Audio Engineering Society

Convention Paper 10438

Presented at the 149th Convention
Online, 2020 October 27-30

Audio in the New Millinneum - Redux

James Moorer[1]

[1] *JAMMINPOWER.COM, 1111 Alligator Drive, Panacea, FL 32346*

Correspondence should be addressed to Author (jamminpower@earthlink.net)

ABSTRACT

In 2000, I was invited to present the Heyser lecture, entitled “Audio in the New Millennium” [1]. As it was the turn of the Millennium, I chose to review the progress of digital audio over the preceding 20 years of my own career, then extrapolate forward to 2020. In the process of doing so, one of the trends identified in my lecture stood out as being of fundamental importance. This was that innovations in all the aspects of modern computer technology – computation, networking, disc storage – had come to be driven almost entirely by the entertainment industry – digital audio, digital video, and computer games. The situation of 1980 when audio and video fed on the scraps of technology developed for military and industrial uses has now been reversed: military, industrial, and scientific computing is now driven entirely by the needs for entertainment technology. This paper will review the prognostications of the 2000 lecture and take stock of the progress, noting specifically which predictions hit the mark (a few of them) and which did not (most of them). More strikingly, a number of trends and developments in entertainment that were entirely unanticipated in 2000 have moved to center stage as drivers of technology.

1 Introduction

Picture the world of audio around 1980. There were no cell phones. There was no MP3, no CDs, no DVDs. Computers that people could afford were pathetic things with very limited audio, if it had audio at all. Telephone answering machines used analog cassette tapes to record messages. The Sony Walkman portable cassette player was the new hit product in 1979 and was the first portable “on-demand” audio device. Keyboard instruments that were small enough to carry with you were electro-acoustic devices, like the Fender Rhodes, which had a distinctive if inflexible sound. There were electronic music synthesizers, but they were large, expensive, and hard to use. There were professional-quality tape recorders, but they were similarly unavailable to most folks.

Figure 1**:** The world of audio in 1980 had no personal computers, no CDs, and no cell phones.

Technological innovation in that era was largely driven by the military and industry. Those of us working on digital audio had to use A/D and D/A converters that were designed to measure, say, air pressure in a nozzle, or sonar-return signals. We used computers that were designed to do spread-sheets or inventory control for large corporations. The limitations of these devices quickly became evident to everyone working in the field. For instance, the speed of the arithmetic on mainframe computers was severely limited with operations taking many clock cycles. Recall also that most computers were large, bulky and expensive devices.

Figure 2**.** Computers were huge expensive devices that were for industry and military applications.

Around 1972, I built an A/D and D/A system for audio for the group at Stanford that went on to form CCRMA, the Computer Center for Research in Music and Acoustics [2]. The best A/D converter I could find was the 14-bit Analogics MP2914A successive-approximation converter. It was designed for instrumentation. The specifications for the device were not in terms of frequency response, harmonic distortion or other audio-related criteria but instead were in terms of how long it took to settle to a voltage corresponding to the number presented to the device.

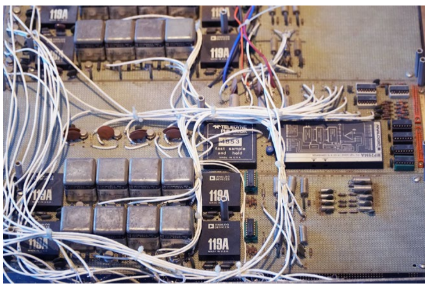


Figure 3**.** The converter system I built at Stanford used modules that were intended entirely for instrumentation rather than for audio.

There were computer games in 1980 – the Atari 2600 came out in 1976. Custom-designed arcade games were available somewhat earlier. All of these featured what you might term “primitive” graphics engines.

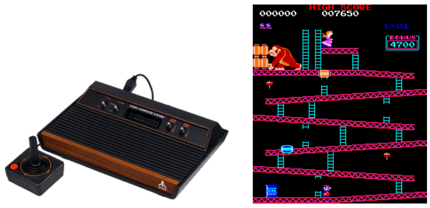


Figure 4. Atari 2600 released in 1976 featured primitive but serviceable graphics.

2 1984 – Miracles Happen

I suppose that actually the miracles started in 1983 with the release of “Return of the Jedi” and the THX Logo Theme [3], but let us concentrate here on a few pieces of transformative technology that all arrived in 1984 – the CD, the Macintosh, and the Yamaha DX-7.

