise masker. The secondary peak above 1 kHz at 80 and 100 dB is likely due to a non-linear intermodulation product, as the very narrow bandwidth of masking noise is not perceived as a steady sound level. The masked thresholds without this effect are indicated by the dashed lines. As the masker level increases, the upper skirt of the threshold broadens toward high frequencies. This effect is commonly referred to as the “Upward Spread of Masking.”
Thresholds of tones masked by white noise (after Zwicker and Feldkeller). The lowest curve is the threshold in quiet (no masker present). The masking pattern remains flat versus frequency up to 500 Hz. Above 500 Hz, the masking pattern rises with increasing frequency at a rate of about 10 dB/decade. This experiment, in fact, reveals the Critical Bandwidth versus frequency.
The alteration of the hearing threshold around a sound can be thought of as a filter. Such a filter is referred to as a Critical Band, if equally distributed noise within that band is judged equally loud independent of bandwidth. Critical bands are not fixed filters, but form around the stimulus frequency. Critical bands for pure tones centered at 70, 250, 1k, and 4k Hz are shown. Note that critical bands are a measure of loudness or audibility not pitch discrimination.
In Psychoacoustics, the **Bark** scale (or Critical Band Rate scale) is used for frequency. The Bark scale is named after Barkhausen, who introduced the Phon. Critical bands at octave band centre frequencies are plotted on the Bark scale as the masking patterns of narrow-band noise of 1 Bark bandwidth. The dashed curve is the hearing threshold in quiet. On the Bark scale, critical bands appear as equal bandwidth at all frequencies. Although not fixed, for purposes of analysis, the critical bands can be thought of as a filter bank in the human hearing system.
Narrowband noise centered at 1 kHz is presented in 8 increasing relative bandwidths as shown. Each presentation is level adjusted for equal power. The loudness increases when the bandwidth exceeds a critical band (ca. 20%).
Critical Bandwidth versus frequency. At frequencies up to 500 Hz, critical bandwidth is constant at approximately 100 Hz. At higher frequencies, the bandwidth is relative (i.e., constant percentage bandwidth, approximately 20%). Therefore, above 500 Hz, the critical bands are often approximated with 1/3 octave-band filters, which have a relative bandwidth of 23%. ¼ octave is 17.3 %. The error in using 1/3 octaves to approximate critical bands is less than 3 dB above the 125 Hz band (Cohen & Fielder). There are two equations for the width of the critical band that model data taken from listening tests that give substantially different results:

Zwicker (Munich, 1961): $f_c = 25 + 75[1 + 1.4(f/1000)^2]^{0.69}$
Moore (Cambridge 1990): $f_c = 24.7[1+4.37(f/1000)]$
Loudness Model processing. First, the spectrum is corrected for the sound field (free field, diffuse field, or headphone listening) using a filter or look up table. For example, a narrow band of noise centered at 1 kHz results in the spectrum level shown in the upper left corner. This spectrum level causes the perceived excitation level, depicted as a masking pattern versus Equivalent Rectangular Bandwidth (ERB) or Critical Band. The excitation level in each ERB is converted to Specific Loudness as shown in the lower left graph. Bands above 500 Hz use the same conversion. The lower right graph depicts the Specific Loudness pattern, \( N' \), in sones/ERB. The Specific Loudness relates loudness to excitation within a single critical band or ERB. The Total Loudness, in sones, is the integral (area under the curve) of the Specific Loudness in all ERBs (or critical bands) and considers masking in adjacent critical bands (see ANSI S3.4, ISO 532b (DIN 45631)). The ANSI S3.4 method of Glasberg and Moore differs from the ISO 532b method of Zwicker, and will yield somewhat different results for the same input spectrum.
Simultaneous Masking causes low level sounds to be inaudible if they occur at the same time as the masker. Forward Masking occurs after the masking sound. In this case, the strength of the masking effect decays with time after the masking event. The processing latency in the ear is both frequency and level dependent, and louder and/or higher frequency sounds are transformed into sensation faster than quiet and/or low frequency sounds. This enables Backwards Masking, where soft sounds occurring before the masking event are inaudible.
