Any Song, Any Time, Any Where

Kees A. Schouhamer Immink, AES Fellow.

Abstract
The author will take a look into the past, and will try to find essential contributions to the digital audio 'revolution'. In addition, he will try to identify some clues that could make it possible to understand the directions of digital audio engineering in the future.

The invention of PCM

In 1952, the year when the Audio Engineering Society (AES) was founded, the life of an electronics engineer was relatively simple as there were not many practical options for analog telecommunications or recording, but as we shall see, a few key developments would change that simple life. Those were the days that shellac gramophone and AM radio ruled. The 12'' Long Playing Vinyl disc had just been brought to market. Spinning at 33 rpm this new high-density disc could play up to 25 minutes per side. RCA produced the 7'' 45 rpm vinyl disc. It took many years before 78 rpm to disappear because the new vinyl formats needed new equipment on which to play them. But after a slow beginning, the two new vinyl formats would rule until the Compact Disc took over in 1982. Magnetic tape audio recorders were used in the professional studios for sound recording. Les Paul created audio dubbing, and brought this technique to very high artistic and engineering standards. Video recorders using helical scan heads were first brought to market by Ampex.

Telecommunications, voice and telegraphy, was mostly done by cable or radio. AM and FM broadcast radio brought sound directly from the studios to the listeners. All signals were analogue, and in 1952, digital technology was a long way to go. But times were changing, and the good observer could have noticed the many signs of change.

The Brit Alec Reeves (1902-1971) invented pulse-code modulation (PCM) in 1937 when he worked for the International Telephone and Telegraph Company. This is a very early invention in the history of electronics, since it is only a few years after Edwin Armstrong invented wideband FM, a method of high-quality radio broadcasting. Reeves, instead of following tradition by sending an electrical current being proportional with the sound level, proposed that the electrical sound signal be sampled and digitised at regular intervals. Then the analogue value of each sample would be rounded to the nearest integer value, which, in turn, is represented by a binary number and transmitted as unequivocal on-off pulses. In principle, the binary, two-level, signalling was a return to the simple, robust technique used by the telegraph. Noise immunity and fidelity benefit tremendously because the sound signal is no longer stored as a delicate analogue signal but as a much more robust sequence of binary numbers. Because PCM is a method of representing an analogue signal in digital form, it is particularly well adapted to work directly with digital data-processing equipment. A notable disadvantage of PCM is the required high (analogue) bandwidth of the transmission or storage system. Sending recognisable speech meant networks and radios would have to carry millions of pulses a second. Though Reeves' extraordinary patent showed how this might be done in theory, the valve-based technology of the time could not do the job. In his 1937 patent, Reeves formulated the major advantages of digital PCM transmission, namely

Quality depends on conversion steps ONLY
Quality independent of transmission media
Compatibility with different media and traffic (video, audio, and data)
Low cost
New features can easily be embedded.

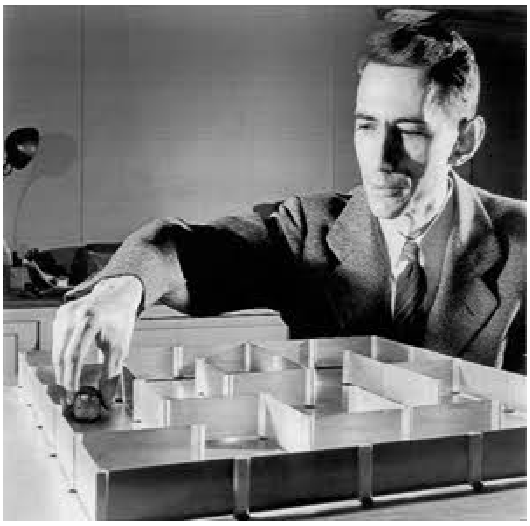
These are gigantic and also visionary conclusions. Reeves showed his enormous engineering foresight, as there are two essential assumptions that are implicit in the above characteristics. Firstly, each quantized sample can be transmitted with arbitrarily small probability of error. It was absolutely not clear in 1937 that this could be accomplished in theory; let alone that he, or others, knew about practical methods for achieving error-free transmission. There was no research, not even by an obscure Russian mathematician, on the topic of error correcting codes. It would, as we will see shortly, take another ten years and a world war, before research on error free digital communication would take off. Secondly, he assumed that conversion from the analog to the digital domain, and vice versa, could be done, either in theory or practice, with arbitrary small accuracy by use of sufficiently frequent sampling, and by quantizing each sample with a sufficiently large number of levels. Early theoretical work by mathematicians had been published, but we may well assume that Reeves was unaware of that literature. Before we take a look at the two basic laws that underlie the existence of digital sound, we divert for a while to a great engineering achievement, namely the sigsaly digital encryption system.