Figure 4. 1984 featured the introduction of the CD, the Macintosh, and the Yamaha DX-7, marking some of the first technology with digital audio and audio quality as a design goal.

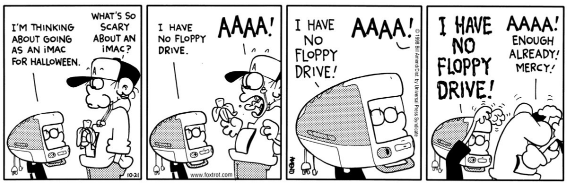
The CD players featured the first D/A converters *specifically designed to audio specifications*, rather than for instrumentation. The same is true of the DX-7, although I feel the real contribution of this synthesizer is that it is probably the first fully-portable, affordable keyboard musical instrument with a wide range of sounds. It could do a passable imitation of a Fender Rhodes or a Hammond B3 organ, as well as brass, strings, piano, and completely synthetic sounds. For the first time, you could fit an instrument in the back of a Volkswagen that was capable of providing the sounds the music makers of the 1980's craved.

I add the Macintosh to this list, not because it had an audio-quality D/A, but more because it was the first computer with the modern window-based interface that had sound integrated into the operating system at a very fundamental level. Sounds accompanied every aspect of Macintosh operation and have become so common today we don't even notice them. For example, when you take a picture using your cell phone, nobody blinks an eye when it makes the sound of an old Kodak camera shutter. In fact, it is weird when you turn the sound off. The Macintosh quickly became the platform of choice for artists and musicians, and to a significant extent, its descendants remain so today.

It is a bit off-topic, but I should mention that 1984 also saw the debut of the laser printer. It may be a bit of a stretch to describe the laser printer as part of the digital entertainment revolution, but it surely did play a significant part in the more general digital/computer revolution of the 1980s.

3 What Starts in Music Doesn't Stay In Music – or Entertainment

It wasn't long before this technology was co-opted by the computer industry. By 1987, the CD-ROM reader was introduced, and soon, all computers were being delivered with them. The number of CD-ROM readers quickly outnumbered the number of audio CD players in the world. The CD recorder made its debut around 1990 [4] as a strictly professional device for "proofing" CD releases. It was not until somewhat later that CD-R devices made it into personal computers. This led ultimately to the decline of the venerable floppy disk drive in the late 1990s.



Another example of a technology that didn't stay in the entertainment industry might include the 8mm video tape. It led to the 8mm digital video tape that was also used for multi-track (up to 8) digital audio recording. Digital 8mm recording soon became a standard for computer hard drive backup.

4 Graphics Processors

The arcade game computers in the 1970s required extensive special-purpose hardware to synthesize the images *in real time* needed for effective game play. The 1980's brought the first round of *graphics coprocessor chips* which could be placed directly on the computer motherboard. Curiously enough, these early coprocessors were not designed for gaming but more for vector graphics uses, such as CAD systems. This perhaps explains their somewhat limited adoption in personal computers of the day.

Ultimately the drive for better graphics came from the entertainment industry. The MMX instruction set was introduced in 1997 *directly into the CPU chip* specifically for accelerating computer games. It had little use in word processing and no use whatsoever for spread sheets, payroll, or inventory control. By the 1990s, it was clear that the computer industry had become something much more than industrial support machinery. It was now a delivery medium for entertainment.

The rise of the modern Graphical Processing Unit (GPU) began in the 1990's but it is descended from a decades-long progression of graphics processors of varying architectures. It took a number of technological advances to make the GPU practical – namely the availability of large capacity digital RAM for high speed frame buffering, and the development of special-purpose chips for video rendering. These developments brought it from the professional domain (flight simulators, medical imaging, *etc.*) ultimately to the consumer. The video game industry became the main economic driver of ever-cheaper and more powerful video rendering engine in the form of GPUs.

Note that processing for video has continued to follow these two tracks simultaneously – chip manufacturers put more and more numerical compute power directly in general-purpose CPU chips, and GPU manufacturers put more and more power into GPUs as well. Both industries seem quite healthy today.

5 Don't Forget Cell Phones are Audio

Although the cell phone was first shown in 1973, the digital cellular networks didn't proliferate until the 1990s. Among other things, it awaited the development of the CELP encoding in 1985 [5]. Today, the cell phone is the most widely used piece of audio technology. It has expanded beyond "simple" communication to the role of personal assistant. Modern "smart" phones also have graphical processing units (GPUs) mostly for entertainment purposes such as gaming and video streaming. There are thought to be something like 13 billion cell phones in the world, which is curious since it is almost double the number of people in the world (!).