The target signal is 10ms bursts of a 2 kHz tone in 10 decreasing steps of -4 dB. The target signal is then presented followed after a time gap, tau, by a narrow band burst noise centered at the same frequency (1900-2100 Hz). The time gap is successively, 100 ms, 20 ms, and 1 ms. The target signal becomes inaudible at a higher level for smaller time gaps.
Forward-masking thresholds for a 10 ms tone masked by a 250 ms noise burst as a function of the delay, tau, between the masker termination and the target tone. The level of a just audible tone impulse is plotted as a function of the delay time after the end of the masker for masker levels, $L_{WR}$, of 40, 60 and 80 dB SPL (RMS). Note the logarithmic time scale. Since the resulting curves are thresholds, it is seen, for example, that an tone of 70 dB presented up to 15 ms after a 60 dB (peak) masker would be inaudible. For comparison, an exponential decay with a 10 ms time constant is shown (dashed line).
The duration of the Forward-masking effect is also dependent upon the duration of the masker. The level of a just audible tone-burst is plotted as function of the delay time after the end of the masker for masker durations of 200 ms and 5 ms. Note the logarithmic time scale. In this case, release from masking occurs more rapidly for the shorter duration masker. This effect is also highly non-linear with level. If the level of the masker is decreased to 40 dB SPL, release from masking occurs in approximately 7 ms for the 5 ms masker, but occurs at about 20 ms for the 200 ms masker.
The perception of echoes or reflections depends upon the relative delay between the primary and secondary signal. The perceived direction of origin of a sound source is determined by the first arrival for signals of equal level. A phantom center image will shift toward the earliest arriving sound if a delay is introduced into one of the channels. Delays up to 5 ms can be compensated for by increasing the level of the delayed sound. For delays greater than 5 ms, a level difference of 10 dB will maintain the centered phantom image. Conversely, echoes delayed less than 25 ms and lower in level than the direct sound are generally not perceivable. This is known as the Haas or Precedence Effect. For example, the application of a small delay to the surround channel signals can be used to improve the directionality and clarity of front channel sounds in surround sound decoders.

**Experiment 1:** A stereo signal is heard with equal level impulses in both channels. The signal begins with a 60 ms delay to the Right channel, but each successive presentation decreases the delay by 10 ms until both impulses are heard simultaneously. The delay to Left channel is then increased in 10 ms steps. The apparent source location moves from the left to the right side.

**Experiment 2:** A stereo signal is presented with impulses in both channels and a constant delay of 15ms to Right channel. The level of the Left channel is held constant while the level of the right channel is increased by 2 dB with each successive presentation (from -8 dB to +10 dB). The apparent source location remains on the Left side until the level is sufficiently high, when the source then moves to the right side.
**Binaural** means “involving the use of both ears.” Most directional information is gathered binaurally. At low frequencies, the wavelength of sound is much larger than the head, so the sound can easily **diffract**, or bend around the body, which poses no significant obstacle. In this case, the ears perform a spatial sampling at two points within the same period. The primary lateral cue is the **Interaural Phase Difference (IPD)**. In general, localization at low frequencies is poor, decreasing with decreasing frequency, as the wavelength becomes larger than the size of the room. For this reason, the location of a hi-fi “subwoofer” is not critical. Example frequency is 200 Hz.
In the mid-frequency range, the wavelength of sound is nearly the same as the size of the head. A delay is perceived as sound passes from right to left. The primary lateral cue is the Interaural Time Delay (ITD). The transition between localization mechanisms is gradual. Example frequency is 2 kHz.
At higher frequencies, the wavelength of sound is smaller than the head. The head is a significant obstacle, attenuating the sound as it passes from right to left. The primary lateral cue is the **Interaural Level Difference (ILD)**. Phase discrimination is poor since spatial sampling occurs across several periods and is therefore ambiguous. Again, the transition between localization mechanisms is gradual. Example frequency is 6 kHz. Phase discrimination becomes poorer with increasing frequency. This phenomenon is exploited in lossy perceptual coders (e.g., MP3, AC-3, AAC, etc.), enabling the high frequency information from multiple channels to be combined or “coupled” together and phase information discarded.
