Digital Pioneer Tom Stockham

by Daniel Levitin

The recording and music industries would have been very different without Tom Stockham. Audio buffs and electrical engineers know Stockham as the man who introduced digital recording and tapeless editing to the world. Then there was his patent on the homomorphic compander, a device which can be thought of as the forerunner to DBX and Dolby noise reduction systems.

Stockham received his bachelor's, master's and doctoral degrees from M.I.T. He has received numerous awards for his contributions to audio technology: The Poniatoff Gold Medal from SMPTE, a Gold Medal from the AES, an Emmy, and many others. NARAS awarded him a Grammy in 1993 for his "pioneering role in the development and advancement of digital recording."

Currently a professor of electrical engineering at the University of Utah, Stockham also runs his own consulting firm, Stockham Technologies, the focus of which is applications in vision and genetics. Daniel Levitin recently caught up with Dr. Stockham to discuss the development of digital recording, his views on the digital/analog debate, and other audio matters.

DL: What was your role in the development of what we now know as digital audio?

TS: Let me go back a little ways. The first time I ever saw digital audio actually working was in 1962. I'd been a student at MIT in the fifties, then I went into the Air Force. When I came back to MIT, as a new assistant professor, I went around visiting the labs and I found peopledoing interesting work. Some of the students were using a computer called the TX-0, and they had connected a tape drive to the machine and a microphone and speaker to an A/D and D/A converter. What you could do was talk into the microphone and the tape would record all of that in binary. It was probably using 6 to 11 bits, with a sampling rate around 10,000 samples per second. After you had finished recording, the thing would rewind and play it back.

DL: We're talking about a machine that took up the better part of a large room, right?

TS: Right. Then the TX-2 was developed, which was an even larger computer. It had 65K memory with 36-bit words.

This was the first really large magnetic core memory computer. This was back in the days before the byte, with 7-track tape, and all of that.

DL: Where did the converters come from at that time?

TS: A company named EPSCO created an A/D-D/A converter you could buy. It was capable of 11 bits, I think, and was pretty high up there in sampling rate - maybe 22K. It's possible I saw some demonstrations of these in the late 50s, around 1959. Bernie Gordon, who was the CEO at EPSCO, was playing with digital audio back then.

DL: What did you do next?

TS: Well, we used the TX-2 and digital audio quite a bit. I worked on a project with Amar Bose, who was at MIT at the time, and we used it to design the first Bose speakers.

DL: Were the first Bose speakers built around the same principles as his later ones?

TS: Yes, it was his idea that the loudspeakers were

limited by the room acoustics, not so much what was in a well designed loudspeaker; that the problem was mostly outside, not inside the speaker. Of course, one could build a terrible loudspeaker to claim otherwise. But his idea was that if you built a loudspeaker properly, most of the distortion was put forth by the room.

DL: Why would room acoustics be so singularly important in a home listening environment? If I have musicians actually play in my living room, it sounds great. Why do room acoustics play such a critical role if I merely put speakers - point sources, essentially - in the room at the same locations where the live musicians were?

TS: Because your living room then becomes a second venue. If musicians are recorded in a concert hall, and you try to play back the sound of that hall in your living room; the second venue problem is created.

DL: So you've got one set of acoustics stacked on top of another; competing.

TS: Yes, that's right. That's why live performances sound so much better than anything you have ever heard. And the second venue problem hasn't been solved, even today.

DL: It seems like one solution would be what contemporary recording engineers have done, at least for popular music. If you record everything in a controlled studio environment, close miked, you would avoid this competition of acoustics...

TS: That is a partial solution to the second venue problem, but it's not a total solution.

DL: Because of the inherent artificiality of studio recordings?

TS: Suppose we let the listening venue be the one that we want. Let's try to change the original venue somehow such that when you combine a typical listening room venue with the changed original venue, the whole thing turns out more like a single venue when it's played back. As far as I know, nobody's ever shown that such trials works well, or even better. It's somewhat better, but it's not the solution we really want.

Bose found a living room that he felt was reasonable, and he put a loudspeaker in the corner. It was a prototype of the first one his company sold, the 2201. You may remember that it was an 1/8 of a sphere and fit in a corner. We set up microphone and recorded some music produced by the speaker in this "ideal" listening room.