6 Audio in the New Millennium

This brings us to the year 2000 and the presentation that kicked off this entire inquiry. The core of the original presentation were four charts. Each chart showed the rise of capacity that I personally experienced over my career in digital audio. I chose, somewhat arbitrarily, the four metrics of CPU power, hard drive capacity, LAN speeds and WAN speeds. I measure CPU speed by the number of million-point FFT-based convolutions that can be performed in real time. This is a simple 1-number characterization that does involve both raw processor arithmetic speed but also memory access speeds. The FFT is, in some ways, a "worst-case" test for caching schemes because of the difficult pattern of memory references. For hard drive capacity, I made two separate graphs – one of simple capacity, and another of capacity over physical volume – that is, hard-drive density. I wanted to track LAN speeds and WAN speeds because these are enabling technologies for studio (local) digital audio communication, patching, and transfer and for regional (non-local) transfer. For WAN speeds, I anticipated a quantum leap in delivery speed due to the adoption of fiber into the home.

7 CPU Speed

Figure 5 shows my original chart for the rise in CPU speeds with an annotation showing both the prediction and the achieved rates indicated in parentheses.

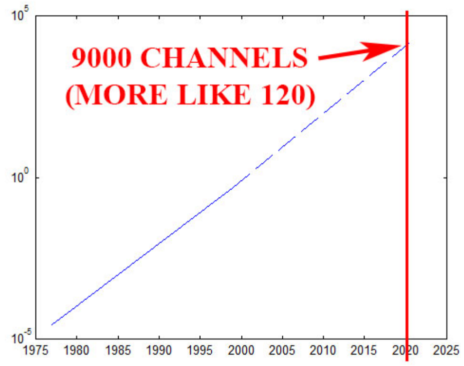


Figure 5: Rise in CPU speed in terms of number of channels of FFT-based convolution in real time. Values past year 2000 are linear extrapolations predicting 9000 channels by 2020. Modern CPUs achieve about 120 channels but that is using only a single core.

A linear extrapolation from the year 2000 suggested that we would be processing 9000 channels of concert-hall reverberation in real time by 2020. It didn't happen. A modern Core-i7 chip achieves about 120 channels in real time. In all fairness, this is using just a single core. The computer on my desktop has 12 cores. With hyperthreading, it runs about 18 times faster than a single core which gets us up towards maybe 2000 channels. This calculation does involve a great pile of assumptions that I won't detail here but have to do with Amdahl's law. I haven't tried the experiment, but the actual rate will probably be somewhat less than 2000 channels.

8 Array Processing Again

Note that the paper in 2000 completely missed the significance of the rise of the GPU and did not really anticipate the modern SSE2 and SSE3 array-processing instruction sets [5]. *The observed rate of 120 channels of real-time convolution does not include any array-processing advances – it is entirely the result of clock speed increases and superscalar architecture*. A speedup of about a factor of 3 can be attained (per core) by making careful use of the array processing architectural and algorithmic advances (SSE3 and the FFTW algorithm [7]).

Note that the SSE3 array processor extensions are so powerful that many computer games with very elaborate graphics can run entirely within the processor chip without GPU help. *There is little motivation for Intel and AMD go to the trouble and expense of adding all this numerical processing horsepower to general-purpose computer chips except for entertainment – streaming video and video rendering for computer games.* For the most elaborate photorealistic real-time rendering, the GPU still provides the best results.

As the available compute power of GPUs has increased, there has been increasing interest in using this power for general-purpose computing. Although the architecture of the GPU is not designed for general computing problems, it can give dramatic speedup of many numerical tasks, such as matrix calculations, spectrum analysis and neural net computations. Here is another case where a device conceived *entirely* for computer graphics and driving by the entertainment industry is finding increasing use outside of that industry. Virtually every computer today has a GPU coprocessor, from cell phones to tablets to laptops to desktops. Note that this is *in addition* to the on-chip array processing instruction sets.

