

THE DEMOCRATIZATION OF AUDIO MEDIA

Rick Shriver

Ohio University

April, 2012

Preface

As I began this paper, I felt compelled to confirm my understanding of the notion of “democratization.” I thought I knew what it meant to “democratize.” I had read some of the dozens of articles on the “democratization” of other forms of media and art, which suggested that democratization meant to “put the power with the people.” (Ogorek, 2009.)

Thus I am positing a working definition of the “democratization” of audio media as “putting the creation and dissemination of audio content into the hands of the consumer.” Further, I am contrasting “democratic media” with “aristocratic media,” “plutocratic media,” or “elitist media,” where filters are in place to regulate which work is ultimately disseminated to the public.

When an art form or medium is democratized, many have worried that the quality of the content is lowered (arxitecture.org, 2011.). That is the position I am taking in this paper: the tradeoff that we make when we remove the filters, and put the power into the hands of the people to create and distribute media content, some of the consumer-created content will be of lower quality.

Many authors have further suggested that a (political) democracy will eventually collapse or transform itself into some other form, and that all such transformations are cyclical. So we can relax, this will pass.

A handwritten signature in black ink, appearing to read "Rich Swen". The signature is fluid and cursive, with a long horizontal stroke at the end.

In 2009, while attending a recording workshop in Nashville, Tennessee, I heard (bluegrass artist) Ricky Skaggs recite an anecdote about performing and recording with his sons. He said that it was very gratifying to share such an artistic pursuit as music performance and recording with one's children, but that he was often vexed at his sons' acceptance of poor quality recordings. Finding a classic recording by one of country music's seminal artists, they would hand him an MP3 player and gush, "Dad, check this out! Isn't this great?"

Most often Skaggs said he would respond by wincing and saying, "No, it's not great. It sucks." He would then locate the same recording on vinyl, put it on the turntable and say, "Now listen to this, this is great."

What Ricky Skaggs found objectionable in his sons' recordings were the artifacts introduced by the digitization and compression required to convert the recordings to the ubiquitous MP3 file format. That anecdote is really the genesis of this paper, as I began to consider the effects of the proliferation of the MP3 player, its associated file type, and the common mode of delivery for those files: the Internet.

My thesis is simple. Twenty-first Century consumers willingly accept lower fidelity recordings than those made in the mid-to-late Twentieth Century – because they know nothing better. Or, in the vernacular, "these kids wouldn't know a good recording if it hit them in the ear buds."

Consider the quaint and archaic vinyl phonograph record, which even with its inherent shortcomings of surface noise, fragility and noticeable degradation over time and repeated

playing, is capable of a “theoretical” frequency response range from 7 Hertz to as high as 50 Kilohertz (7Hz – 50KHz), and a dynamic range of over 75 decibels. (Grey, 1997.) Of course not all vinyl recordings are created equal, and playback equipment varies wildly, but the potential fidelity for vinyl seems practically limitless when contrasted with the MP3, which depending upon the codec employed may allow for frequency response ranges in the neighborhood of 16 Kilohertz (16KHz), but nearly always exhibit significant roll-off beyond that. (Miller 1999.)

One could argue that the low frequency response roll-off for MP3s is irrelevant for a couple of reasons. Much of contemporary audio content is derived from compact disks. The *Red Book* standard for CDs sets 20 Hertz to 20 Kilohertz (20Hz-20KHz) as the frequency response range. (Collins, 1998.) Thus, compact disks that are mastered to *Red Book* standards do not contain frequencies beyond those limits. Most equipment manufactured for playback of consumer digital audio is built around that same standard. CD players just don’t reproduce anything beyond those limits, and it would be the rare set of ear buds that could reproduce below 20 Hertz.

Beyond those inherent technological limitations, this paper will further argue that the overall quality of available audio content has dropped in some part due to the explosion of “amateur” content. In the Twentieth Century there was a clear delineation between “amateur” and “professional” recordings. Amateur recording equipment was generally incapable of achieving the quality that was considered acceptable for mass distribution.