The next step was to find the impulse response of the room. By setting off a spark in the corner where the speaker had been, and recording it over and over again, the spark permitted us to determine the noise free impulse response of the room. In this way, we had the impulse response of the ideal speaker as a reference.

Next we took the music we had used earlier and convolved it with the spark recordings we had made. The whole idea was to see how much poorer the loudspeaker was than the spark (the ideal speaker). We found that by using a filter, properly designed, the two were very close. It seemed to us all but a very few of the speaker problem had been removed. Now I'm going from memory here; if Bose was sitting here he might be yelling that's not the major point. As exciting as these results were, the point is we used digital audio for this experiment back in 1963-64.

DL: What was your next encounter with digital recording?

TS: TS: We knew how expensive digital audio was back then, and I just didn't pursue it much more for awhile. Starting in 1966 I spent 2-1/2 years at MIT's Lincoln Labs working on digital signal processing.

DL: What got you interested again?

TS: One day two very well known MIT professors came into my office because I had been asked to write a problem for the doctorate exam that year. They said, "we want to ask you a question about this problem you put in here on the exam. What is this *about*?" Now I had created a digital version of an RC circuit in the problem. I figured since digital signal processing was going to be an important part of the future, that the students were being taught all about this. And these two professors just said, "*Oh*. O.K." and they walked away. But this told me that the people who were really in-the-know didn't understand this; they weren't aware of what it meant.

DL: You left MIT in 1968 and joined the Computer Science faculty at the University of Utah. What prompted the formation of Soundstream in 1975?

TS: Malcolm Low (the L in KLH) was here at Utah and helped set up Evans and Sutherland Computer Corporation; they pioneered 3-D computer graphics. One day in 1974, Malcolm came over to my house and said, "you know it's time to start a digital audio company." I told him he was crazy, but one thing led to another, and we were in business a year later.

Our purpose in starting Soundstream was to develop a system for the home that would play back digital recordings. We knew there was a chicken and egg problem, so we started out by creating the professional equipment that would be needed to create the recordings themselves.

We developed some 16 bit A/D-D/A systems and we put together a machine that could record and playback. It had an instrumentation tape recorder - this is a recorder that's used for technical experiments, and such - and a large box of electronics that went with it.

DL: And storage was all on magnetic tape?

TS: Right. As far as I know there no one was using anything but magnetic tape in this arena. However, the Japanese had built a number of different digital audio recorders and playback systems in their labs and they'd bring them around occasionally at AES conventions, but that was growing very slowly and very internally there. I don't think they thought there was a market for it! I think they thought they needed to be up to date with the technology, but they weren't really anxious about commercializing it; you couldn't find anybody at Japanese companies who was talking about selling it. It was more for demonstration, "here - have a look at the future." Denon was doing the most; they were going out in the field with 14-bit recorders and making LPs with them. Then they would use their digital recordings to show how good their analog audio equipment was. They'd put one of these digital recordings on their equipment to show how great it sounded. They were not trying to interest the recording companies in any way I could detect. However, I was. We worked for three years before we made a paid commercial recording.

DL: You are credited with making the *first* commercial digital recording...

TS: Right. We took our machine and we did a digital recording at Santa Fe of an opera in 1976. Of course, by then, we'd made a lot of digital recordings, and so had many others, but only in laboratories. Santa Fe was the first real world recording we made. Everything worked perfectly. Then we demonstrated the recordings at the AES convention in the fall of 1976.

DL: Did you then try to get record companies and recording engineers interested?

TS: Doug Sax was doing a lot of direct to disc recordings, but he wasn't interested in our digital machine. So we went to Crystal Clear records and we did a recording of Virgil Fox, and those recordings were stunning - very, very interesting. That started our cash flow going. We also recorded Arthur Fielder and the Boston Pops. These weren't released right away though. The first commercially released digital recording was for Telarc, Frederick Fennell and the Cleveland Symphonic Winds, recorded in spring, 1978. The thing became afire then; people who had never talked to me before started calling me on the phone and saying, why didn't you tell us it could be this good? That was when people got washed, when people really understood what we were doing and what it meant. By 1981, we had at least 500 digital masters in our vaults that came from various sources.

