

A NeuroSafe Brochure

The Science of Human Sound Experience

William Softky

This Brochure describes how the history of sound engineering created audio technologies which actively undermine human sonic perception and connection. If you understand how recorded and transmitted sound fails our nervous system's deepest needs, you can appreciate how NeuroSafe technologies and environments can help.

Neil Young was a famous rock musician in the 1970s, specializing in live performance and weird acoustic spaces, like the echo-filled iron sawdust burner I once camped in as a kid. In a recent interview for The New York Times Magazine¹, he claimed that digital compression technology — CD, MP3, streaming — undermines human dignity. Of the thousands of comments in response, many readers denounced him as emotional, anti-scientific, a Luddite and even partly deaf. But might Young know something the rest of us don't?

Put another way, if a sensitive, world-acclaimed innovator denounces his industry and its technology for undermining human dignity and brain function, something big is up. Who could be more qualified than a world expert — with loads of experience and no incentive to fib — to call the alarm about widespread technological damage?

Young isn't the first to denounce digital and acclaim analog. Legions of self-proclaimed audiophiles have lamented the loss of vinyl LPs since digital CDs first appeared in the 1970s. Likewise, legions of people who grew up talking for hours on old-fashioned analog telephones hate talking on digital cellphones now. For decades, there has been a deep-set conflict between those who claim analog has some ineffable "presence" missing in digital versus those with technical know-how who can explain how analog and digital actually work. In essence, the producers of digital tech claim it is flawless, while the most sensitive consumers claim it is awful. They can't both be right.

¹ New York Times magazine, August 20, 2019

There is a truism that technology often fails to live up to its promise. But utterly failing at its central task is something else. Car buyers don't claim cars fail to move and shoe wearers don't claim shoes cut their feet. Yet consumers and especially producers of the most spiritual and emotional human communication — music — routinely say digitization destroys emotion. For those to whom emotion matters, this problem is bigger than cars or shoes. Science needs to get emotion right, and sound is in the way.

What Ordinary People Say

Since we're talking about everybody's emotions, not just Neil Young's, let's look at the spectrum of reaction to his interview. About half of commenters agreed with him, often saying lyrical things like, "In analog music, you can hear 'subtle differences,' so voices and instruments 'sound more natural,' even 'glorious,' with more 'color,' 'depth,' and 'beauty.'" Other readers wrote, "Just because one cannot hear or discern the missing info doesn't mean it does not exist" and that poor "audio quality leads to 'listener fatigue.'"

The anti-Neil Young comments don't sound as much like music appreciation. Rather, they have the snide sound of an annoying know-it-all engineer. To wit: "Digital recordings are better and hold more information," "Fidelity on a basic CD is 'altogether superior' to any vinyl LP," and the "Nyquist theorem proves that analog information is perfectly preserved in digital form, and naysayers 'disagree with math and science.'"

While such comments can be annoying, they do invoke a kind of truth sacred to technologists: Certain kinds of math and science are objectively true. You can't ignore physics and stay intellectually legitimate.

Of course, if you invoke the wrong law of physics, your ideas might be useless, or worse. In the case of sound and sound reproduction, there are two laws of nature at work, the first of which is the Nyquist Theorem mentioned above.

Overall, the comments cluster into complementary philosophies, which might be called theory vs. experiment." They raise the deepest issues possible pertaining to interpreting reality: How much should we trust the laws of nature versus how much should we trust our own sensory experience? Of course, people who notice differences in sonic texture will say so, as the reader comments above refer to the experiential data of "color," "depth" and such. But other people, those who know certain laws of

physics always apply, will assert the primacy of natural law bluntly, as one might assert the self-evidence of $2+2=4$.

The unfortunate asymmetry of such discussions is that while one individual sensory experience doesn't speak to anyone else's, and thus doesn't insult them, the reverse isn't true. Invoking a law of nature that says certain experiences can't possibly exist does, in fact, insult anyone having that experience. In such discussions, the "scientists," in effect, call the "musicians" stupid, insensitive and/or hallucinating.

It's not just musicians who don't want to be told by scientists that their experiences are hallucinations; it's even other scientists. For instance, neuroscientists know their results must accord with the laws of physics, but the reverse isn't true. Physics doesn't have to "obey" neuroscience laws, because there are no neuroscience laws.