The dashed green curve is the on-axis free field acoustic response to the ear entrance. As frequency increases, there is a pressure build up at the head where the wavelength is on the same order of magnitude as as the wavelength of sound. The red curve with circle markers is the response from the ear entrance to the ear drum. The increase in response at 3 kHz is due to a Helmholtz Resonance formed by the ear opening and the volume of the ear canal. The solid blue curve is the total response from an idealized on-axis free field point source to the ear drum. This response is also known as the Orthotelephonic Response. The increase in sensitivity around 3 kHz provides a S/N advantage for speech consonants. Data measured on a Brüel & Kjaer Type 4128 Head And Torso Simulator, free field.
The response of the ear is not the same for all angles of sound incidence. This graph shows the response in the median horizontal plane at 15° azimuth intervals. Dips at 4 kHz are cancellations due to shoulder reflections. Differences in response at higher frequencies are interpreted by the brain to give directional information. This is particularly important for elevation cues. Processing signals by filtering using HTRFs enables spatial illusions to be created where the sound seems to originate from a virtual source location. This is the principle exploited in surround sound “virtualizers”. Data measured on a Brüel & Kjær Type 4128 Head and Torso Simulator, free field.
Rather than free field, the response can be measured from in a diffuse field. This is equivalent to the power sum of the response from all angles of incidence. For microphones, this also called the Random Incidence Response. The response shown is taken from IEEE 1652, but is functionally equivalent to the data of Shaw (JASA 1974).
The **Auditory System** can be broken into three parts as shown. The anatomical dividing points are also indicated. Each part has a different function in the hearing process. These functions are described in detail in the following slides.
The human ear consists of three main parts: The Outer Ear, The Middle Ear, and the Inner Ear. The function of these parts is described in detail in the following slides.
Elements of the **Pinna** and **Outer** ear. At very high frequencies, reflections from the elements of the pinna provide up/down localization cues. The Pinna also functions as a horn or impedance matching device for sound entering the ear canal.
The outer ear consisting of the **Pinna** and the **Auditory Canal (External Acoustic Meatus)**, collects airborne sound waves which vibrate the **Tympanic Membrane**, or ear drum, which is the interface to the Middle Ear. The Middle Ear acts as an impedance matching device and acoustical-to-mechanical transducer, converting the sound waves into the mechanical motion of the **Ossicles**: The **Malleus**, the **Incus**, and the **Stapes**. The Ossicles – the three smallest bones in the human body – operate as a set of levers. To better understand their function, they are sometimes referred to as the Hammer, Anvil, and Stirrup, respectively. The middle ear is a sealed chamber and is therefore sensitive to changes in atmospheric pressure. The **Eustachian Tube** provides an equalization path to the throat to compensate for changes in ambient pressure. The ossicles transfer the vibration to the Inner Ear, which consists of two separate systems: the **Semi-Circular Canals** and the **Cochlea**. The semi-circular canals are three mutually perpendicular fluid-filled rings for controlling balance and orientation.
The Cochlea is a fluid filled, snail-shaped tubular structure which is divided longitudinally into two parts, the Scala Vestibuli and the Scala Tympani, by the Basilar Membrane. The Scala Media is the cavity of the cochlear duct containing the Organ of Corti. It is closed at the Helicotrema. The stapes rests on the Oval Window, at the entrance to the cochlea. In response to an acoustic stimulus, the stapes vibrates and disturbs the fluid in the Cochlea, which in turn distorts the basilar membrane. Arrows indicate the direction of sound propagation inside the Cochlea. The Round Window is an opening through the bone that separates the middle ear from the Scala Tympani. It also absorbs the incoming wave to prevent a reflection back through the Cochlea.
Inside the Scala Media, the **Organ of Corti** contains hair cells that respond to motion along the basilar membrane. The hair cells register this motion and transform it into nerve impulses which are then transmitted to the brain along the **Auditory Nerve**. Sensation occurs in the brain.