Telarc and Soundstream brought digital recording to the world. Other people had it, but they were just using it internally. Then, in 1982, Jack Renner [CEO of Telarc] put out the first CD. The role I played, along with my people of course, was to commercialize digital audio and to have it used by the recording companies; not just by the technical people. There's been a lot of talk about who made digital audio first. We had the first commercial digital recorders; 3M was second in developing theirs.

DL: You also were a pioneer in digital editing.

TS: Yes. The funny thing is that even today, people are using editing systems that are very primitive, and our editing system was up and running back when we were in Santa Fe. We had a totally computerized editing system; this meant you didn't have to have tape swishing back and forth; this means you could start at the back of the recording and do your editing backwards if you wanted. We invented hard disk editing, and we were using it back in 1975.

In a sense, we not only pioneered tapeless editing, but we were the only people who were in it. Soundstream stopped business in 1980, but the technology found it's way to other companies.

DL: Let me ask you some hardware questions about these first digital recorders. What was the sampling rate?

TS: At different times we had 3 different sampling rates. When we first put the converters together, we thought that our market would be radio, that people at radio stations would want this. So we were working with a 15Khz bandwidth then, because FM has a 15K bandwidth. We made some recordings with that bandwidth, but not many, not more than 3 or 4. I'm pretty sure that the one at Santa Fe was using that.

DL: So, you're talking about a sampling rate of 32K or so? **TS:** It would have to have been larger than 32K, that's cutting it a little close. We used 37,500. After that we had a converter that was something like 47 KHz - I don't know where that came from, it was some fraction of something - but it was above 44.1. Then all the rest were 50KHz, and anybody that's got one of our machines right now is at 50.

DL: How many Soundstream machines were sold?

TS: Between 15 and 20. I have no idea how many are still in use. My guess is essentially none. We sold our most of our inventory to Bertelsmann. They loved our editor, and worked one of them constantly for eight years. They just put it away in April of this year. I think that editor worked on thousands of CD masters.

DL: Did all of the machines use the original Honeywell transport?

TS: Yes.

DL: Did you have to modify it?

TS: Yes. Basically, what we did was put in a head for 16- track work plus two side channels for SMPTE. We usually didn't use that, we didn't have too many jobs where we had to sync with anybody.

DL: How did you handle error correction?

TS: We recorded a given track on two tape tracks and separated them as far as possible from each other. So for example, we might record on tape track number 1 and 8, and the signals were identical. This correction and detection system was very very good. The only one I can think of that was better was the one that comes with the CDs. We had no problems with this scheme of writing everything twice. The logic of determining if there was an error was very simple and very reliable.

DL: So the way it worked is that you just compared one track to the other?

TS: Yes, we'd see if the two codes were identical. If they weren't identical, we knew which one was bad, because if there was a drop-out the energy on the tape was too low. Incidentally, none of our clients ever found a digital error in any of the tapes we made. And I'm sure there was a terrabit of stuff by the time we quit.

DL: What did early digital sound like?

TS: It sounded great to me, but don't ask me. You know what a golden ear is. Well, I have a *lead* ear. Jack Renner always tells me I have a good ear, but I never would say that's the case. I enjoyed the music and it was definitely superior to anything I had heard before. Our tape recorders were made not just by listening, of course, but by a group of people listening and very, very careful study of whether the thing was theoretically correct.

DL: Did Soundstream ever get around to looking at digital audio for the home? Something like CDs?

TS: That was the idea from the beginning. Malcolm Low brought the idea up in our very first discussions in 1974. In 1980, we merged with Digital Recording Corporation, and the name became DRC/Soundstream, a public company. I did not play a heavy role in the development of the digital player they were trying to put together, but we *were* trying to develop something like the CD.

DL: When you say it was like the CD, do you mean it was optically-based?

TS: Yes, it was. I was running the recording part, so I wasn't in charge of the design there. They wanted to create a "record" that would be 3" by 5", the size of a 3x5 card that you could put in your shirt pocket and carry around with you. In this mode, it wouldn't be the card that moved, it would be the reader that moved. A unit was built and it did work, but it was abandoned when the CD emerged and the design race was over.

DL: There are still a great many people who say that analog sounds better than digital. The battle has been playing out in some of the high-end audio journals, such as *The Absolute Sound*, as well as in the pages of the pro-audio and musician magazines...

TS: Are the people who are saying this in control of a company or a business? Because of course, companies have to make money or they die. And you know, just before people are going to be killed or murdered, they will say *anything*.