Pulse Code Modulation could not be implemented economically until the invention of the transistor decades later. For military operations, however, economy is not always at a premium. As we all know, the war was very fruitful for technology. A transmission system based on full-fledged PCM was first developed at Bell Labs for the complex and very advanced ‘Sigsaly’ radio voice encipherment system on which Churchill and Roosevelt talked in total secrecy for during World War II. A team of top engineers at Bell Labs started to build the first digital telephone system. Churchill and Roosevelt used the top-secret SIGSALY telephone encryption system to discus war matters. Terminals were in London and the Pentagon, Sydney, and other places. The older analogue scrambling technology was not sufficiently secure, and indeed by 1941 it was broken by the Germans. Sigsaly’s technology is absolutely fascinating. The basic technologies employed were not simply improvements upon prior art; they were fundamentally new and absolutely necessary for the system to work. The concepts were proven in the laboratory, but before the system was ready for final development and deployment, there were several important system features that needed refinement. Just one of the many major problems that the Bell team faced with was the generation of the secret key. The basic requirement for the key was that it should be completely random, and must not repeat, but could still be replicated at both the sending and receiving ends of the system. To that end, random wideband noise was generated by a mercury-vapour rectifier vacuum tube. The noise signal was sampled every twenty milliseconds and the samples then quantized into six levels. The level information was converted into channels of a frequency-shift-keyed (FSK) audio tone signal, which could then be recorded on standard vinyl phonograph records. Two exact copies of the master record were made: one for the sender and one for the receiver. The random digitised signal from the record was combined, *scrambled*, with the digital speech signals, and the scrambled audio signal was transmitted to the receiver. The receiver subtracts the same random signal from the received signal, and obtained the original, unscrambled, speech signal. However, in order to do so, both ‘random’ signals had to be generated in exact synchronism at both the sender’s and receiver’s site. The playing of the two records with the random signal was accomplished by the use of very precisely driven turntables driven by a stable electric motor. For stability reasons, the motor was kept in constant operation, and the frequency for it was derived directly from the terminal's frequency standard, a crystal clock oscillator. For security reasons, the records were used only once, and immediately destroyed following a session.

The block diagram of sigsaly’s voice speech encoder (vocoder) resembles the block diagram of modern audio compression technology such as the one used in a modern mobile telephone The voice spectrum from 250 Hz to 3000 Hz was divided into ten channels; and all ten channels were sampled. Two channels sampled voice pitch and unvoiced background hiss.

The hardware of that steam voice encoder and decoder did not fit into a standard living room. Each terminal used forty racks of equipment weighing about fifty tons, and filled a large room. The installation in London was so large that it required the basement of Selfridges Department store. The Bell engineers could be proud on their product: the Germans did in fact monitor "Sigsaly" for two years, but were never able to unscramble the voice communications.

The period of time during and immediately following World War II was incredibly ripe for technology. Claude Shannon (1916-2001) published his landmark article "*A mathematical theory of communication"* in 1948, which laid the foundation of Information Theory. Information theory is primarily concerned about the *mechanics* of information handling. The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have a meaning: for example it is an audio or video signal. This 'content' aspect of communication is irrelevant to the engineering problem of error free



communication. By the way, Shannon coined the term 'bit' for an elementary measure of information. Shannon proved mathematically that communication over noisy channels is possible with arbitrarily small probability of error. He did not give a clue how engineers could reach the promised land of error free communication in practice. Real error correcting codes were first invented by Hamming in 1948; his code could correct single bit errors. The mathematicians Irvin Reed and Gus Solomon formulated the error correcting capacity of their renowned codes in paper in 1960. Only five pages were needed to present their results. They rigorously showed that Reed-Solomon codes can correct any number of errors within their designed error correction capacity, but it took some fifteen years before mathematicians found an efficient algorithm for decoding such codes. It took another ten years before the codes were practically applied, first in magnetic audio tape recorders, later in CD players, and thereafter in essentially all other digital communication and storage products. The error correction of a $25 DVD player can correct about 100 byte errors per data block.