Over the last twenty years affordable computer-based Digital Audio Workstations (DAWs) have undoubtedly allowed greater access to high quality recording technology for increased numbers

of enthusiasts. Readily available consumer-market technology can record, edit and reproduce audio with frequency response and dynamic range that will far exceed the specifications for *Red Book* CDs or any form of broadcasting, including HD. Presently an “off-the-shelf” computer in the \$500 price-range running on the *Windows*TM operating platform (with appropriate peripherals), and loaded with free software is capable of recording multitrack audio that exceeds the frequency response range of the human ear. The user interface on most recording software is intuitive enough for novice users to quickly grasp. File management, storage and editing are accomplished using menu-driven or “drag and drop” commands familiar to any computer-user. “Home made” audio content can easily be burned to compact disk or uploaded to the Internet.

Contrast the foregoing with the technical requirements for multitrack recording twenty-five years ago. In 1983 TASCAM introduced the Model 38, which was an eight-track reel-to-reel recorder that used half-inch tape. “The introductory price was under \$3,000. The 38 filled a niche for the serious entry level recordist.” (Tascam, 2010.) Add the cost of a stereo mastering deck, a modest mixing console, cables and microphones, and the cost of a “prosumer” multitrack rig was in the range of \$10,000. And still, as the niche term “prosumer” implies, the technical specifications of this level of equipment was only adequate, and far from audiophile quality.

In the same way, small inexpensive digital recorders that store data on solid state *secure digital* or *compact flash* media have made it easier and cheaper to capture location recordings. The quality of those recordings can be much higher than that of previous tape-based technologies, such as reel-to-reel tape, cassette or microcassette. Recorders costing less than \$100 can record

several hours of audio at “CD quality,” a term that has, somewhat ironically, now come to represent the highest fidelity we can expect.

This easy and affordable access to recording technology has been hailed by some as a renaissance in the recording arts. With technology readily accessible, affordable and much easier to operate, “anyone” can record audio content, and many people are doing exactly that. Social networking sites like *Facebook* are full of examples of amateur recordings, as are media sharing sites such as *YouTube* and *VIMEO*.

But this paper is arguing that because of the proliferation of this affordable and easy-to-use technology, and the subsequent explosion of “amateur” content, there has been a concomitant lowering of expectations, standards, and in what is “acceptable quality.” That is, the lowering of the standards of acceptable quality of audio content is partly driven by the very accessibility being fostered by decreased costs and increased availability of recording technology. Whereas recording equipment was once costly and arcane, it is now available at the local discount department store and requires little or no instruction to operate. In other words, anybody with a few hundred dollars can be in the recording business.

This phenomenon is what some media scholars hailed as the “democratization” of media a few decades back. Some visionaries predicted that the lowered costs and easy availability of media technology would allow the citizenry to create and distribute content, unfettered by the requirements and conventions of the “mainstream media.” Well, the future is here and the

proletariat is in fact creating and distributing content. Unfortunately, as Ricky Skaggs has pointed out, much of that content sucks.

Some may argue that violating the proscriptions of the mainstream recording industry is just what is needed. Indeed, there is always a place in the arts for innovation and revolution. But many of us would agree that when, in the words of communications scholars of the twentieth century, “Noise obscures the message,” there is little communication occurring. Put another way, if the listener is so distracted by the distorted kick drum or has difficulty hearing the vocals buried in the mix, the impact of the message in the lyrics is lost. As an instructor in audio recording, I have observed this phenomenon frequently. The student in a recording class is so thrilled to have successfully captured the sound of all the instruments in the band; they fail to notice that the _____ (insert: levels, panning, equalization, effects, mixing, distortion, background noise, etc.) distracts the listener.

Note that I am limiting this discussion to the compromising of technical standards, not performance. Clearly, one could easily make a similar case for the consumers’ willing acceptance of poor performances as Internet sites have become clogged with “viral videos” of off-key caterwauling. Witness Rebecca Black’s auto tuned monstrosity, “*Friday*.”