Just for reference, I measured the speed of a million-point convolution on the GPU. The one I used was the NVIDIA Titan Black, which does have quite fast double-precision. It was about 100 times faster than the single core (unaccelerated) rate using a modern CPU. When compared to a CPU implementation accelerated by SSE3 and FFTW, that goes down to a bit over 30 times as fast. Mind you, that is still something like 12,000 channels of concert-hall reverberation in real time. This is a theoretical number – it is not clear we could actually achieve that many channels of audio in and out of the GPU memory that fast (over 1G bytes/second), or where that much data would come from. The point of calculating this number is just to give a hint about the amount of compute power available on a GPU.

Let me repeat that the GPU exists almost entirely to accelerate video rendering for computer games – that is, for entertainment purposes. Everything else (neural nets, statistical and matrix calculations, *etc.*) is gravy. Industry now feeds on the droppings of entertainment technology.

9 The Hard Disc

Figure 6 shows the graph from the 2000 paper annotated with the capacity predicted by linear extrapolation of the 1980-2000 period out to 2020. The number in parentheses represents the capacity that is readily available today.

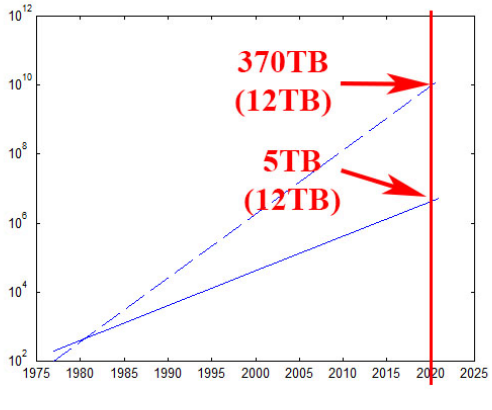


Figure 6 – Increases in capacity of hard drives. The 2020 paper had

> predicted a dual path for hard drives – one path of big, bulky drives, achieving maybe 370 TB in capacity by 2020 and the other path of ever smaller drives achieving maybe 5TB by 2020. No such bifurcation happened. 5.25” hard drives of 12TB capacity are readily available.

In 2000, I had noticed that the hard-drive manufacturers were taking two separate paths to hard-drive design – higher capacities were being achieved, but the drives were physically large, whereas there was another trend of ever-smaller drives like the 1” IBM Microdrive [8]. I made two separate predictions, one of sheer capacity irrespective of physical size and one normalized by size. This led to a prediction of large drives of something like 370 TB of data, and tiny (1”) drives of maybe 5TB. Instead, the industry standardized on the 5.25” format (of varying thicknesses). Today, drives of 12 TB are readily available. The capacity of the largest drives did not increase as much as predicted, and drives stopped getting smaller once the 5.25” format was adopted.

I should mention, however, that one thing I did not anticipate was the remarkable rise in transfer speed, mostly due to the adoption of magnetoresistive head technology [9] that became commercially available around 2004. This technology gave a huge boost to the hard drive industry. It seemed like overnight, we went from the old 9 Mbits/s drives to drives capable of up to 500 Mbits/s. For a while, it took relatively exotic computer interfaces (like 800 Mb/s P-1394 “firewire”), but the now-ubiquitous SATA and USB-3 interfaces are inexpensive and achieve quite high transfer rates.

As a personal note, I look back on the endless hours and sleepless nights we spent (before the year 2000) trying to get reliable random-access hard-drive recording and playback of 8 channels of audio with a single drive. The drives of today easily achieve seamless transfer of hundreds of channels of audio.

10 LAN Speeds

In 2000, the 100 Mb/s ethernet was widely available. The state of the art was probably represented by the 1 Gb/s fiber channel, but that was limited to professional or industrial installations. Still, 100 Mb/s is good for maybe 80 channels of reliable audio communication. Figure 7 is the chart from the 2000 paper, annotated with today’s results:

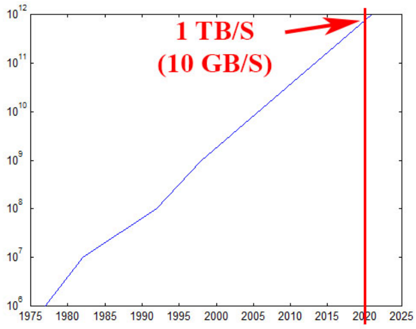


Figure 7 – LAN speeds. Extrapolating from the 100 Mb/s rate of 2000 gave a prediction of 1 Tb/s by 2020. Today the 10 Gb/s ethernet is widely available.