DL: Well, there are musicians and artists, respected recording engineers...

TS: Well, that could be, but that's not the question. The question is, does the digital recording sound like the original source? Now because it sounds different from the analog recording, it might be more pleasing to some listeners and less pleasing to others. As far as I can tell, there isn't any defect hanging around in digital.

Have you seen that book about recording by Gaisberg? It's an old book from the 40s. He recorded Caruso way back and he talks about the transformation between acoustic and electric recording as it occurred in 1925. He points out that most of the recording engineers had to give up their trade when that happened because they didn't know what to do with the new recording technology. Now I can imagine that more than a few people were upset about that - they lost their jobs. I wouldn't be a bit surprised if some of them said, "Gee, this electric recording doesn't sound as good, does it?"

DL: Well, I always figured that when people say they like analog recordings, what they like is actually the harmonic distortion. The terms they use, like *warmer, softer, less harsh,* suggest that this is what they like about analog. Distortion can sound warm because it muddies up things. Digital has no distortion, so they don't like it; they're not accustomed to hearing recorded music that way. Do you think this is part of it?

TS: Oh, absolutely. In fact, I think that's almost the *whole* thing. It is that smearing that they like. After I restored the Caruso recordings, we played them back for some collectors and compared it with the originals. Now the originals had lots of surface noise, lots of energy above 8K or so, and the restoration didn't, because it there was no energy in that part of the spectrum in the original recording. And you know what? Many of the collectors liked the sound of the originals *better*, and the only reasonable explanation is because their ears needed to hear energy in this range. I find this very interesting. I should think you would know something about this from your own laboratory work.

DL: Just to exhaust all the possibilities in this digital/analog debate, let me ask you one more question. Is it possible people are hearing artifacts in digital? Bad anti-aliasing, dithering, and so on?

TS: I don't think so. If things have been done right by the people who build the equipment, that's not going to happen. Unbelievable amounts of diligence have been put in to make sure that's not a problem.

DL: Some of the really cheap converters that are built into budget equipment do sound awful, though...

TS: Well, that's a different issue. Obviously you can make something bad if you want to, and if you want to make it cheaper. I feel kind of funny about all this, because like you, I want to find "the answer." But every time I went to find the answer, I haven't gotten cooperation, or the facilities, and so on.

Do you know Lipshitz' work? He has done more than anyone to put this out in the open. He writes in the AES journal. He's tried to make the role of dithering understood by people who haven't understood it. He was president of the AES a couple of years ago. As an academic, I think he is the person in the world who most knows what can go wrong - the unbelievable things that can go wrong in research studies when people try to do them honestly - and dishonestly. Lipshitz talks about the whole notion of telling whether things are different, whether you can hear the difference between A and B. He points out that if you wanted to try and determine if two things are identical or not with 95% confidence, you have to get things right every time in 11 consecutive trials.

DL: You mean because of the statistics of certainty...

TS: Right. So for example, if you listen to three recordings and start pontificating on the differences, and someone comes along who knows whether this is accident or talent, you're not even on first base with this sample of three. If someone is going to come in and listen and "A - B" some recordings, you can't believe they know what they're talking about, or that the difference exists, unless they get it right 11 times in a row.

DL: You were part of the expert panel that examined the Watergate tapes. What was the assignment of the Watergate panel?

TS: It was just a few days before Rosemary Wood came up with the existence of the 18-1/2 minute gap. And so any plans that were being made at that time were cleared up quite quickly right then and there. Our team then spent essentially 6 months writing a report about the gap.

DL: Did you get a hold of the tape and try to recover the portion that was erased?

TS: Oh yes. We did a *very* thorough job trying to recover it. Unfortunately, it was erased by a stenographer's recorder which has a double erase head, and absolutely no human voice sounds were there except in a couple of places where the instrument used was stopped and then started again. But it was obvious, in the final analysis, that the gaps were created by the pushing of a manual button on the recorder. Also obvious was the way in which it was done; without a doubt, it had to have been done by a finger pushing this manual button.

DL: Do you see something replacing the CD soon as new technological breakthroughs are made?

