

Thursday, March 22, 2012 – Stereophile

Parte prima = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-1.html>

Parte seconda = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-2.html>

Parte terza = <http://audioskeptic.blogspot.it/2012/03/final-steps-in-auditory-system-four-key.html>

Parte quarta = <http://audioskeptic.blogspot.it/2012/03/why-we-hear-what-we-hear-part-4-end.html>

Why We Hear What We Hear, Part 1,2,3 & 4

This overview explains why we hear what we hear and how you can manipulate it to get different results from multiple listenings.

Background and Notes about This Tutorial

This tutorial contains information garnered from 26 years as a research scientist at Bell Labs in Acoustic Research at Murray Hill, and its lineal descendants, from a variety of papers, from learning while I was an Audio Architect at Microsoft, and when I was Chief Scientist for Neural Audio. It contains ideas gathered from a variety of papers and experiments, done by many people, over a long period of time.

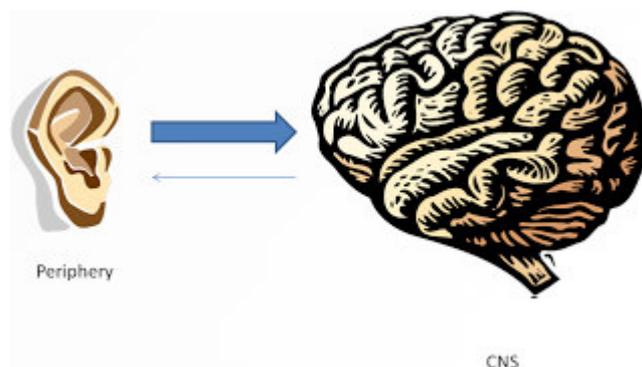
The information presented here is a work in progress, as research in both the hearing periphery, by which I mean the ear, up to and including the cochlea, and the Central Nervous System (CNS), that is the brain, continues today. I will present a high-level overview of a number of very broad set of subjects. These are phenomena that are observed by researchers in both the hearing and cognitive psychology communities. Each of these phenomena warrants in depth study, some of which will be covered by future tutorials.

Keep in mind, this information

- is not inviolate
- is a discussion of phenomena
- is an unknown mechanism, in most cases, once one gets beyond the basilar membrane
- will be revised by further research as time goes on, there will **always** be revisions

An Introduction to the Auditory System

The auditory system consists, in the way it is usually examined, of two parts, the periphery (head, ear, cochlea), and the brain, or Central Nervous system (CNS). The flow of information is substantially one-sided, with the ear providing a lot of information to the brain, and the brain providing very little feedback (some loudness issues, body and head movement) to the periphery, relatively speaking.



What is the Periphery?

It is all of the hearing apparatus that is outside of the brain. I am not, for this purpose, including body and skin sensation, which operate mostly at higher sounds levels than we should be listening to. I'll break this into three distinct sections

- HRTF's, including ear canal (outer and middle ear functions)
- Cochlear analysis (inner ear)
- Reduction of sound into partial loudnesses as a function of time (inner ear)

First, the physical shape of the head, ear, body, and surrounding environment create what's called a Head Related Transfer Function (HRTF) for each ear. You can think of the HRTF as a frequency response that varies with distance and angle between the sound source and the head, providing an Interaural Level Difference (ILD) between each ear. The HRTFs also create the Interaural Time Delay (ITD) information

gathered by the physical acoustics. The ITD and ILD together are the information gathered by physical acoustics for directional processing by the CNS, as I will explain later.

All the effects of the outer and middle ear are being simplified and lumped into the HRTF for this tutorial. Conventionally in the literature, the functions of the outer and middle ear are separated from each other. What does the middle ear contribute to the HRTF? The ear canal influences the HRTF by its resonance, which is a function of its length and width. The eardrum and 3 bones of the middle ear influence the HRTF by changing the impedance of the system and rejecting “near DC” components of sound. In other words, it’s similar to a transformer, modifying the relationship between force and distance as well as filtering out very low frequencies like changes in barometric pressure. The middle ear’s rejection of near-DC signals is absolutely essential; otherwise a passing weather front would amount to a 150dB sound level, intolerably loud and damaging to our hearing.