In this paper, I argue that a second contributing factor to the lowering of the acceptable quality of audio content is simply the inherent limits imposed by all digital technologies. A digital wave is still not an exact duplicate of its analog counterpart. Even at the generally-accepted consumer standard *Red Book* “CD quality” sampling and bit rates of 44.1 KHz at 16 bit, some harmonic

content is lost. In other words, we have learned to expect and obtain less from the audio recordings that we hear because the best digital recordings are still “facsimiles.”

We know that the Nyquist – Shannon principle states that with these response standards of 44.1 KHz at 16 bits, it is possible to capture and reproduce frequencies that exceed most humans’ frequency response range of 20 to 20,000 Hertz. But in the analog world, musical notes and frequencies will interplay and create harmonics and overtones at the multiples, sums and differences of the fundamental frequencies.

Thus it is likely that frequencies are present in a live audio source that exceed 20,000 Hertz, and while technically beyond our response range, those frequencies may further interact to produce subtle colorations that are within our range. Therefore, even when digital audio files are stored as uncompressed “.WAV” files, they are unlikely to accurately reproduce all harmonic content present in the original signal. Since presently the highest “standard” is that of *Red Book* “CD quality,” we might conclude that even a professionally recorded compact disk may not contain all the harmonic nuances present in the live signal.

It is this logic that has helped drive digital audio technology to ever-higher sampling and bit rates (e.g. 96 KHz at 32 bits). And yet, no matter the technical limits of the digital process, it will always be a binary reconstruction of an analog event.

Furthermore, even if original content material is recorded at those highest sampling and bit rates, and stored in a format that preserves all the inherent frequency content, at some point that

material will be probably be distributed to the consumer. And then the trouble begins. As discussed above, much of the audio content that is intended for consumer use is further compromised when more aggressive compression schemes are applied. Compression formats applied to MP3 files employ algorithms that take advantage of the ear's tendency to mask certain sounds. For example, *temporal masking* occurs when a very loud is sound is heard immediately before or after a much quieter sound, say within .1 second. The auditory system will not respond to the quieter sound; it is not "heard." (Skoglund, 2000) Curiously, the same phenomenon will cause the ear to also mask a quieter sound that occurs before a louder event.

Another auditory phenomenon known as *frequency masking* causes certain frequencies to be heard well, while others are masked. Audiologists have well documented that the human ear acts as a comb filter, tuning out high and low frequencies, while making the mid-ranges easiest to hear. (Skoglund, 2000) Again, the compression algorithms for .MP3 files incorporate routines that recognize and simply eliminate frequencies that are less likely to be picked up by the human auditory system. Of course the first content to be purged will be the highest frequencies, followed by low frequencies, according to the parameters set by the software's operator.

It is because so much of the audio content that is consumed by the general public is delivered via .MP3 files that I posit my premise that consumers have come to accept and expect audio content that is not of the highest fidelity. That is, one of the effects of the proliferation of digital technology is a compromise in quality resulting from the need to limit the size of the files for Internet delivery.

The literature of audio recording contains frequent examples of advice on file management schemes for Internet delivery which involve limiting sampling rates to as low as 16,000 Hertz. This limits the high frequency response of those files to a mere 8000 Hertz!

A generation of younger audio consumers is unaware that better quality is possible because the most common method for the acquisition of audio content is the Internet. While empirical research on the topic is only beginning to appear, anecdotal evidence abounds that younger consumers are quite willing to accept lower quality recordings. The mere fact that the .MP3 file format is becoming the de facto standard for delivery and consumption of music should indicate that today's audiences are not the critical listening audiophiles of the late twentieth century.

It is well documented in the trade publications of the music and entertainment industries that downloading has far eclipsed all other playback media as the preferred method for today's consumers to acquire music. A.C. Nielsen, *Billboard*, and others have been tracking the steady decline of CD sales, and the associate rise in sales of downloaded music over the past decade. In February, 2011, *Digital Music News* reported on statistics gathered from the *Recording Industry Association of America* (RIAA, 2010), which showed that in 2010 approximately one billion singles were sold online. (Osorio, 2011) During the same year, nearly 400-million albums were sold online, which exceeded the number of CDs sold by nearly 100-million.

