

The world beyond 20kHz

by David E. Blackmer

Using a study of the human hearing mechanism as his foundation, Earthworks' founder David E Blackmer presents his arguments for, and his vision of, high-definition audio.

THERE IS MUCH controversy about how we might move forward towards higher quality reproduction of sound. The compact-disc standard assumes that there is no useful information beyond 20kHz and therefore includes a brick-wall filter just above 20kHz. Many listeners hear a great difference when 20kHz band-limited audio signals are compared with wide band signals. A number of digital systems have been proposed which sample audio signals at 96kHz and above, and with up to 24 bits of quantisation.

Many engineers have been trained to believe that human hearing receives no meaningful input from frequency components above 20kHz. I have read many irate letters from such engineers insisting that information above 20kHz is clearly useless, and any attempts to include such information in audio signals is deceptive, wasteful and foolish, and that any right-minded audio engineer should realize that this 20kHz limitation has been known to be an absolute limitation for many decades. Those of us who are convinced that there is critically important audio information to at least 40kHz are viewed as misguided.

We must look at the mechanisms involved in hearing, and attempt to understand them. Through that understanding we can develop a model of the capabilities of the transduction and analysis systems in human audition and work toward new and better standards for audio system design.

What got me started in my quest to understand the capabilities of human hearing beyond 20kHz was an incident in the late eighties. I had just acquired a MLSSA system and was comparing the sound and response of a group of high quality dome tweeters. The best of these had virtually identical frequency response to 20kHz, yet they sounded very different.

When I looked closely at their response beyond 20kHz they were visibly quite different. The metal-dome tweeters had an irregular picket fence of peaks and valleys in their amplitude response above 20kHz. The silk-dome tweeters exhibited a smooth fall off above 20kHz. The metal dome sounded harsh compared to the silk dome. How could this be? I cannot hear tones even to 20kHz, and yet the difference was audible and really quite drastic. Rather than denying what I clearly heard, I started looking for other explanations.

WHEN VIEWED FROM an evolutionary stand point, human hearing has become what it is because it is a survival tool. The human auditory sense is very effective at extracting every possible detail from the world around us so that we and our ancestors might avoid danger, find food, communicate, enjoy the sounds of nature, and appreciate the beauty of what we call music. Human hearing is generally, I believe, misunderstood to be primarily a frequency analysis system. The prevalent model of human hearing presumes that auditory perception is based on the brain's interpretation of the outputs of a frequency analysis system which is essentially a wide dynamic range comb filter, wherein the intensity of each frequency component is transmitted to the brain. This comb filter is certainly an important part of our sound analysis system, and what an amazing filter it is. Each frequency zone is tuned sharply with a negative mechanical resistance system. Furthermore, the tuning Q of each filter element is adjusted in accordance with commands sent back to the cochlea by a series of pre-analysis centers (the cochlear nuclei) near the brain stem. A number of very fast transmission-rate nerve fibers connect the output of each hair cell to these cochlear nuclei. The human ability to interpret frequency information is amazing. Clearly, however, something is going on that

cannot be explained entirely in terms of our ability to hear tones.

The inner ear is a complex device with incredible details in its construction. Acoustical pressure waves are converted into nerve pulses in the inner ear, specifically in the cochlea, which is a liquid filled spiral tube. The acoustic signal is received by the tympanic membrane where it is converted to mechanical forces which are transmitted to the oval window then into the cochlea where the pressure waves pass along the basilar membrane. This basilar membrane is an acoustically active transmission device. Along the basilar membrane are rows of two different types of hair cells, usually referred to as inner and outer.

The inner hair cells clearly relate to the frequency analysis system described above. Only about 3,000 of the 15,000 hair cells on the basilar membrane are involved in transducing frequency information using the outputs of this travelling wave filter. The outer hair cells clearly do something else, but what?

