ANALYSIS OF ZERO-LEVEL SAMPLE PADDING OF VORBIS AND OPUS ENCODERS FOR THE .OGG FILE CONTAINER

by

MAX BOTHE

B.S., University of Colorado, 2018

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Science

Recording Arts Program

2020

This thesis for the Master of Science degree by

Max Bothe

has been approved for the

Recording Arts Program

by

Catalin Grigoras, Chair

Gregory Wales

Cole Whitecotton

Date: December 12, 2020

Bothe, Max (M.S., Recording Arts Program)

Analysis of Zero-Level Sample Padding of Vorbis and Opus Encoders for the .OGG File Container

Thesis directed by Associate Professor Catalin Grigoras

ABSTRACT

Zero Level Samples can be added to the beginning and/or end of audio files by the encoder being used and are read as silence in the audio information streams. These samples can be added into audio files for several different reasons that revolve around giving the encoding algorithm time to process the information that is to be encoded. With each subsequent generation of compression, it is expected that the number of Zeroes added to the beginning and end of each audio file would change. The purpose of this study is to observe how different audio encoding programs pad these zero-level samples into audio files, and to see if the different encoders have an observable pattern between each generation of compression.

The form and content of this abstract are approved. I recommend its publication.

Approved: Catalin Grigoras

ACKNOWLEDGEMENTS

I would like to thank Catalin Grigoras, Jeff Smith, Cole Whitecotton, and Leah Haloin for the patience in helping me get through this program.

TABLE OF CONTENTS

CHAPTER

CHAPTER I

INTRODUCTION

A Brief History of the OGG Container; Vorbis and Opus Encoders

The OGG container format scheme is maintained by the Xiph.Org Foundation. Created in 1993 and originally called OGGSquish, the Xiph.org Foundation wanted to create a compressed audio format that would work for modern audio applications, as well create open-source alternatives to proprietary technology such as MP3. As of 2009, the OGG container format became the umbrella term for all multimedia codecs for the Xiph.Org Foundation[6].

Vorbis compression was created as an open-source alternative to MP3. Officially released in 2002 by the Xiph.Org Foundation to be used with their OGG container format, Vorbis lacked the mainstream backing of other compression schemes such as MP3 but was popularized by developers for being completely open source, as well as allowing a variable bit rate, versus MP3's constant bit rate, allowing for smaller files sizes[4]. Vorbis is currently used by websites such as Bandcamp and Wikipedia, as well as streaming services such as Spotify.

Opus was released in 2013 and was a compression originally created for Voice-over Internet Protocol (VoIP). Now, Opus is used for both audio and video compression and is used in a variety of audio and video streaming products such as WhatsApp, Playstation 4 voice communication, and SteamOS Video/Audio streaming[5].

Zero-Level Sample Padding

Part of the process of coding and decoding lossy compression formats can be to pad new files with zero-level samples (ZLS). ZLS are read as absolute silence in the audio information stream. The number of zeroes added or subtracted from the beginning or end of each audio file can vary depending on the codec used, as well as what audio program was used to encode them.

The number of ZLS can also vary between the left and right channels. Previous research done by Schroeder and Boehm on the MP3 codec tell us a few reasons why these zero-level samples can be introduced. Digital audio is processed in blocks of samples and, because of this, codec algorithms need to create a buffer before work can be done on the signal. Frequency-domain processing can also cause a delay in the signal processing, since all signals must pass through a filter-bank which can be implemented in various ways. The third reason for this delay is due to some encoders needing look ahead time for their algorithms to decide on how to best tackle the information ahead[3].

Purpose of this Study

The purpose of this study is to observe whether the number of zeroes padded by different encoders, in this case Vorbis or Opus, will produce a pattern that is measurable between subsequent generations of compression. This study will observe the ZLS at the beginning and end of both the left and right channels for each generation of compression across multiple recorders, using multiple different audio programs. If a pattern of ZLS can be observed between generations of recordings, it may help create new ways of authenticating audio recordings.

CHAPTER II

MATERIALS AND METHODS

This study took five different common handheld recorders, each producing two original recordings, and compressing those recordings for four generations of new audio files using multiple different encoders and programs. The devices used were:

Table 1: Hardware Used

Hardware	MP3 Settings	WAV Settings
American Recorder	128kbs, 16-bit, 44kHz	16-bit, 44kHz
Olympus 700M	256kbs, 16-bit, 44kHz	16-bit, 44kHz
Olympus 750M	128kbs, 16-bit, 44kHz	16-bit, 44kHz
Sony UX512	128kbs, 16-bit, 44kHz	16-bit, 44kHz
Tascam GTR1	128kbs, 16-bit, 44kHz	16-bit, 44kHz

Each recording device produced two files, one MP3 file and one WAV file. All MP3 files were produced at 128kbs, 16-bit, 44kHz with the exception of the Olympus 700M which was recorded at 256kbs. All WAV files were 16-bit, 44kHz. All files were recorded in stereo.