Inside the Scala Media, the **Organ of Corti** contains hair cells that respond to motion along the basilar membrane. The hair cells register this motion and transform it into nerve impulses which are then transmitted to the brain along the **Auditory Nerve**. Sensation occurs in the brain.
Research indicates that not all hair cells, or cilia, have the same function. The Inner Hairs Cells are responsible for converting motion along the basilar membrane into nerve impulses. This portion of the system has a dynamic range of approximately 50-60 dB. To accommodate the larger range of audible sound, it is believed that the Outer Hair Cells control dynamics – which in turn reduces sensitivity – for progressively louder sounds.
The diagram shows the malleus, incus, stapes, and “unwrapped” cochlea. Psychoacoustic masking experiments reveal that frequency analysis occurs along the length of the basilar membrane inside the cochlea. Each region is sensitive to a particular frequency band. Interestingly, high frequencies are processed nearest the oval window, at the entrance to the cochlea, while lower frequencies are processed further down the length of the basilar membrane. This means that the **Group Delay** through the cochlea is not uniform with frequency, since high frequencies are processed first. It also means that the regions sensitive to high frequencies are more prone to potentially damaging loud impulsive sounds, as the ear typically underestimates the level – and damage potential – of short duration sounds.
There is a finite processing delay from the time sound enters the ear under sensation occurs. 1) indicates the acoustic propagation delay from the outer ear to the ear drum. A deceleration can occur due to the temperature increase at the body (20°C to 37°C). 2) is the Middle Ear mechanical response. 3) represents frequency analysis and transduction in the cochlea. 4) is the processing time for sensation to occur at the brain. The Stapedius Reflex has a time constant of 10ms for high intensity sound and up to 150ms for lower intensity sounds. This small muscle responds to intense sounds by swinging the stapes away from the Oval Window, limiting the motion of the ossicles and thereby protecting the inner ear from damage, especially at low frequencies. Unfortunately, this reflex is too slow to respond to impulsive and high frequency sounds. Short, impulsive sounds are transmitted without obstruction to the inner ear, which is exposed to the full power of the signal, even though perceptually, the sensation – which takes place later in the brain -- may not be loud. As will be shown later, the damage risk of impulsive signals is severe.
The deflection of a hair cell is limited. Exposure to loud sounds for extended periods of time causes hair cells to break off. Hair cells also become brittle with aging and are prone to loss. Although there is some redundancy, loss of enough hairs cells in a specific region will cause hearing loss at the frequency range corresponding to that region along the basilar membrane. Hair cells, like other nerve cells, do not grow back after loss.
Audiology is the study of hearing. In Audiology, the graph of hearing range is inverted to show Hearing Loss instead of Sound Level. Note that the Vowels are lower in frequency and contain the majority of speech power but a minimum of intelligibility information while Consonant sounds are higher in frequency and contain less energy, but the majority of speech intelligibility. E.g., the difference between “tap”, “cap”, and “cat” or “gift”, “sift”, and “drift”, etc.
Hearing Loss can be due to many causes. Birth defects, infection, reaction or side effects of drugs or medication, normal aging, and noise exposure can all cause hair cell loss. A Temporary Threshold Shift can occur after exposure to loud sounds. This is due to fatigue or reversible injury. Permanent Hearing Loss can occur after exposure to loud noise for extended periods of time. OSHA and ISO have standardized Noise Dose criteria for steady noise exposure, trading level for time, e.g., 100% Noise Dose = 90 dB SPL (or 85 dB SPL) for 8 hours. Permanent Hearing Loss falls into two categories: Conductive and Sensorineural, described in detail in the following section. The Schwalbach Test is a diagnostic bone conduction test that is used to distinguish between conductive and sensorineural hearing loss.
Conductive Hearing Loss occurs at the Tympanic Membrane and Middle Ear and represents about 5% of hearing loss cases. Major causes are birth defects, ear wax (cerumen), infection to the outer or middle ear, hardening or perforation of the ear drum, cysts in the middle ear (cholesteatoma), or bony deposits at or near the stapes (otosclerosis). Conductive Hearing Loss represents an “open circuit” where the transmission of sound energy as vibration in the ossicles is prevented from reaching the inner ear. It can often be successfully treated medically.