Sampling Theorem

The Sampling Theorem was independently formulated by at least four scientists, namely Whittaker, Shannon, Kotelnikov, and Nyquist. Whittaker was probably the very first who presented the theorem in 1915. Shannon, who first formalized Whittaker's result and presented the Theorem to a wider, engineering, audience, when he gave this Theorem as a kind of 'bonus' in his 1948 landmark article.

The Sampling Theorem in Shannon's words reads: If a function *f(t)* contains no frequencies higher than ω (the 'bandwidth') it is completely determined by giving its coordinates at a series of points spaces $T = \pi/\omega$ seconds apart.

The sampling theorem is by far the most eminent equation in digital audio engineering as it mathematically proofs that an audio, or other continuous-time signals, can be completely restored in case it is observed (sampled) at a rate at least twice the bandwidth of that audio signal. The mathematical elegance of the sampling theorem might suggest that analogue-to-digital (AD) conversion or vice versa is a straightforward procedure in practice. Reliable converters, however, are difficult to construct, and they are still the subject of both academic and industrial study. The conversion steps have always been very difficult. While it has been understood for a long time that for example dither can be used to eliminate distortion in AD converters, it has not always been deployed. As a result, PCM high-quality audio systems have received criticism over the years.

It is of some interest to observe that engineers were very early with the sampling of electronic signals. The sampling of signals goes back almost 100 years when multiple simultaneous telegraphy signals were transported over a single telegraph line. Clearly, long and under seas telegraphy cables are very expensive, and simultaneous usage of that cable by a group of independent telegraphers would increase the profit proportionally. Synchronous rotators at the sending and receiving end sampled the telegraphy signals. The engineers observed that if the rotation frequency was sufficiently high that the telegraphy signals could be transmitted without distortion. And now we know that it took almost fifty years before that rule of thumb was mathematically proved.

The Roaring Seventies

We now take a big step in time from the fifties to the seventies. Practical progress was made at Bell Labs and other places to practically construct PCM

transmission. Progress was slow as the key component of PCM transmission was very young, and just invented. The transistor, born in 1947, would make PCM possible in practice. Large mainframe computers were transistorised, and later, in the sixties, small so-called mini-computers, for example the famous PDP-8 and PDP-10, were brought to market. The mini computers were cheap and its proliferation triggered research in low-cost data storage. In the early 1970s, Thomas Stockham used mini computers, and regular computer tape storage, for his pioneering digital recording of sound. In 1975, he and Malcolm Low founded Soundstream, where they developed a 16-bit digital audio recorder using a high-speed instrument magnetic tape recorder. Soundstream was the first commercial digital recording company in the United States, located in Salt Lake City. Companies such as Denon had been experimenting with digital recording since 1971, but Stockham was the first to make a commercial digital recording, using his own recorder in 1976, and demonstrating the recordings at the 1976 AES convention. In the mid 70s, digital audio made its way into the recording studios. The following survey taken from an AES Journal shows some of the PCM systems and its key parameters.

Manufacturer	***Type***	***fs***	***resolution***
BBC	Microwave	32	13
Soundstream	tape	37.5/42.5	16
Many	PCM adaptor	44.056	13
Denon	tape	47.25	13