There are about 12,000 'outer' hair cells arranged in three or four rows. There are four times as many outer hair cells as inner hair cells(!) However, only about 20% of the total available nerve paths connect them to the brain. The outer hair cells are interconnected by nerve fibers in a distributed network. This array seems to act as a waveform analyzer, a low-frequency transducer, and as a command center for the super fast muscle fibers (actin) which amplify and sharpen the travelling waves which pass along the basilar membrane thereby producing the comb filter. It also has the ability to extract information and transmit it to the analysis centers in the olivary complex, and then on to the cortex of the brain where conscious awareness of sonic patterns takes place. The information from the outer hair cells, which seems to be more related to waveform than frequency, is certainly correlated with the frequency domain and other information in the brain to produce the auditory sense.

Our auditory analysis system is extraordinarily sensitive to boundaries (any significant initial or final event or point of change). One result of this boundary detection process is the much greater awareness of the initial sound in a complex series of sounds such as a reverberant sound field. This initial sound component is responsible for most of our sense of content, meaning, and frequency balance in a complex signal. The human auditory system is evidently sensitive to impulse information imbedded in the tones. My suspicion is that this sense is behind what is commonly referred to as 'air' in the high-end literature. It probably also relates to what we think of as 'texture' and 'timbre' - that which gives each sound its distinctive individual character. Whatever we call it, I suggest that impulse information is an important part of how humans hear.

All the output signals from the cochlea are transmitted on nerve fibers as pulse rate and pulse position modulated signals. These signals are used to transduce information about frequency, intensity, waveform, rate of change and time. The lower frequencies are transduced to nerve impulses in the auditory system in a surprising way. Hair cell output for the lower frequencies are transmitted primarily as groups of pulses which correspond strongly to the positive half of the acoustic pressure wave with few if any pulses being transmitted during the negative half of the pressure wave. Effectively, these nerve fibers transmit on the positive half wave only. This situation exists up to somewhat above 1kHz with discernable half wave peaks riding on top of the auditory nerve signal being clearly visible to at least 5kHz. There is a sharp boundary at the beginning and end of each positive pressure pulse group, approximately at the central axis of the pressure wave. This pulse group transduction with sharp boundaries at the axis is one of the important mechanisms which accounts for the time resolution of the human ear. In 1929 Von Békésy published a measurement of the human sound position acuity which translates to a time resolution of better than 10 μ s between the ears. Nordmark, in a 1976 article, concluded that the intramural resolution is better than 2 μ s; intramural time resolution at 250Hz is said to be about 10 μ s which translates to better than 1° of phase at this frequency.

The human hearing system uses waveform as well as frequency to analyze signals. It is important to maintain accurate waveform up to the highest frequency region with accurate reproduction of details down to 5 μ s to 10 μ s. The accuracy of low frequency details is equally important. We find many low frequency

sounds such as drums take on a remarkable strength and emotional impact when waveform is exactly reproduced. Please notice the exceptional drum sounds on The Dead Can Dance CD Into the Labyrinth. The drum sound seems to have a very low fundamental, maybe about 20Hz. We sampled the bitstream from this sound and found that the first positive waveform had twice the period of the subsequent 40Hz waveform. Apparently one half cycle of 20Hz was enough to cause the entire sound to seem to have a 20Hz fundamental.

The human auditory system, both inner and outer hair cells, can analyze hundreds of nearly simultaneous sound components, identifying the source location, frequency, time, intensity, and transient events in each of these many sounds simultaneously and develop a detailed spatial map of all these sounds with awareness of each sound source, its position, character, timbre, loudness, and all other identification labels which we can attach to sonic sources and events. I believe that this sound quality information includes waveform, embedded transient identification, and high frequency component identification to at least 40kHz (even if you can't 'hear' these frequencies in isolated form).