Note that 80 channels of audio is good for “small” studios with less than maybe a dozen workstations but is inadequate for larger studios with multiple rooms. It is tempting to think of designing a large studio with a single SAN (Storage Area Network) holding all the audio, but 100 Mb/s isn’t enough to support such an architecture.

Fiber has continued to progress in speed, but only by continuing to increase the complexity, such as by the use of wavelength-domain multiplexing. For general use, ethernet through UTP (unshielded twisted-pair, such as the standard CAT6) has become the standard. The ethernet standard is defined up to about 400 Gb/s, but the commercial implementations top out at 10 Gb/s. This speed makes centralized storage theoretically possible but generally system designers still choose local storage for recording and mixing with large online storage just for backup and moving data from one task to the next (*i.e.,* from recording and editing to final mix).

In 2000, I did not anticipate the rise of wireless technology, which is ubiquitous today. It is still not as fast as UTP connections, but many installations achieve more than 100 Mb/s – about where UTP connections were in 2000.

11 WAN Speeds

In 2000, the fastest widely-available data link into the home or office was by DSL, which topped out at maybe 300Kb/s. That sounds downright quaint today, although at the time, it seemed astonishing. Figure 8 is the chart from the 2000 paper, annotated with 2020 results.

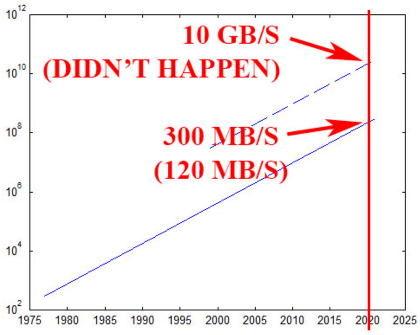


Figure 8 – WAN (download) speeds. In 2000, the generally-available state of the art was the 300Kb/s DSL line. Extrapolating to 2020 predicted a rate of 300 Mb/s. Rates of 120 Mb/s are available in most locations. The quantum leap to fiber-optics in the home did not happen.

Today, the cable modem to my house achieves about a 120 Mb/s download rate (16 Mb/s upload). Note that this (theoretically) is capable of transferring over 100 channels of audio in real time.

In 2000, I was anticipating a "quantum leap" to fiber directly to the home. It didn't happen. Bummer. It still might at some point in the future. Despite this, we have easily enough bandwidth to upload and download gigabits of data, audio or otherwise.

As a side note, I might point out that the ratio of about 10:1 between LAN and WAN speeds has held pretty constant over the last 4 decades with only momentary deviations. WANs were a bit slow around the year 2000 but have picked up speed since.

12 Streaming Media Happened

I wrote one of the first drivers for the Arpanet in the early 1970s. We did transfer audio files around (mostly speech), and some experiments were done in real-time speech communication. The 50Kb/s data rate of the original Arpanet was so slow that it was hard to look ahead to a time when that would not be a limitation. Today's streaming media requires the confluence of three technological improvements: (1) inexpensive compute power (2) massive increases in inexpensive disk storage capacity, and (3) reliable data rates into the home of greater than 10 Mb/s. This number is somewhat arbitrary – it is the data rate of video from a DVD, which was considered to be "high quality" for quite a long time. There have been several improvements in video coding since the MPEG-2 used on the DVD.

In 2000, I did predict that media, both audio and video, would be delivered by IP, although I wasn't clear on the exact form it would take. I did not anticipate, for instance, streaming through a cell phone. I also had a hard time figuring how the home computer would fit into the picture. I did not anticipate that cable boxes would receive not only the cable but would also be connected to the internet[1]. Note that cable TV, though digital, is still not IP. It remains a special-purpose multiplexed format. I was expecting the cable in that form to

[1] This is doubly ironic – at home, my internet is delivered by the cable, but my TIVO doesn't implement the DOCSIS 3.2 protocol necessary for extracting the internet connection directly from the cable it receives. Thus my need for a cable splitter, separate cable modem and CAT6 cable to the TIVO.

disappear and replaced entirely by video-over-IP. Although this is an important trend, the TV cable (and corresponding “cable guy”) persists, although many households rely entirely on internet for their entertainment delivery.

I also had no idea that the internet would be distributed through what are essentially the wireless telephone exchanges to cell phones. In 2000, there didn’t seem to be much interest from the cell phone companies in delivering internet access through their networks. That changed soon after 2000. This is also interesting because the transport protocols are very, very different. The telephone networks use ATM. It took a while for “IP over ATM” or even “TCP over ATM” to get fast and reliable. I find it interesting that cell phones have three different radio links – cell-phone tower radio links, wireless internet, and Bluetooth – each with its own RF transceiver and antenna, all in the same hand-held package. The engineering challenges to make this work are significant.