TS: I haven't been following this, but the concept that some people have had - namely that you could have a "CD" on a chip - is still pretty far off. You have to take into account the "25-year syndrome." The syndrome is that you cannot change recording media faster than once every 25 years; in particular, there was the Edison era, which was about 1877 to 1900; then the disc reigned as a primary medium, and of course it wasn't electric for 25 years; then the electric recording survived until 1947, when the LP came out. (I'm going to put the medium of tape aside here.) Then came the CD in 1982. It was a longer stretch that time, maybe because of the two media being around, the cassette and the lp. And of course, another retardant [for the 25-year rule] was the advent of stereo in the middle there. But the industry isn't going to put up with another major change for another 15 years or so. You have to realize that everything before the CD was needle and groove, so it might even be more robust this time because the technology is so different now. But I'm sure there *will* be a change, counting when it started in 1982 and then adding 30 years.

DL: That long? What if technology allows for a dramatic change in format or type? What if you could get the playing time up to 10 hours, or the size down to 2 inches; or if the indexing gets better?

TS: I don't think that size or any of those other things is going to have the weight that a new change in sound quality would. The things you mention are just conveniences in my world. I don't think they would have the thrust to create a major change.

DL: You think that change is driven by sound quality?

TS: I do, I really do. When you look back historically, it was sound quality that drove the changes; the first records were so much quieter than the Edison cylinders, which were very, very noisy. And the advent of vinyl was a huge quality improvement.

DL: You were first author on a famous IEEE paper in 1975, "Blind Deconvolution Through Digital Signal Processing." The article describes your work in restoring those old Caruso recordings, but it also talks about work you've done in enhancing blurred visual images. What is the connection between your work in audio and in vision? Have they cross- pollinated each other?

TS: Yes, because in fact the whole technology is blind deconvolution. In both modalities [sound and vision], our work was based on the ideas put out in an earlier paper, "Nonlinear filtering of multiplied and convolved signals," Oppenheim, Schafer and Stockham, published in the IEEE journal in 1968. That was a very large bomb on the EE plateaus. The notion that you could do linear filtering for non-linear systems wasn't universally well received. The whole thing stood on Al Oppenheim's doctoral thesis - and you see, it is really very simple. Mathematicians have known for years that you could make a transformation from multiplication to addition; even children learn this in school: it's just the logarithm. There's a theorem in modern algebra that says that if you have a vector space - and this is a modern algebraic vector space, not what you would talk about if you were doing electrical engineering with electricity and things like that - Anyway, if you have a vector space, then if the rule for combining vectors is not type A, you can make it be type A by a unique 1-1 transformation creating another vector space with a different rule for combining things. What that says of course is that if you have something that doesn't combine using addition, but using some other transform, you can force it to use addition.

DL: That's the key to the deconvolution problem, then. In the particular case of the Caruso recordings, you have two convolved signals; the signal from Caruso is convolved with the response characteristics of the old mechanical horn recording mechanism.

If I understand you now, you're saying that you can take this convolution, apply an FFT to create a multiplicative function, and turn it into an additive function. Once you have the latter, it is trivial to separate the two functions, allowing you to restore the sound of Caruso's voice.

TS: That's right. As you know, if you apply an FFT to a convolution, the convolved signals are then no longer convolved; they're multiplied. Because you've gone from the time domain to the frequency domain. And when you do that, as everyone knows, you go from convolution to a product. Then you just take the log and you've got a sum, and you can you apply regular linear theory.

There are two arenas for making an interesting practice for using these things. One is taking multiplied things and making them additive; the other is taking convolved things and making them additive.

DL: In the paper you talk about deblurring photographs as being the same problem conceptually as the dereverberation you did in sound. Ideally, you would have several recordings of Caruso with the same horn impulse response, and analyzing these would allow you to extract out the horn response - because it would be common to all of the recordings. But in this case, you only have one example, so you sliced up the image into a bunch of smaller frames, assuming that whatever it is that created the blur will exist commonly in each frame. Is this how NASA deblurs Mars pictures?

TS: Well, I haven't had much contact with them, so I don't really know how they do it. But, yeah, I imagine it must be very similar.

DL: Presumably, NASA has thousands of pictures of Mars all taken with the same camera from a similar angle, so the problem should be easier for them.

TS: Yes, you're right. And this is fundamentally a deconvolution problem of the type we've been talking about, but they don't need to chop up the individual image so they can get much better resolution than if they only had one picture.