Neuroscience is an experimental discipline, 150 years of accumulated interesting results, with very little agreement beyond the usual mantra that "more research is necessary." In the hierarchy by which abstract theory intersects with specific experiments, math trumps physics, physics trumps chemistry, chemistry trumps biology and biology trumps psychology. So even inside science, as a general rule, established laws of math and nature trump experimental results.

In that sense, the discussion between Young and his detractors is truly scientific. And, in that sense, science can resolve it. But first, a little history.

The Laws of Sound and Math

While Thomas Edison is credited with inventing both movies and phonographs, the two techniques couldn't be more different. Movie technology freeze-frames analog continuous-time into discrete, discontinuous snapshots. In film, analog 3D life is compressed into separate analog 2D pictures. Time, on the other hand, is chopped up (i.e., digitized). Film captures individual photographs (i.e., two-dimensional gray-scale (analog) images spaced a few dozen milliseconds apart).

Contrast those intermittent photographs with phonographs, in which time is continuous. Phonographs, the ancestors of vinyl LPs, originally scratched tiny but continuous sound vibrations directly, mechanically, onto a cylinder of tin foil, then replayed them in reverse, with a horn-shaped "loudspeaker" — like the old Victrola players listened to by dogs in the ads (above the slogan, "His Master's Voice," because dogs actually could recognize people in those recordings). The cylinders had only one

groove, hence they had no stereo and no possible representation of space. But they recorded enough pure, continuous analog time to evoke instant recognition, which is the sweet elixir of the nervous system for humans and animals alike.

During the 20th century, analog sound recording vaulted from foil cylinders to plastic platters, ever larger and spinning ever slower as technology improved, at 78, 45 or ultimately 33 1/3 revolutions per minute (33 1/3, the slowest speed, was called “long-playing,” or LP). At roughly the same time, analog sound transmission wrapped the world in two other ways.

The continuous-wave analog radio (first AM and then FM), invented first by Nikola Tesla and popularized by Guglielmo Marconi, mimicked the abilities and peculiarities of analog recording with its tubes, crystals and resonators. While audio recording moved sound across time, radio moved it across space and could have an enormous impact when blanketing a country. Orson Welles’ voice and genius — on the radio — took less than an hour to convince America we were under Martian attack. Franklin D. Roosevelt and Winston Churchill’s voices, also on radio, inspired citizens in time of war. So did Adolf Hitler’s and Benito Mussolini’s. Broadcast voices move whole peoples.

Analog emotional resonance also worked through phones. The old-school “landline” phones (aka, the plain old telephone service, or POTS), connected two people in real-time through a pair of dedicated, analog copper wires. Those of us who grew up with POTS phones happily spent hours on them because the connection was so good: no dropping, no gargling, no lost words, no weird sounds or silences. Ask anyone who lived that to confirm it. Talking on POTS phones was like having your conversational partner whisper in your ear.

Here’s one bit of historical trivia: Even the earliest connections in the 1920s were so continuously lifelike that musicians who had access to shared “party” lines would often practice music over the phone, rather than travel to each other’s houses. But that practice disrupted paid phone traffic. Jam-sessions and day-long conversations over party lines eventually became so widespread that the telephone company, Ma Bell, had to run a campaign to discourage them, like this amusing promotional message from the 1940s: *Bobby Gets Hep.*²

Even before the invention of stereo, the decades-long experience optimizing monaural sound in all its forms taught technologists a lot about human hearing. They

² Bell Telephone System, 1946

discovered that our ears can't hear at all above 20,000 Hz (back-and-forth cycles per second), so they knew not to waste any effort at all above that threshold. They also discovered that we only really need frequencies below about 3000 Hz to understand spoken speech, that the consonants have much higher frequencies than the vowels, and that consonants can never be dispensed with (while vowels can, as in whispering).

Then stereo arrived, enabled by the singular invention of having the same needle move in two directions at once, 90 degrees apart, to carry two independent channels. Because individual vibrations from the same needle in the same groove stayed in stereo synchrony at the micro-second level, someone listening to a high-fidelity player could actually hear where in space the sound came from, not just what object made the sound.

After stereo, the next innovations were digital ones: how to digitize sound (analog-to-digital conversion, or ADC), how to reconstitute it for recognition (digital-to-analog, or DAC), and later how to most cost-effectively "compress" the sound in between for storage and transmission. Engineers finally learned how to throw large parts of the signal away without anyone noticing. There were thus new choices about new informational concepts like bandwidth, frequency limits, bit-rates, sampling rates, encoding depth and, most crucially, deciding what is "signal" vs. what is "noise." Making new choices required new principles.