TO FULLY MEET the requirements of human auditory perception I believe that a sound system must cover the frequency range of about 15Hz to at least 40kHz (some say 80kHz or more) with over 120dB dynamic range to properly handle transient peaks and with a transient time accuracy of a few microseconds at high frequencies and 1° - 2° phase accuracy down to 30Hz. This standard is beyond the capabilities of present day systems but it is most important that we understand the degradation of perceived sound quality that results from the compromises being made in the sound delivery systems now in use. The transducers are the most obvious problem areas, but the storage systems and all the electronics and interconnections are important as well.

Our goal at Earthworks is to produce audio tools which are far more accurate than the older equipment we grew up on. We are certainly pushing the envelope. For example, we specify our LAB102 preamp from 2Hz to 100kHz \pm 0.1dB. Some might believe that this wide range performance to be unimportant, but listen to the sound of the LAB102, it is true-to-life accurate. In fact the 1dB down points of the LAB preamp are 0.4Hz and 1.3MHz, but that is not the key to its accuracy. Its square wave rise time is one quarter of a microsecond. Its impulse response is practically perfect.

Microphones are the first link in the audio chain, translating the pressure waves in the air into electrical signals. Most of today's microphones are not very accurate. Very few have good frequency response over the entire 15Hz-40kHz range which I believe to be necessary for accurate sound. In most microphones the active acoustic device is a diaphragm that receives the acoustical waves, and like a drum head it will ring when struck. To make matters worse, the pickup capsule is usually housed in a cage with many internal resonances and reflections which further color the sound. Directional microphones, because they achieve directionality by sampling the sound at multiple points, are by nature less accurate than omnis. The ringing, reflections and multiple paths to the diaphragm add up to excess phase. These microphones smear the signal in the time domain.

We have learned after many measurements and careful listening that the true impulse response of microphones is a better indicator of sound quality than is frequency amplitude response. Microphones with long and non-symmetrical impulse performance will be more colored than those with short impulse tails. To illustrate this point we have carefully recorded a variety of sources using two different omni models (Earthworks QTC1 and another well-known model) both of which have flat frequency response to 40kHz within -1dB. (Fig. 1: QTC1 vs 4007). When played back on high-quality speakers the sound of these two microphones is quite different. When played back on speakers with near-perfect impulse and step response, which we have in our lab, the difference is even more apparent. The only significant difference we have

been able to identify between these two microphones is their impulse response.

We have developed a system for deriving a microphone's frequency response from its impulse response. After numerous comparisons between the results of our impulse conversion and the results of the more common substitution method we are convinced of the validity of this as a primary standard. You will see several examples of this in Fig.2.

Viewing the waveform as impulse response is better for interpreting higher frequency information. Lower frequency information is more easily understood from inspecting the step-function response which is the mathematical integral of impulse response. Both curves contain all information about frequency and time response within the limits imposed by the time window, the sampling processes and noise.

The electronics in very high quality sound systems must also be exceptional. Distortion and transient intermodulation should be held to a few parts per million in each amplification stage, especially in systems with many amplifiers in each chain. In the internal circuit design of audio amplifiers it is especially important to separate the signal reference point in each stage from the power supply return currents which are usually terribly nonlinear. Difference input circuits on each stage should extract the true signal from the previous stage in the amplifier. Any overall feedback must reference from the output terminals and compare directly to the input terminals to prevent admixture of ground grunge and cross-talk with the signal. Failure to observe these rules results in a harsh 'transistor sound'. However, transistors can be used in a manner that results in an arbitrarily low distortion, intermodulation, power supply noise coupling, and whatever other errors we can name, and can therefore deliver perceptual perfection in audio signal amplification. (I use 'perceptual perfection' to mean a system or component so excellent that it has no error that could possibly be perceived by human hearing at its best.) My current design objective on amplifiers is to have all harmonic distortion including 19kHz and 20kHz twin-tone intermodulation products below 1 part per million and to have A-weighted noise at least 130dB below maximum sine wave output. I assume that a signal can go through many such amplifiers in a system with no detectable degradation in signal quality.