Sensorineural Hearing Loss occurs in the Inner Ear and is caused by damaged hair cells. Sensorineural Hearing Loss represents about 95% of all hearing loss cases. Causes include aging (Presbycusis), noise exposure, and disease. Hair cell loss is PERMANENT!
The **Audiogram** is a graph showing hearing threshold level as a function of frequency. The minimum audible sound pressure level at each frequency is compared to the standardized average threshold of young persons with no hearing impairment. The difference is reported as the **Hearing Loss (HL)** at each frequency for each ear. The tests are conducted by an **Audiologist** using an **Audiometer** under quiet conditions in a sound proof booth to ensure that ambient noise does not invalidate the test. Tones at various levels and frequencies are presented to the listener through headphones or insert earphones. The subject responds using a handheld button indicating if the stimulus was audible. The resulting **Audiogram** is a graph of the listener’s **Hearing Threshold**. An additional diagnostic is to measure the **Threshold of Discomfort**.
Typical forward sloping mild high frequency sensorineural hearing loss. Note that the ability to hear consonant sounds at normal level is impaired. Reverse Slope Losses are less common. Notches in the audiogram ("Cookie Bites") are usually due to impulse noises such as gunshots. A Reverse Cookie Bite can also occur, but is rare. Example 1 is music (Berlioz, Symphonie Fantastique). Example 2 is the same music processed through a mild sloping hearing loss simulation. Example 3 is speech. Example 4 is the same speech processed through a mild sloping hearing loss simulation (Moore 1995, Tracks 2, 12, 62, & 72).
A flatter moderate sensorineural hearing loss. Note the increased downward shift of the hearing threshold and the loss of dynamic range, particularly at high frequencies. Most losses are not uniform across frequency. Audibility of consonants at normal levels is lost as well as the audibility of some vowel sounds. A person with this audiogram will have difficulty understanding speech in noise without a Hearing Aid. Hearing loss beyond this is considered Profound (or Severe). Example 1 is the same music as in the previous slide, but processed through a moderate flat hearing loss simulation. Example 2 is the speech sample from the previous slide, but processed through a moderate hearing loss simulation (Moore 1995, Tracks 4 & 64).
Another diagnostic test available to the audiologist is the measurement of **Loudness Growth** at various frequencies. Half-octave bands of random noise at various levels are presented to the subject. The subject presses a button indicating their judgment of the loudness. The yellow curve indicates loudness growth for a normal hearing person. The red curve indicates the loudness growth for an impaired listener. Soft sounds are inaudible while loud sounds are perceived as “too loud” due to the reduced dynamic range caused by hearing loss. **Recruitment** is the loss of the inner ear’s natural compressive function. Hair cell loss can occur anywhere along the length of the basilar membrane, at regions sensitive to different frequencies. Therefore, the most effective known method of compensation is to preprocess the signal before the ear with a multi-band wide dynamic range compression hearing device. This processing is similar to the encoding of ‘companding’ noise reduction systems.
Hearing Protectors  attenuate sound before it enters the ear. They are required in noisy occupations such as construction and at airports. The effectiveness of ear plugs and ear muffs depends on well they fit and how carefully they are worn. A loss of attenuation can occur if the path to ear is not tightly sealed. A foam insert ear plug and a circumaural ear muff are shown.
Different types of hearing protectors are available. As with any acoustic barrier, different types of ear protectors have different curves of attenuation versus frequency. An appropriate protector should be chosen for the sound field and exposure in a given environment. The Musician’s Plugs (e.g., Etymotic Research ER-15) are notable in that they have a “flat” attenuation versus frequency. Since high frequencies are not excessively attenuated, speech intelligibility does not suffer. Note that for optimum performance, Musician’s plugs require custom ear molds the same as a hearing aid.
References

ANSI S3.20 – Psychoacoustical Terminology.


ANSI Standard S3.4 – Procedure for the Computation of Loudness of Steady Sounds.