I anticipated that internet radios would be ubiquitous as well, and would be combined with FM radio receivers. That one didn’t happen either. I do, however, enjoy listening to internet radio and have been delighted with the international nature of the available broadcasts.

The RIAA has a nice chart [10] showing each technology (LP, Cassette, CD, various forms of download/streaming and more), when it came into use, and when it declined. Note that according to Variety Magazine, video streaming accounts for a bit over 60% of the total internet bandwidth [11]. Think for a moment about what this is telling us. The internet began as a strictly military way to maintain communications in a fallible environment. When I was working on it in the mid 1970’s, I considered it just a way for those of us in the big research facilities to send email and bits of software to each other. Even Tim Berners-Lee [12] developed the world-wide web as a means for academic researchers to share technical documents and information. None of us anticipated the huge commercial adoption of the network, which has now become our “second brain” and trusted assistant. The fact that 60% of the total bandwidth of the internet has been taken over by entertainment is more than significant – it is stunning. It tells us something important about ourselves.

In 2000, I also thought that the increase in network bandwidth would have meant the end to data compression in audio. It didn’t happen. I think this is at least partly the acceptance that AAC or MP3-encoded audio (or AC3 for 5.1 surround audio) is “good enough”. There are still artifacts, but one seldom hears them without direct side-by-side comparison with uncompressed audio.

13 The Streaming Revolution

The streaming revolution started with music downloads – Napster in 1999 and iTunes in 2001. This led to the rise of the MP3 player, such as the iPod and others. Napster was put out of business in 2001 for not securing the rights to the music – an error that Steve Jobs avoided with iTunes. At the time, I personally found it astonishing that Jobs had managed to make deals with the major record labels. In 2003, the FairPlay DRM system that helped protect the artists and labels came into commercial use. Today, most people do not bother to build libraries of audio files but instead subscribe to “jukebox” services like Pandora. These have the advantage that you can call up any song in its massive database quickly. Even if you had enough memory to hold all the audio tracks that exist, you probably wouldn’t want to. Most folks would prefer to use that storage for things like photos and videos from their friends and family – items that you can’t get from any download service.

Video streaming was adopted perhaps a bit more slowly than audio, presumably due to the technological limitations noted before. Decoding the video takes compute power, and it took a while for the processors to get powerful enough to decode high-quality video in real time. As noted before, it also requires a relatively-high bandwidth connection into the home. That notwithstanding, streaming video displaced the DVD for commercial delivery of video to the extent that the last of the major big-box DVD chains like Blockbuster went out of business in the years around 2015 or so (again, see [10] for a more detailed timeline). The last rotating medium, the BLU-RAY discs, continue to be used today for high-capacity applications, such as high-definition feature films and computer games, but it is very much an isolated niche player.

It is interesting that, like streaming audio, most people no longer keep large libraries of digital video entertainment. As noted before, we reserve that capacity for videos from family and friends.

14 Hits and Misses (Mostly Misses)

Of the four dimensions of technological progress that I introduced in 2000, only two of the admittedly primitive predictions came anywhere near the mark – hard drive capacities did continue to expand, and download speed to the home, at least in or around big cities, did hit the 100+ MB/s prediction (close enough to the predicted 320 MB/s that I call it a win). The miss in compute power is at least partly because I don't take multi-core or GPUs into account. If you add either one of these to the equation, then overall compute speeds do get close to the predicted 9000 channels of concert-hall reverberation. Let me point out again that I am using concert-hall reverberation as a metric for compute power – not to suggest that there is a need for that many channels of reverberation calculations.

Probably one of my biggest personal disappointments is that little or no progress has been made on "intelligent assistants". I thought surely that, for instance, music mixes would become so complex that significant sub-mixes would have to be delegated to automatic AI-like processes. This didn't happen. From what I can tell about the industry, for the most point we have decided that we have enough capacity and flexibility to do what we want. That is, nobody seems to want to do a 300-channel mixdown even though we can. Part of that is surely because of the limits of human hearing – there is no way the human ear can tease apart 300 separate voices. After all, it only took 30 voices in the original THX logo theme to make a texture thick enough to obscure most of the internal motion of the voices.