Note also that the statistics reported through the sources like *Billboard*, the *RIAA*, and *A.C. Nielsen* are only for legally downloaded music through "reporting" sources. It is also well documented that many "traditional" recording artists and the *RIAA* believe that illegal

downloading of music is rampant and may well account for nearly as much as 63% of all music downloaded in the U.S. (RIAA, 2011)

In fact, the importance of Internet distribution has become so great that the music industry and the industry press have had to evolve to accommodate it in other ways. Where once the *Billboard Hot 100* was the bellwether for popular music, today that is no longer the case. In November 2010, popular teen idol Justin Bieber's video posted on *YouTube* was the most viewed *YouTube* video ever. But at the same time, Bieber's song was not even among the top 100 popular songs on *Billboard's* charts. This trend has prompted *Billboard* to create the "Social 50" to track net play on sites like *YouTube*, *Facebook* and *Twitter*. (*Billboard.com*, 2011)

Some of us worry that the degradation of quality of audio content that is available to the consumer is an unfortunate trend. Will this make teaching the recording arts irrelevant and eventually obsolete? As educators, we would like to believe that we teach the skills and knowledge that will allow trained people to create higher quality media content than the untrained, and that higher quality content is of higher value. But we are now in a situation where untrained recordists are capturing, sharing and even selling audio content, much of which is below any previous standards of quality. Even though current technology allows a skilled operator to "correct" many problems in poor audio content, one could just as easily argue that proper training could have prevented many of those problems. Inadequate levels, excessive background noise, inconsistent levels and other troubles are easily fixed. It's just that often they are not.

Finally, we must also recognize that this is the “slippery slope” about which we hear so much. Like any artistic endeavor, what is “right” is always in the ear of the creator.

History suggests that most trends are cyclical. Political democracies come and go. It may be that the democratization of audio media may ultimately lead to its reformation in the form of a class system; an elitism. In this scenario, the technocrats will again produce the best content, and reap the rewards. It just may be that there was, still is, and will be some value in learning the art and science of recording, mixing and mastering.

Collins, M. (1997, January). *Compact disk formats explained*. Retrieved from <http://www.soundonsound.com/sos/jan98/articles/cdformats.htm>

Fee, W. R. . (1933). *The transition from aristocracy to democracy in new jersey, 1789-1829*. Somerville, N. J.: Somerset Press.

Kramnkik, Is. (2007, February 20). Compact disc [Web log message]. Retrieved from <http://red-book.blogspot.com/>

Miller, P. (1999). *An examination of the correlation between perceived sound quality and frequency response of current mpeg audio encoders*. Retrieved from <http://www.mp3-tech.org/tests/pm/Introduction.htm> Grey, K. (1997, May 3). *Producing great sounding phonograph records*. Retrieved from <http://www.recordtech.com/prodsounds.htm>

Ogorek, K. (2007). *Ithedemocratizationpublishingof*. Retrieved from <http://www.authorsolutions.com/uploadedFiles/Whitepapers/PowerToPublish.pdf>

Osorio, A. (2011, February 17). *The history of recording industry sales, 1973-2010*. . Retrieved from <http://www.digitalmusicnews.com/stories/021711disruption>

RIAA, . (2010). *Frequently asked questions*. Retrieved from <http://www.riaa.com/faq.php>

Skoglund, J., & Kleijn, W.B. (2000). On time-frequency masking in voiced speech. *Speech and Audio Processing, IEEE Transactions*, 8(4), 361-369.

Social 50. (2011, April 23). Retrieved from <http://www.billboard.com/#/charts/social-50>

The c20 democratisation of culture. (n.d.). Retrieved from <http://www.architecture.org.uk/ad1.htm>

Tascam 38. (2008). Retrieved from http://www.hifi-museum.com/js/teac/teac_fs.htm

Tascam professional/teac. (2008). Retrieved from <http://tascam.com/Company.html>

Windows is a registered trademark of Microsoft Corporation in the United States and other countries.