Based on the existing, already-deep understanding of single-channel "mono" sound, the founding principle of compression became fidelity: to preserve the best possible rendition of sound consistent with our ears' limits, in the service of better recognition, identification and description of the source. Who is making the sound? What are they saying or singing? What do they mean?

Of course, balancing bandwidth vs. fidelity required engineering tradeoffs, based on needs and costs. They did so by respecting the old laws of math and physics, plus the new "laws" of information. These laws, discovered about this time by mathematician Claude Shannon, turned the woolly concept of "information" into a quantified technological term by defining bits, bandwidth, signal and noise in terms of analog statistical probabilities. Weirdly, while Shannon's equations are central to the discrete bits and messages of digital communication, the equations themselves are analog, crossing the divide between continuous nature and manmade messages plucked out of nature.

Another law, the Nyquist Theorem, was more human-centered. People already knew from experience and experiment that high-frequency signals carry more information

than low-frequency ones. This is why speech needs consonants more than vowels, and why hi-fi recordings sound more realistic than low-fi ones. Shannon knew there were deep mathematical reasons for that experience and, ultimately, proved a theorem showing that if you sample signals at a regular frequency f , the maximum possible information transmission I (in bits) is given by the equation: $I \leq f/2$.

This is the Nyquist Theorem, and it implies that digital sampling can be perfect in certain ways. Shannon's proof of was named after Harry Nyquist, who had decades earlier proved part of it. In particular, the Nyquist Theorem revealed that a human ear, whose upper frequency seems to be capped at 20,000 Hz, could not possibly use signals sampled above 40,000 Hz. (This is why CDs and WAV files sample at 44 kHz, slightly beyond that limit). The fact that individual human ears can't hear above that range — while bats can — is the warrant behind the claim that a sample rate up to 88kHz is useless unless you're a bat.

A Triumph of Technology

Back in the 1970s, as a teenager, I fancied myself a nerdy audiophile. Half my spending money went to flat-frequency-response headphones and fancy amplifiers with high linearity, low noise and wide dynamic range. Even the earliest CDs brought amazing improvements in all those measures. So I tossed my record collection. I knew, based on our human 20,000 Hz maximum, that digital recording was objectively better, if not truly perfect.

Actually, it's not quite that simple. Headphones and earbuds, being smaller and quieter, do indeed give better sound per dollar, but they move with your head and remove the bass notes from your skin. Digital CDs — being digitized but not otherwise compressed — still sounded nearly perfect to me, but not so MP3s, AACs and, later, streaming, all of which made audio more portable and convenient but at some cost. It is very clear to both my senses and my intellect that too much compression really does damage sound quality — no one disagrees with that part.

So, I grew up experiencing two technology transitions: from analog LPs and phones to digital CDs and voice-over-internet (VoIP), which sounded fine, then from those to highly-compressed MP3s and cellphones, which definitely sounded worse.

Yet even the bad-sounding conveniences won in the marketplace, so much so that now most of the music and sound we consume is both digitized and compressed. Clearly, having 5,000 songs in your pocket, as the first iPod advertised, is collectively

more attractive than having a few really good songs at home on a turntable. The triumph of science was to invent and deploy all those technologies, so people could choose between fast/convenient/coarse (like iPods) vs. slow/inconvenient/refined (like CDs).

The Paradox of Musicians vs. Engineers

The vast majority of low-quality sound, whether on iPods or cellphones, comes from over-compression. But the age-old debate revived by Neil Young isn't about over-compressed music like streaming, but about uncompressed digital music like CDs. That's the mystery: Why do engineers think it is perfect while musicians think it is awful?

That's a truly scientific paradox, involving scientists themselves. Why do Young and many, many musicians and sound artists insist, in their full professional capacity, that even the best digital recordings damage presence and emotional content? This standoff pits sensitive listeners who can't explain their perceptions scientifically against actual scientists, who, in spite of insensitive ears, claim those perceptions are physically and mathematically impossible. What if the scientists missed something?

The Mystery of Missing Microtime

My own scientific career might shed some light on this. As a kid, I actually enjoyed solving the infamous physics "word problems" that intimidated most of my peers. I cranked tirelessly not just on my high school's "problem of the month," but on constructing (analog) electronic circuits like TV jammers, blinking tie-clips, librarian tormentors and wireless bugs. (My only digital project was a clock radio: I etched the copper circuit board, soldered in the transistors and integrated circuit, and mounted it all in a wood cigar box.)