I recently completed a long-form computer music piece entitled "The Man in the Mangroves Counts to Sleep" [13]. At a several points in the piece, the number of active voices exceeds 100, reaching a maximum of about 150. I found it impossible to perform a mixdown of the entire piece all at once. I took to doing what is often done in the film industry, which is to make a number of "sub-mixes", each one of which was no more than maybe 16 or 32 channels. The final mix then was balancing 5-10 submixes. That is, instead of tackling the mixdown problem all at once, I used a version of "divide and conquer" to break the problem down into manageable pieces. This makes things more complicated. When something goes wrong at the last moment, you do have to dig around to figure out which submix has the mistake, correct the mistake, rebuild the submix, and go on. It does add some steps to the process, but it allows a small mixing team (*e.g.* one person) to do a very large mix, though with some difficulty.

I was also a bit disappointed that "running transforms" haven't become more popular despite their advantages, such as running constant-Q, constant-bandwidth, and Bark scale transforms at the same time, all allowing perfect reconstruction. To help encourage their adoption, I wrote a summary of what I know about running transforms that is available on the web [14]. Maybe this will help spark some applications.

I had advocated the use of higher sampling rates because they provide spectral "head room" that allows us to use more nonlinear processing. High sampling rates are readily available today, but they are surrounded with some amount of controversy, and the motivating nonlinear processing algorithms have largely failed to arrive.

15 The Audio Beam-Forming Fizzle

In the 2000 paper, I waxed poetic about the possibility of self-locating microphones and speakers, possibly using IOT (Internet of Things) and GPS technology. I did some amount of development along this line that I summarized in a paper [17]. It didn't happen. Or, at least, it didn't happen the way I was hoping for. There are many examples of multi-speaker installations, and the 64-channel Dolby ATMOS theater sound system is probably the most prominent example of this. My biggest disappointment is that we aren't using beam-forming technology in any meaningful way. This is perhaps because it depends on exact and precise control of phasing of each transducer that is hard to obtain in practice, combined with the possibly limited payoff. I had envisioned a physically large microphone array where we would record hundreds of channels, but then, using beam-forming, would be able to "zoom in" on individual sound sources, such as musicians in an orchestra *after the recording is done*. It would appear that it is way too easy to just set up microphones in front of each of the soloists than to worry about some gargantuan and twitchy microphone array. All this notwithstanding, I persist in visions of "transducer wallpaper" that would continuously calibrate itself using super- or sub-sonics and would thus be capable of forming a precise audio image at a spot. The mathematics for beamforming is well known and is mostly just awaiting a convincing application. Some limited commercial applications in the home have been introduced. Some of the modern "sound bars" use limited kinds of beam-forming to "throw" the image to one side or the other. Perhaps another decade will be enough for this technology to become common. Certainly, we can expect the "weak" form of this – large numbers of speakers for home theater systems – to become more available.

16 The Explosion and Fragmentation of Media Distribution

I had thought in 2000 that media distribution and consumption would get concentrated into a very small number of devices. A case can be made that this has indeed happened, but we have to admit that the smart phone has taken the lion's share of the roles. Today's call phone is perfectly capable of streaming audio, video, and computer games. Somewhat more elaborate versions of all these can be found in game consoles, digital cable boxes with recorders, internet radios, and other fixed-location devices. The death of cable TV is predicted every year but cable still seems to hang on. In 2000, I was complaining that the last classical FM radio station in my area had ceased broadcasting. Today, several streaming audio sources can provide endless hours of just about any type of music desired – classical, rock, K-Pop, Kwela or whatever. Similarly, I have been delighted that more and more vintage material is available for on-demand streaming so that we have the classics available anytime and pretty much anywhere. Note that we can also get international broadcasts when desired.

Probably the most disturbing trend in modern media is not technological but more a matter of rights, and that is the fragmentation of distribution – not of distribution technology, but of distribution services. In 2000, most people had cable television from a single local provider into the home with relatively few add-on services, such as HBO, that required additional payment. Today, there is no one portal that can deliver *all* the entertainment that is produced today. We end up subscribing to a number of different sources for our entertainment needs. One might think that these services would specialize, so that people in general would only subscribe to a few of them. This does not seem to be the case. There are now many channels with overlapping audiences. The result is that most people can not afford to see absolutely everything they might like to see – some compromises must be made. I can only hope that there is some consolidation in the industry in the future so that we can get, say, 80% of our desired media consumption through a single moderately-priced service.