Second, the cochlea filters the time signal into many overlapping bands. The cochlea is a complicated organ that performs a mechanical filtering of the signal entering the ear. It filters the incoming signal into heavily overlapping critical bands or Equivalent Rectangular Bandwidths (ERBs). This filtering gives us our frequency sensitivity above about 100Hz.

Third, the signal in each ERB is compressed to give us a reduction of sound into partial loudnesses as a function of time. This ties changes in atmospheric pressure to sensation. A Sound Pressure Level (SPL) is a physical quantity, a measured change in physical atmospheric pressure. A partial loudness is a measure of the sensation level detected by an inner hair cell, a perceptual quantity. An intensity increase of 10dB creates a change in sensation level of about a factor of 2. A signal on one ERB does not compress the signals that do not enter into the same filter, so we are left with an interesting effect: signals are compressed inside an ERB, but partial loudnesses from 2 ERBs add! The sum of all the partial loudnesses is in fact what we think of as loudness.

Please notice that I’ve used two words in a very specific way. To be clear,

- **loudness** is the level you hear, the sensation level that gets shipped to your CNS
- **intensity** is the measured signal level in the atmosphere, the SPL

The two do not track very well, even with knowledge of the compression mentioned above. The only time they track each other is when the frequency response of the two signals being compared have the same shape and only the total energy of the two signals are different. A measure of SPL is not enough to determine loudness by itself.

For further information about loudness, see the loudness tutorial of April 2006 at the [Pacific Northwest AES web site](#) (Loudness Tutorial PowerPoint presentations [without sound](#), [with audio](#)). There are a variety of slide decks there on the subjects of audio and hearing, which you may find interesting.

What is the Central Nervous System?

The central nervous system (CNS) consists of the end of the auditory nerve where it connects to the brain and then the brain. It carries out, in some fashion that we do not well understand, the following operations:

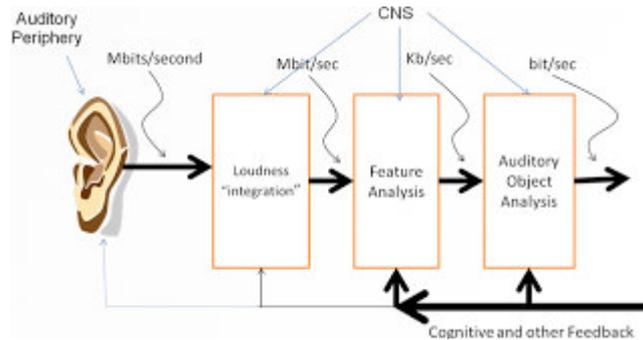
- Reduction from partial loudness to auditory features
- Reduction of auditory features to auditory objects
- Storage in short-term and long-term memory

There are some well-known issues with the CNS. In particular, it’s very flexible in the way it interprets input. This is sometimes referred to as “plastic”. What plastic means is you can change what you listen to, what you look at, what you smell or feel, and the CNS, by design, combines information from all sensory modalities. It does this all of the time under all circumstances. Everywhere. All the time.

Plasticity, by itself, creates a problem with isolating what’s going on in any one bit of audio, be it due to equipment, performance, or what-have-you, because the mere knowledge of which of a set of things you are listening to will change the way you think about it, and hence what you pay attention to. This is not a question of “Heisenberg” kinds of uncertainty, rather it is a simple case you notice the things you focus on. What information gets to the CNS and when does it get there? Anything detected by the auditory periphery, the visual periphery, knowledge of what button you pushed on the amplifier, the color of the speaker grill cloth, and so on gets to the CNS. However, what is specifically important for auditory sensation is the information from the auditory periphery.

Here, we are not going to consider body sensation (although it is certainly germane at low frequencies at high levels, or at ultrasonic frequencies at extreme levels). We will also leave out extremely intense LF and VHF signals, which can be detected by other means; these are extreme conditions and should not generally be experienced by a listener.

To summarize, the CNS and its connection to the audio periphery can be illustrated as



Four Key Points

Here are four key points about the auditory periphery.

- The auditory periphery analyzes all signals in a time/frequency tiling called ERBs or Barks.
- Due to the mechanics of the cochlea, first arrivals have very strong, seemingly disproportionate influence on what you actually hear. This is actually useful in the real world.
- Signals inside an ERB mutually compress.
- Signals outside an ERB do not mutually compress.