DL: In audio, is surface noise an example of an additive function?

TS: Yes. And there are analogies in vision, as well.

In black and white photography with film, how much silver do you have to put in the film so that when you examine it on a light table it appears to be the same as the original scene? The answer is, the log of the exposure. So photography is multiplicative. But you can transform it into being additive by taking the log.

In audio, an automatic gain control is multiplicative. The 1968 paper discusses this, and describes four of our experiments. The first one was an automatic gain control where you'd take the log of the signal, process it linearly, and come back out again and exponentiate the result. That gives you an automatic gain control.

DL: This sounds like a compander.

TS: Right, that's what I did, I made a compander.

DL: Well, in fact, didn't you make the first compander? **TS:** Yes, I did, but only the first *homomorphic* compander. Other companders were developed starting in the early part of the century. The telephone company used them, for instance, to reduce line noise.

DL: How did the commercial companders, the DBX, for instance, differ from yours?

TS: I think the DBX units had a great similarity to the original compander I built.

DL: Maybe you could describe, for the readers who don't know, what a compander does and how it works.

TS: If you have a dynamic range of *X* db you can make it have a dynamic range of *X/2* db, or any other dynamic range you want to have. I patented a way to make a compander that would do that kind of thing. It takes the complex log of the signal, then it puts the real part of that through either a linear or non-linear filter. Next it exponentiates the filtered signal, and then restores the sign by multiplying the exponentiated filtered signal by the imaginary part of the complex log.

The filter is designed to be a low frequency attenuator if the compander is in the compression mode, and a low frequency amplifier if the compander is in the expansion mode. It makes a very nice compander. But not good enough for ultra high-fi. That's what made me do what I did in developing digital recording.

DL: How is this different from compressors and expanders such as the recording studios use?

TS: Well, those are designed to alter the signal for particular purposes. A compander should allow you to take a signal, compress it and expand it and have it be intact, so it's just like the original.

DL: You said it was clear to you it would be impossible to make one with high enough fidelity, but then Ray Dolby did make one.

TS: The compander that Dolby made was very, very good work. You know how that works. When you have tape noise, you make the dynamic range of the signal less, so everything's louder - the loudest things are now just as loud as they were, but the softest things are 50 db louder. That means that when you go and expand, the noise has been pushed through a floor that you didn't have before. Unfortunately because the signal is bipolar it's very hard to control. Because the signal is positive then negative and so on, you have to have a very quiet switch. I'll tell you, making digital recording work is a lot easier than making this work. Dolby made this work, because his things aren't that decontrolling. He doesn't try to do too much compression.

DL: What was the engineering breakthrough that allowed him to make this?

TS: His brain, I think. He's a very creative guy, very creative. He was a star, you know, in the TV era. He was one of the people who put together TV recording in the 50s, so that you could have delayed television broadcasts; he was part of the Ampex group that did that.

DL: Have you heard Dolby SR?

TS: It's not bad. It won't do what digital will do.

DL: Some people say it is better than digital, that it's *smoother, silkier, warmer*...

TS: Well, we're back to what we were talking about and the dichotomy between fidelity and what you like. Also, I haven't had a chance to look very deeply inside the machines it's competing against, the Sony and Mitsubishi digital recorders, so I don't know what's in them; I haven't had a chance to see for myself whether they work right or wrong, so I can't comment.

DL: Do you think this is a case similar to the Caruso restoration where your collectors liked the noisier recordings better?

TS: There is definitely a relation. When we first were using the DBX, we did some experiments where we recorded one tape with DBX and the other without. We had it set up so you could switch between them for playback. When you turned on the normal one, everyone would be happy. But when you got rid of all that noise and the hiss went away, it was dull and uninteresting. Then, when I said, "let's compare the companded one with the *original*," and I threw the switch, they couldn't tell the difference!

This is a matter of "fidelity" versus "what you like." I want to be sure that you understand I have *no* feelings of any type that people shouldn't like what they like. But I do get upset when people don't understand that what you like might not be exactly like the original you were trying to put back together in the recording process. There's nothing wrong with not putting it back together the same as it was, it might even be a lot more fun. But if you're talking about fidelity, you shouldn't say that you don't have it when you do. It is important to distinguish between whether what you're talking about is coherent or not.

Back to Daniel Levitin's Homepage