My do-it-myself attitude lasted through graduate school, where I stumbled onto a deep mystery about neurons in the brain. It turns out that neurons care about time-differences a thousand times tinier and faster than neuroscientists thought.

The reason is simple: Information content goes up with timing precision. If brains are to process information efficiently, they must be sensitive at least to milliseconds, if not to microseconds. To discover on my own that neuroscience had missed the mystery, microtime reminded me both of nature's delicate machinery and of the

narrowmindedness hobbling publish-or-perish research. I saw my own field of science miss something huge.

After graduate school and a post-doctoral fellowship, I returned from academia to my boyhood stomping ground in Silicon Valley. There, I spent 15 years working as a “tech turnaround artist,” meaning I either saved hopeless technology projects or at least understood why they couldn’t be saved. In facing challenges that other strong professionals had not resolved, I re-learned the principles of solving very, very hard problems — principles I first learned from those physics word-problems in high school. The most basic principle is common to puzzles of all types: What clues haven’t we used yet? In particular, is there another law of nature we forgot that might apply?

“Where” Beats “What” in Hearing

At the beginning of this article, I mentioned that sound recording involved two laws of nature, with the Nyquist Theorem being one of them. The other is the speed of sound in air: speed of sound = 330 meters/second = 1 mm in 3 microseconds.

Why might this matter? The science of sound recording got in its present mess by sticking with what it was good at, rather than what it needed to know. Science was good at dealing with one variable at a time, a single monaural sound channel, via needles, wires, speakers and, ultimately, Shannon and Nyquist. At that time, science didn’t know what brains do, so it skipped that part.

A more brain-centered, human-centered approach to sound³ recognizes that the main task of a brain is to manage the vibrations of a physical, three-dimensional body. Part of that task involves making sense of vibrations outside the body, both in recognizing what thing made a sound and, more especially, inferring where the sound came from.

Here’s why: Imagine you’re alone and frightened in the woods, in the dark, with threats nearby. Suddenly, crack! A twig snaps close by. At that moment, which would matter more to you: where the sound came from or what type of wood the twig was made of?

The best way to locate sounds is to use the whole body — ears, skull, skin, even guts — since the entire body contains vibration sensors. The brain’s main job is making sense

³ See NeuroSafe brochure *The Science of Sound Healing*

of vibrations throughout the body, eyeballs to toes to eardrums, all consistent, all at once. One single vibratory image unified from skin and ears.

Headphones and earbuds fracture that unified sensory experience. Normally, your skin still absorbs vibrations from the outside, consistent with what you see. But with headphones covering them, your ears process entirely different signals injected directly into the perceptual space inside the head. That new sound image bypasses skin and eyes, while still being superimposed in front of you in space, on top of real sound sources. That physical impossibility sounds interesting, but it is the deepest kind of hack a brain can suffer, short of drugs. Consuming separate, inconsistent sensory streams that create competing maps of space violates a brain's design.

For localizing sound, though, the skin is merely helpful, because the ears do pretty well on their own. Your ears (and brain) deduce the location of a sound by using the microsecond difference in arrival times of soundwaves at two ears. (Acoustic scientists call this discrepancy the inter-aural timing difference. Using the sound-speed equation above, they might enjoy the "word-problem" of calculating how a centimeter-level separation you can hear translates into the microtime difference your ears make use of).

I wrote my PhD thesis on neural pulses, so I know they last about a millisecond each. That's a thousand microseconds, hundreds of times longer than the few microseconds we need for sound location. On the face of it, neural pulses seem too clunky and bulky to carry such tiny, delicate messages.

Fortunately for us, neurons do not encode sound like CDs do. Digital recording samples the sonic world every 23 microseconds, whether or not a sound is there. On the other hand, each neuron waits for a soundwave to arrive and then fires a single pulse. And all the neurons do that. So a soundwave striking lots of neurons at once will fire lots of pulses at exactly the same time. That synchronous volley arrives at the brain, which infers from the volley the time the original soundwave arrived, if necessary down to the microsecond level.

How does all this translate into the language of technology? The guiding principle of a nervous system is to record only a single bit of amplitude at the exact time of arrival. Since amplitudes are fixed, all the information is in the timing.

On the other hand, the guiding principle of digitization is to record variable amplitudes at fixed times. For example, sampling with 24-bit amplitude resolution,

every 23 microseconds (44 kHz). Since sample times are fixed, all the information is in the amplitude.