The first arrival information in a given ERB, where the first arrival is the first signal in approximately the last 200 milliseconds, is emphasized by the mechanics of the cochlea. The compression of signals inside an ERB starts to take effect about 1 millisecond after the arrival of the sound, so the first part of a sound sends more information to the brain. This turns out to be very useful to us in distinguishing both perceived direction and diffuse sensation.

The partial loudnesses from the cochlea are integrated somewhere at the very edge of the CNS such that some memory of the past is maintained for up to 200 milliseconds. Level Roving Experiments show that when delays approaching 200 milliseconds exist between two sources, the ability to discern fine differences in loudness or timbre is reduced. It is well established you need very quick, click-less switching between signals when trying to detect very small differences between signals, otherwise you lose part of your ability to distinguish loudness differences.

Final Steps

There are two steps remaining in what you hear, both of them executed by the brain in a fashion I do not care to even guess about. The first step is analysis of the somewhat-integrated partial loudnesses into what I call "auditory features". There is a great deal of data loss at this juncture, about 1/1000th of the information present at the auditory nerve remains after feature reduction, the rest being integrated or discarded into the information that remains. At this level, feature analysis is extremely plastic and can be guided by learning, experience, cognition, reflex, visual stimuli, state of mind, comfort, and all other factors. The features last a few seconds in short-term memory.

In the second step, the information from feature analysis is again reduced in size by about 100 to 1000 times, and turned into what I refer to as "auditory objects". These are things that one can consciously describe, control, interpret, etc. Words are examples of things made of successive auditory objects. This process is as plastic as plastic can be, you can redirect yourself cognitively, be directed by visual stimuli, guesses, unconscious stimuli, including randomness, and is the final step before auditory input can be converted to long-term memory. Interestingly, this process can promote a feature to an object if you consciously focus on that feature. It is this process that is most affected by every other stimulus, including those generated internally. It is at these last two points, where short-term loudness is reduced to features and then objects, where we integrate the results from our senses and our knowledge, regardless of our intent or situation.

What This Means about What You Hear

The plasticity of forming auditory features and auditory objects has some very strong implications. If you listen for different things, you will hear different things. This is not illusion, it is not confusion, it is not deception, it is not hallucination, it is simply how your brain works. This has a particularly important implication for audio enthusiasts, which is **expectation will always cause you to hear things differently**. The effect of expectation is not always positive, or negative, for example you may not hear what you expect to hear, but expectation can not, not now, not ever, be consciously filtered out of your hearing experience. Expectation is why audio testing needs to be what's called "double blind". Double blind does not mean that you close your eyes, wear a blindfold, or have anything to do with vision. Blind means that when you are trying to identify something you do not know which of several (usually two) signals it is, even though you

can, or should be able to, switch to each of the known signals at will for reference and comparison. Double means that any experimenter you interact with in any fashion also doesn't know which you're hearing. Lots of work has shown that cues from an experimenter will lead you to answer differently. Interestingly, you may not agree with the experimenter in a single-blind test, you may take the other answer than the one the experimenter prefers, but your performance will be affected.

This can be demonstrated by the old "backward masking" demonstration that was originally brought up in a different context, where it was alleged that there was satanic content to part of the song "Stairway to Heaven" when it was played in reverse. Interestingly enough, when the reversed song is played to an unsuspecting audience (this has happened quite a few times in lectures), the audience hears nothing, or a very few random syllables. When, however, the "words" are presented, the audience hears the words clearly and correctly, even though the actual sound presented to the ear is not changed. This is a direct result of how we understand speech, when expecting speech with some particular content, we guide our feature extraction and object resolution to find the parts of our audio surroundings in order to extract the information, and in this case can do so even when there is no such information. This kind of effect is echoed in "EVP" arguments, wherein sounds, whispers, etc., are supposedly heard in recordings of radio static or other random noise.

In summary, the processing done in the brain is exceptionally plastic, and can be guided by a variety of things. The result of all that processing is what we actually, consciously hear.

- If you listen to something differently (for different features or objects)
- You will **remember** different things
- This is not an illusion
- If you have reason to assume things may be different
- You will most likely listen differently
- Therefore, you will remember different things

In short, if you want to be sure of what you heard due to only the auditory stimuli, you need to arrange a blind test of an appropriate nature. Such tests are neither easy nor simple to arrange, but are the best, and perhaps only way to avoid inadvertent self-deception.

The end