High-quality PCM audio requires a significantly larger bandwidth than a regular FM audio signal. For example, a 16-bit PCM signal requires an analogue bandwidth of about 1-1.5 MHz, and, clearly, a standard analogue audio recorder could not meet that requirement. The obvious answer, at that time, was to use a video tape recorder, which is capable of this high bandwidth, to store the information. Such an audio recording system therefore includes two machines, namely the PCM adaptor and the video tape recorder. A PCM adaptor has the analogue audio (stereo) signal as its input, and translates it into a series of binary digits, which, in turn, is modulated into a pseudo-video signal. The pseudo-video signal can be stored on any ordinary analogue video tape recorder, since these were the only widely available devices with sufficient bandwidth. This helps to explain the choice of sampling frequency for the CD, because the number of video lines, frame rate and bits per line end up dictating the sampling frequency one can achieve if wanting to store two channels of audio. As we all know the sampling frequency, 44.1 kHz, was also adopted in the Compact Disc, as at that time, there was no other practical way of storing digital sound than by a PCM Converter & video recorder combination. The sampling frequencies of 44. I and 44.056 kHz were thus the result of a need for compatibility with the NTSC and PAL video formats used for audio storage at the time. The Sony 1600 was the first commercial video-based 16-bit recorder, and continues in its 1610 and 1630 incarnations. Key parameters: sixteen bits quantisation and the sampling frequency is 44,1 (or 44.056 for NTSC) kHz. The PCM adaptors could only store a stereo signal, and could not be used for studio multi-track recording.

Much later we have witnessed the advent of dedicated professional digital multi-track recorders such as Mitsubishi’s ProDigi format and Sony’s DASH format. These recording machines accommodated the obligatory 44.1 kHz, but also 48 and 32 kHz as sampling rate. Digital mixing tables and other digital equipment, such as reverberation, were introduced, and have become indispensable tools in modern sound studios.

The Big Bang: Compact Disc

From 1973 to 1976, two Philips engineers in Eindhoven were given a mandate to develop an audio-only disc based on optical videodisc technology. They started by experimenting with an analogue approach using wide-band frequency modulation. The problem with this was that it was not really much

more immune to dirt and scratches than an analogue LP record, although there was a certain improvement in sound quality, so they decided to look for a digital solution as electronics technology had become ripe for such a step. They were successful, and around 1977 Philips, Sony and also other companies demonstrated the first prototypes of a digital sound system using a laser disc. In 1979, Sony and Philips decided to work together, and they set up a joint group of engineers, whose mission it was to design the standard of the new digital audio disc. Philips had lost the market for the videodisc, but had considerable optical expertise, as well as expertise in servo systems and digital and analogue modulation systems. Sony's huge expertise in error correction, PCM adapters and channel coding would complement this ideally. A reasonable summary would be that most of the 'physics' was provided by Philips and the digital audio experience by Sony. The Compact Disc Digital Audio System was first brought to market in 1982. Sony was the first company to develop a portable CD player in 1985. The new audio disc was enthusiastically received and its handling quality received particular praise. Two years later, in 1985, the CD-ROM (read-only memory) was introduced. With this it was now possible to disseminate massive amounts of computer data instead of digital sound. A user-recordable CD for data storage, CD-R, was introduced in the early 1990s, and it became the *de facto* standard for exchange and archiving of computer data and music. The CD and its later extensions have been extremely successful: in 2004 the annual worldwide sales of CD-Audio, CD-ROM, and CD-R reached about 30 billion discs.

Twenty years after CD

We are now in a very exciting time: on the one hand we see the advent of systems with extremely high audio quality demanding super high-density recording systems such as DVD, the successor to the CD. On the other hand, we see systems offering low quality ‘compressed’ sound requiring low-density storage.

An audio compression code such as, MP3, which is short for MPEG-1 or MPEG-2, Layer-3, compresses sound files by removing the inaudible information from the recording. The result is a smaller audio file, which can be transferred more efficiently without ‘any’ depreciable loss in sound quality.

Compressed audio files are about a factor of ten to twenty smaller than the original CD file. Any fears the record industry had about digital recording in 1992 now seem insignificant with the advent of compressed audio, MP3, and the explosion of the Internet. Students who had an Internet connection and a computer first used compressed audio. The situation changed completely when a large public started to use mobile players using solid state or hard disk drives (Apple Ipod). These devices have completely replaced Sony's Walkman.

From Hi-Fi to Wi-Fi

The record companies have been using the same business model for almost 100 years. Edison and Berliner introduced this classical business model, when they invented the gramophone. In this model, the distribution of electronic sound has been accomplished by radio or gramophone. We are all accustomed with this model. You go to a shop and you buy a carrier of music, ranging from the cylinder, gramophone, compact cassette, to CD. You take it home and store it at home or in your car. You can play the song you bought any time and anywhere you want. Alternatively, you can listen to music by radio or TV, but here you can choose neither the song nor the time. The digital CD has replaced the analog gramophone in 1982, but, at that time, it did not change the above business model.