17 What Happened and Why?

I have argued that all aspects of modern computer technology is now driven by the needs of audio and video – that is, by our entertainment needs. One of my colleagues quipped that there will only be two professions in the future: health care and entertainment. This may be a bit overdrawn, but it is clear that all aspects of technology – computer chips, display screens, hard drives, networks – are now being forced to advance by the people's need for entertainment. In some sense, I can say that we won! For those of us who spent decades trying to convince people that entertainment technology is actually a thing, it is no longer a hard sell. People hold in their hand every day an entertainment delivery vehicle. Billions (trillions?) of dollar every year go to making the devices and the content of the stream of entertainment we consume. Now we can ask the hard question – why? To hint at an answer, I have to leave the world of verifiable facts and delve into speculation, but here is my answer, such that it is:

They didn't look like us. They slept in trees on the savannahs of Africa. They talked. I imagine that one of the first phrases spoken might have been "tell me a story". And tell stories they did. They told stories of the great heroes that we still see in the skies – they told stories of where their families came from and of the great herd migrations. They told stories of floods and droughts. But most of all, they told stories about each other.

Story-telling is so basic to our existence that we have to remind ourselves how fundamental it is, and how different we are from the animals. Take our nearest relatives – some kind of ape/chimp primate – and ask this question: how often does a juvenile pick up, say, a colorful bug and show it to mom, as if to say "look what I found! What is it?". How often? Never. It is not our opposable thumbs or tool use that makes us human, it is our relentless drive for story-telling. For the most part, we don't pick nits out of each other's hair – we tell them stories. We say what happened today. We ask what happened to them. Sometimes people complain when I describe even the old "dumb" flip-phones as entertainment devices. My answer is that most of what we communicate over the phone has little to getting done what we need to get done to function in this world – it has more to do with telling our stories and soliciting stories from others.

Once you get this notion in your head, you start seeing everything as a form of a story. A commercial is a story. The editor has 900 frames to make a beginning, set the characters and the conflict, and wind it up to a satisfying ending. An enormous amount of the world's creativity goes to formulating compact stories in the hearts of commercials.

Some news professionals bristle at the thought that they are producing entertainment. But the news is exactly entertainment – it may be informative as well, but first and foremost, it is a form of story-telling. The best news stories have a form with a beginning, middle and end. They may not resolve the inherent conflicts, but the best news stories expose them in context to the fullest degree that captures and sustains your interest. There is some hint or expectation that the resolution, if any, will surely come in another news story at some other time, forming a full story-arc from beginning to finish. Some stories may wait hundreds of years for a full exposition, but wait they do – and we remain ready to hear them.

It is our relentless drive to tell our stories and to receive the stories of others that forces our technology forward. What is Facebook if not a place entirely dedicated to the exchange of our stories? Or Instagram? Or Twitter? It is that most fundamental of human activities that propels us forward. "Tell me a Story".

References

[1] J. Moorer, "Audio in the New Millennium" *Journal of the Audio Engineering Society*, Volume 48 Issue 5 pp. 490, 492, 494-498; (2000)

[2] Andrew J. Nelson, "The Sound of Innovation: Stanford and the Computer Music Revolution" MIT Press 2015, Cambridge, Massachusetts.

[3] https://www.thx.com/deepnote

[4] D. P. Cumming, J. Moorer, H. Ogawa "CD Mastering Using a Recordable 'Red Book Standard' CD and Graphical PQ Subcode" Preprint 2967 Audio Engineering Society 89th Convention 1990, Los Angeles

[5] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in *Proceedings of the IEEE International*

Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 10, pp. 937–940, 1985

[6] https://en.wikipedia.org/wiki/SSE3

[7] https://en.wikipedia.org/wiki/FFTW

[8] https://en.wikipedia.org/wiki/Microdrive

[9] https://en.wikipedia.org/wiki/Disk_read-and-write_head

[10] https://www.riaa.com/u-s-sales-database/

[11] https://variety.com/2019/digital/news/netflix-loses-title-top-downstream-bandwidth-application-1203330313/

[12] https://en.wikipedia.org/wiki/Tim_Berners-Lee

[13] https://www.jamminpower.org/man.html

[14] http://www.jamminpower.org/PDF/JAM/Running%20Transforms%202018.pdf

[15] http://www.jamminpower.org/PDF/Phased-Array%20Microphone.pdf