Many audio signal sources have extremely high transient peaks, often as high as 20dB above the level read on a volume indicator. It is important to have some adequate measurement tool in an audio amplification system to measure peaks and to determine that they are being handled appropriately. Many of the available peak reading meters do not read true instantaneous peak levels, but respond to something closer to a 300 μ s to 1ms averaged peak approximation. All system components including power amplifiers and speakers should be designed to reproduce the original peaks accurately. Recording systems truncate peaks which are beyond their capability. Analogue tape recorders often have a smooth compression of peaks which is often regarded as less damaging to the sound.

MANY RECORDISTS even like this peak clipping and use it intentionally. Most digital recorders have a brick-wall effect in which any excess peaks are squared off with disastrous effects on tweeters, and listener's ears. Compressors and limiters are often used to smoothly reduce peaks which would otherwise be beyond the capability of the system. Such units with RMS level detectors usually sound better than those with average or quasi-peak detectors. Also, be careful to select signal processors for low distortion. If they are well designed, distortion will be very low when no gain change is required. Distortion during compression will be almost entirely third harmonic distortion which is not easily detected by the ear and which is usually acceptable when it can be heard.

A look at the specifications of some of the highly rated super-high end, 'no feedback', vacuum tube, power amplifiers reveals how much distortion is acceptable, or even preferable, to some excessively well-heeled audiophiles.

All connections between different parts of the electrical system must be designed to eliminate noise and signal errors due to power line ground currents, AC magnetic fields, RF pickup, crosstalk, and dielectric absorption effects in wire insulation. This is critical.

Loudspeakers are the other end of the audio system. They convert electrical signals into pressure waves in the air. Loudspeakers are usually even less accurate than microphones. Making a loudspeaker that meets the standard mentioned above is problematical. The ideal speaker is a point source. As yet no single driver exists that can accurately reproduce the entire 15Hz–40kHz range. All multidriver speaker systems involve tradeoffs and compromises.

We have built several experimental speaker systems which apply the same time–domain principles used in our Earthworks microphones. The results have been very promising. As we approach perfect impulse and step–function response something magical happens. The sound quality becomes lifelike. In a live jazz sound–reinforcement situation using some of our experimental speakers and our SR71 mics the sound quality did not change with amplification. From the audience it sounded as if it was not being amplified at all even though we were acutely aware that the sound was louder. Even with quite a bit of gain it did not sound like it was going through loudspeakers.

Listening to some Bach choral music that we recorded with QTC1 microphones into a 96kHz sampling recorder, and played back through our engineering model speakers is an startling experience. The detail and imaging are stunning. You can hear left to right, front to back and top to bottom as if you are there in the room with the performers. It is exciting to find that we are making such good progress toward our goal.

I have heard that the Victor Talking Machine Company ran ads in the 1920s in which Enrico Caruso was quoted as saying that the Victrola was so good that its sound was indistinguishable from his own voice live. In the seventies Acoustic Research ran similar ads, with considerably more justification, about live vs recorded string quartets. We have come a long way since then, but can we achieve perceptual perfection? I suspect that truly excellent sound, perhaps even perceptual perfection? As a point of reference you should assemble a test system with both microphones and speakers having excellent impulse and step response, hence nearly perfect frequency response, together with low distortion amplifiers. Test it as a sound reinforcement system and/ or studio monitoring system with both voice and music sources. You, the performers, and the audience will be amazed at the result. You don't have such a system? Isn't that impossible, you say? It is not! We have done it! If you want more information, here are several books which I believe anyone who is intensely involved in Audio should own and read and then reread many times.

An Introduction to the Physiology of Hearing, 2nd edition

James O. Pickles, Academic Press 1988

ISBN 0-12-554753-6 or ISBN 0-12-554754-4 pbk.

Spacial Hearing, revised edition

Jen Blauert, MIT Press 1997

ISBN 0-262-02413-6

Experiments in Hearing, Georg von Békésy

Acoustical Society of America

ISBN 0-88318-630-6

Hearing, Gulick et al

Oxford University Press 1989

ISBN 0-19-50307-3