So unlike digital recorders, nervous systems care a lot about microtime, both in how they detect signals and how they interpret them. And the numbers really matter: Even the best CDs can only resolve time down to 23 microseconds, while our nervous systems need at least 10 times better resolution, in the neighborhood of two to three microseconds. In crass amplitude terms, that missing microtime resolution seems like “only” tiny percentage points. However, it carries a whopping 90% of the resolution information the nervous system cares about. We need that microtime to hear the presence and depth of sounds outside us and to sense others’ emotions inside us.

The old analog technologies, LPs and POTs phones, preserve that necessary 90%. Digitization destroys it. Neil Young was right.

Easy Tests

This being science, all of it is testable. The most dramatic measurements of microtime precision will be found when testing people accustomed to using sound to locate things — that is, in the congenitally blind, especially those who live in quiet, natural places, where tiny distinctions are easier to hear. In general, the hearing of those people will likely represent the most accurate human auditory perception possible.

So, we should test how well they hear locations in space, in ideal circumstances — say, tracking airplanes flying overhead from a quiet, open field (with skin exposed), or distinguishing clicks a few centimeters apart at various distances. Let them compare LPs against CDs and streaming media. Let them compare POTS phones against VoIP and cellphone calls. Ask them yourself. (And don’t trust Apple or Google to help).

The Future of Microtime Communication

In calling out digital music, Neil Young unknowingly highlighted a global public-health catastrophe caused by artificial sound in general. (And by screens too⁴.)

⁴ *Neural Computation* 29, 2293–2351 (2017) © 2017 Massachusetts Institute of Technology
doi:10.1162/NECO_a_00988

Of course, people like artificial or artificially-transported sounds. But as with drugs, our nervous systems sometimes like things that are bad for them. Enhanced, interesting, distracting sounds are no exception. And as with sugar, the market makes money giving people what they want right now.

Likewise, people as a species want to talk, or at least used to want to talk, back when real-life proximity and POTS phones let us do it properly. Unfortunately now, due to straightforward but perverse incentives, networks make money by throttling bandwidth through compression, so calls via cellphones are typically low-quality and shorn of emotional resonance. Humans connect poorly through cellphones unless they know each other really well.

The Coming Microtime Technologies

I predict the emergence of three new technologies that could change the world by reconnecting people.

1) Devices that quantify sound the right way. It shouldn't be hard to create a multi-function "tricorder." It could measure someone's sonic environment in all kinds of ways: decibels (min, max, median, average), frequency distribution, suddenness, repetition and any other signal parameters that matter to ears and brains. Better yet, when paired via a data channel with a matching tricorder on the other end of a phone line, it could track sensory metrics of the call itself, such as latency, latency jitter, hotspots and dead spots in pattern-space and (with stereo) 3D reconstruction resolution. This device would provide sensory-nutrition information, akin to the nutrition labels on foods, enabling healthy decisions.

2) Microtime recording and stereo. A video technology called an "event camera" already exists⁵, which uses pulses much like the nervous system does. Audio pulse-tracking could underlie a whole new form of analog recording, tossing amplitude and keeping microtime instead. When that recording scheme is used for stereo, played back through well-placed speakers, listeners will experience the sharpest, fullest 3D sonic field possible short of real live sound.

⁵ Mueggler et al. *The International Journal of Robotics Research* 2017, Vol. 36(2) 142–149

3) Micropresence = microtime telepresence. Imagine marrying microtime stereo with remote-video “telepresence” for the best interpersonal connection possible over distances. One very good arrangement would be an augmented-reality system (connecting matched rooms) that superimposes your conversation partner’s face consistently and coherently atop your own visual space. Microtime visual cues like micro-expressions will be partly visible even on a 3D face scanned by normal video. When combined with microtime sound properly aligned with the speaker’s mouth and throat, you will experience the most coherent sensori-motor experience possible remotely.

The sooner technology restores the microtime connections that humans need to thrive, the sooner we will thrive again, leaving loneliness behind for good.

The best connection will always be a physical presence and proximity. I expect more “acoustic” music concerts, all-live musicians, no microphones or even hyper-flickering LED illumination. Acoustic dances. Acoustic conferences. It turns out the so-called “emotional resonance” people enjoy together really is a kind of neuromechanical resonance, aided by acoustics and reduced by reproduction. (It’s best experienced in sacred spaces like churches, temples and Auroville’s famous meditation space called the *Matrimandir*. Live silence, like live music, will always connect people the way Neil Young hopes.