In a few years from now, each home and car will have a *digital music library*, which is placed in the broom closet, basement or attic. *Networked audio players* are connected to the music library by standard Internet technology such as UTP cable or Wi-Fi radio. Networked players are in essence dedicated computers that have an Internet connector and some software for playing and selecting the songs. The music is not necessarily of low capacity compressed quality. A digital music library will have a storage capacity of say 1 Terabyte, i.e., 1000 Gigabyte. Thousand gigabytes is equivalent to the storage capacity of 2000 standard full-quality audio CDs. Alternatively it can store 50,000 CDs in compressed quality, which amounts to 5 years of unrepeated around the clock music. All the above electronic equipment is essentially state of the art, and there are no inventions to be made. For a reference: at this moment (2004), commodity PCs sell with disk drives of about 80 Gigabytes, and an extra disk drive of 250 Gbyte will cost about $100 more. Thus, the central storage of one hour of music costs about half a dollar cent, while a CD jewel case (i.e. without CD) costs more.

Alternatively, we may see the creation of music-on-demand services. A digital networked player can also play songs directly taken from libraries somewhere on the Internet using streaming audio. This network will be the biggest jukebox in the universe. No distribution costs, no pressing costs, no returns, no out-of-stock items.

The primary question will be: how much storage will be done at home, and how much will be done elsewhere. In other words, what will be the balance between the audio on demand service and the locally stored music? This, clearly, will depend on the price tags attached to the services, and on the availability of those services. At this moment, it is not clear where the balance will be, but one thing is ultimately clear, in this new business paradigm, there is no place for CDs or super CDs. I predict that they will become obsolete within five years or so. So in a few years, after storing the content on your CDs onto your central music library, compressed or uncompressed, you will bring all your CDs to the attic, and hopefully there is some place next to your dusty collection of old vinyl gramophones that you put there some twenty years ago.

Clearly, the digital technology, specifically the Internet, is, at this very moment, completely changing the classic distribution paradigm. As more and more media content becomes available in digitized form the fear of the content providers due to illegal means of duplication and re-distribution increases. This is particularly relevant when the disk drives used to store the content are external and therefore easily “portable”. As for music piracy, or downloading music for free from the Internet, everyone believes it's a problem. At the end of the day the technology will help songwriters and performers more than it will hurt them, since the cost of music distribution itself is considerably less than in the old model. The difficult task now is for the record companies to catch up to changing technology and consumer demand.

Conclusions

Digital audio has a rich history, and the membership of the Audio Engineering Society has played a key role in its development. Considering the long history of digital sound, it seems to be incorrect to use the term “digital audio revolution”. The term evolution would be more appropriate, unless one specifically deals with the short time span, say ten years, following the introduction of the Compact Disc.

For me it is clear that the future of digital audio will be rich as well, but we have to be very careful that we come up with new products that will attract the main base of consumers, and is not confined to the niche market of audio buffs. Therefore, I believe that, for example, Super Audio CD or DVD-Audio are a niche product, which will not be purchased by the modal audio buff. Such an orthodox audio fundamentalist prefers vinyl anyway. He (it is seldom a she) does not have faith in the sampling theorem. I remember very well that I heard the following argument during a Convention: “May be it is true that mathematics shows that one can faithfully reconstruct the signal between samples, but the lost emotion can never be reconstructed.

References

J.V. Boone and R.R. Peterson, The start of the digital revolution, SIGSALY secure digital voice communications in World War II, Fort George G. Meade, Md., Center for Cryptologic History, National Security Agency, 2000.

K.A.S. Immink, '*The Compact Disc Story*', AES Journal, pp. 458-465, May 1998.

D. Kahn ‘*Cryptology and the Origins of Spread Spectrum*’, IEEE Spectrum, Sept. 1984.

S.P. Lipshitz, *‘Dawn of the Digital Age’*, AES Journal, pp 37-42, vol. 46, 1998.

C.E. Shannon, ‘*A mathematical theory of communication’*, Bell System Technical Journal, 1948.

M. Unser, ‘*Sampling – 50 Years after Shannon*’, Proceedings IEEE, vol. 88, pp. 569-587, 2000.