It is known that the loudness of the file can affect the ZLS padding done by encoders. For this study, all recordings were done at low volume with either ambient noise or talking[7].

The process of creating each generation was as follows: The original recording was loaded into an audio program (such as Adobe Audition) and a first-generation compression was made. This first-generation file was then loaded back into the program and compressed again. This was done until there were four generations of compression.

Compressed files were made into 16-bit, 128kbs .OGG files apart from files compressed using FFmpeg, which were compressed at 500kbs, and NCH Vorbis files which were compressed at 327kbs (variable).

The following programs and encoders were used for the making of each generation of audio file:

Table 2: Programs Used

Program	Codec(s) Used
Adobe Audition 2020	Vorbis
FFMPEG	libopus 1.3
iZotope Rx7	Vorbis
NCH Switch Plus v8.06	Vorbis and Opus
dBPoweramp v16.6 (64-bit)	Vorbis and Opus

Forensic Audio Analysis System (FAAS) created by Catalin Grigoras, PhD., was used to measure the ZLS of each file. Each file was first created into a .WAV file and split into left and right channels by FAAS in order to perform the analysis. Below is the initial ZLS count for the original recordings.

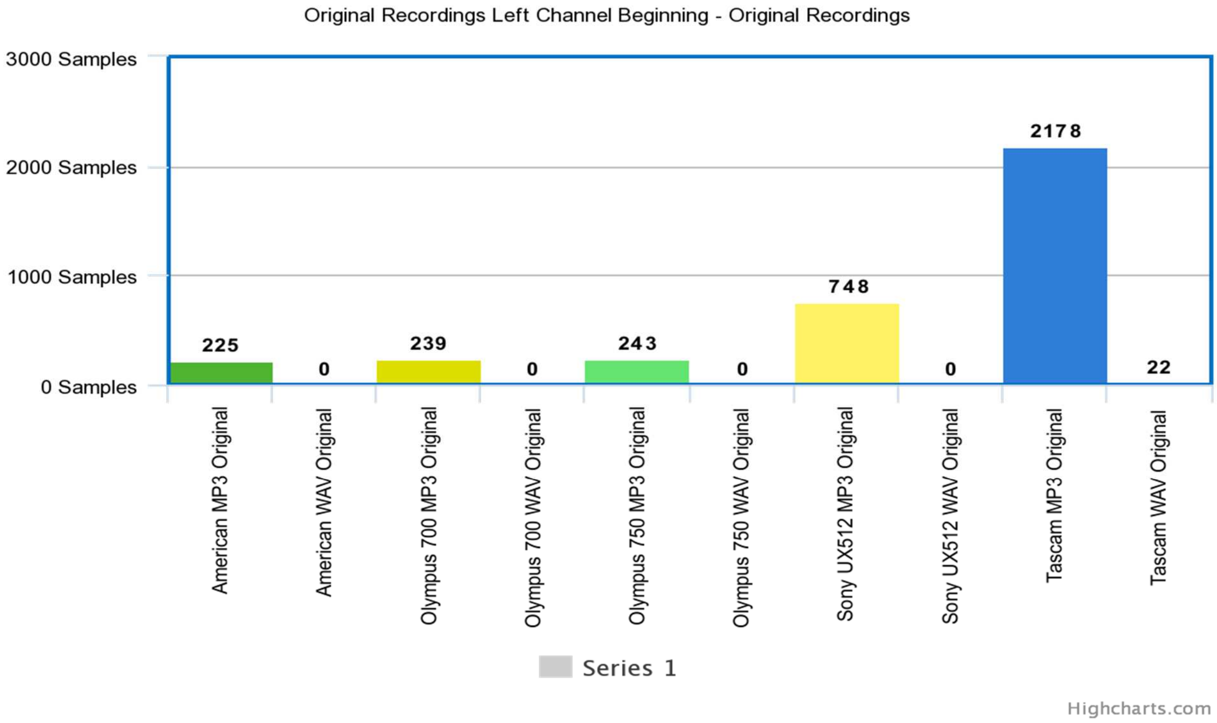

Figure 2.1: Original Recordings ZLS—Left Channel Beginning

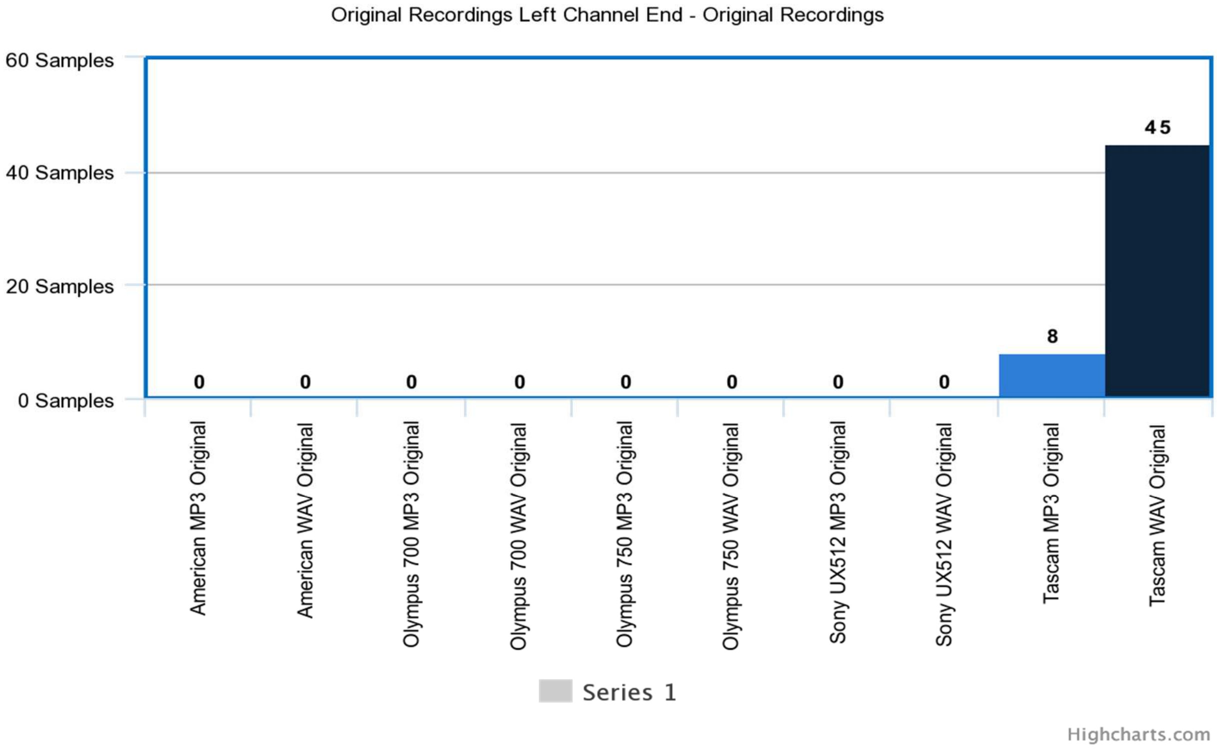

Figure 2.2: Original Recordings ZLS—Left Channel End

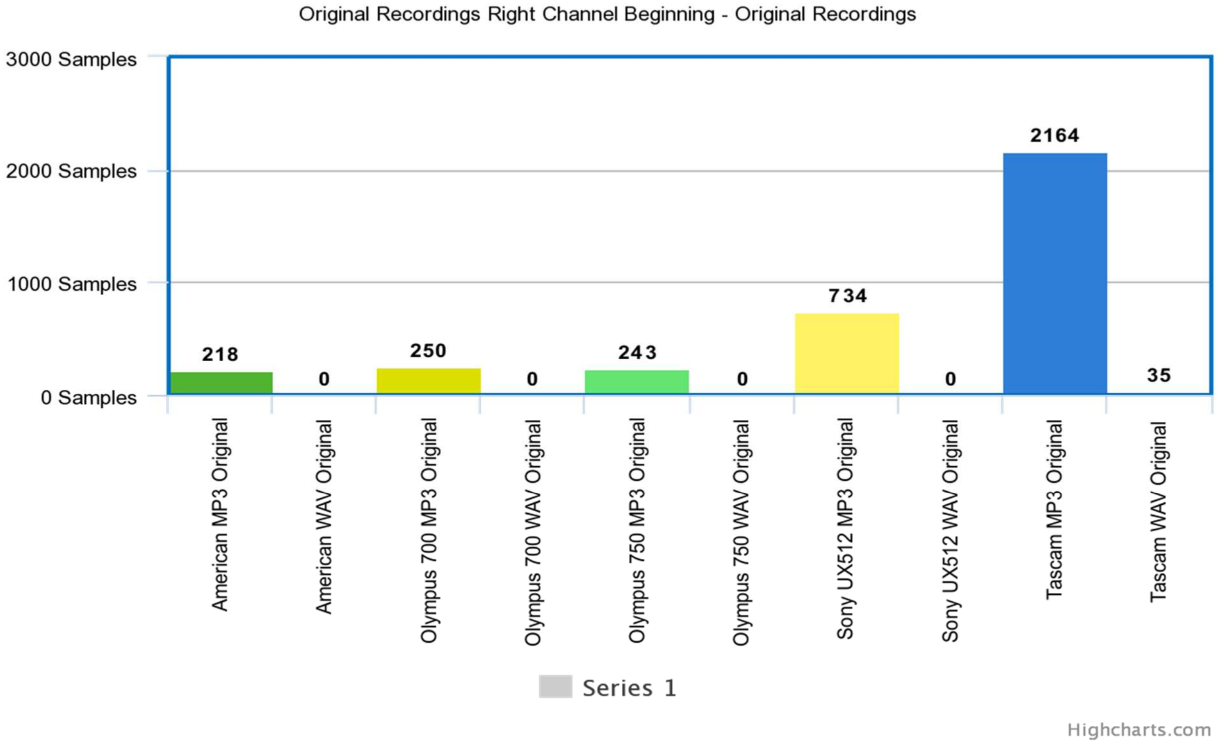

Figure 2.3: Original Recordings ZLS—Right Channel Beginning

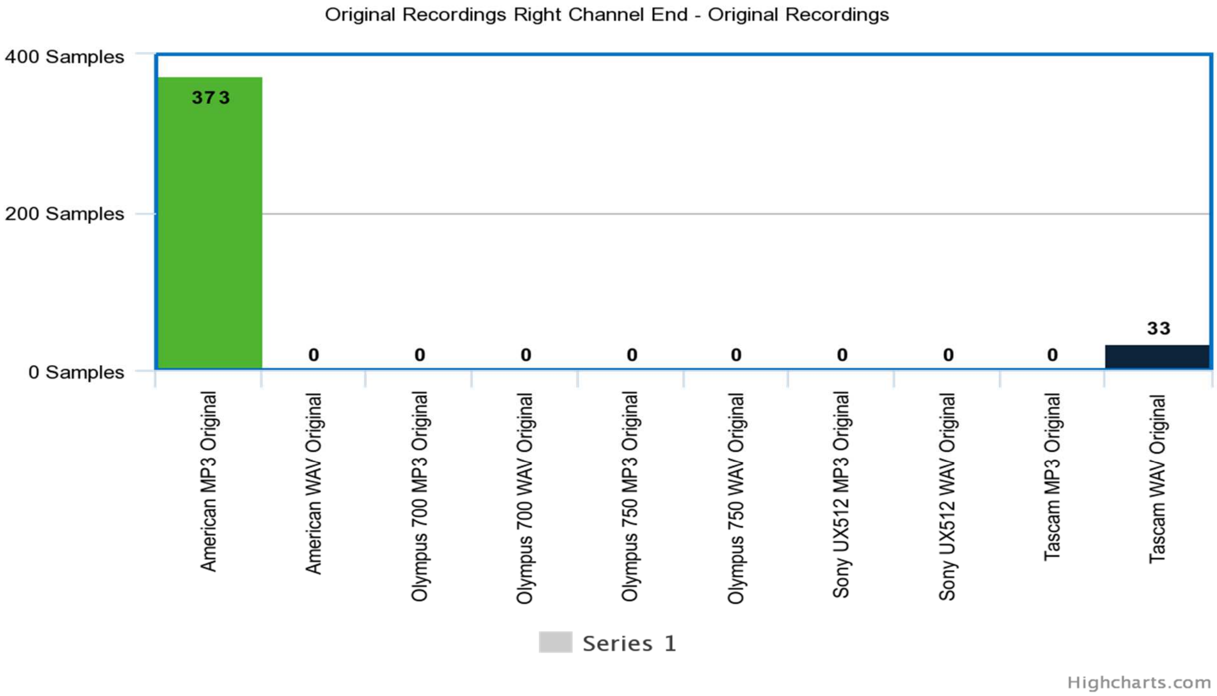

Figure 2.4: Original Recordings ZLS—Right Channel End

We can see with the original files that nearly all of the MP3 recordings have most of the zero-level samples starting at the front of the audio recording, while almost all of the WAV recordings have no zero-level samples. The Tascam was the only WAV recording to produce any zero-level samples, and even then, it was still very few compared to its MP3 recording.

As previous research as suggested, we would expect some zero-level samples from MP3 recordings.

CHAPTER III

RESULTS

Each histogram will show either the beginning or end of a recording, as well as being separated by the left and right channels for each recording. This will allow us to observe how each part of each file was affected by individual programs and codecs and compare each audio file and recording side-by-side. Each device and recording will be displayed with each generation (G) starting with the first to the fourth.

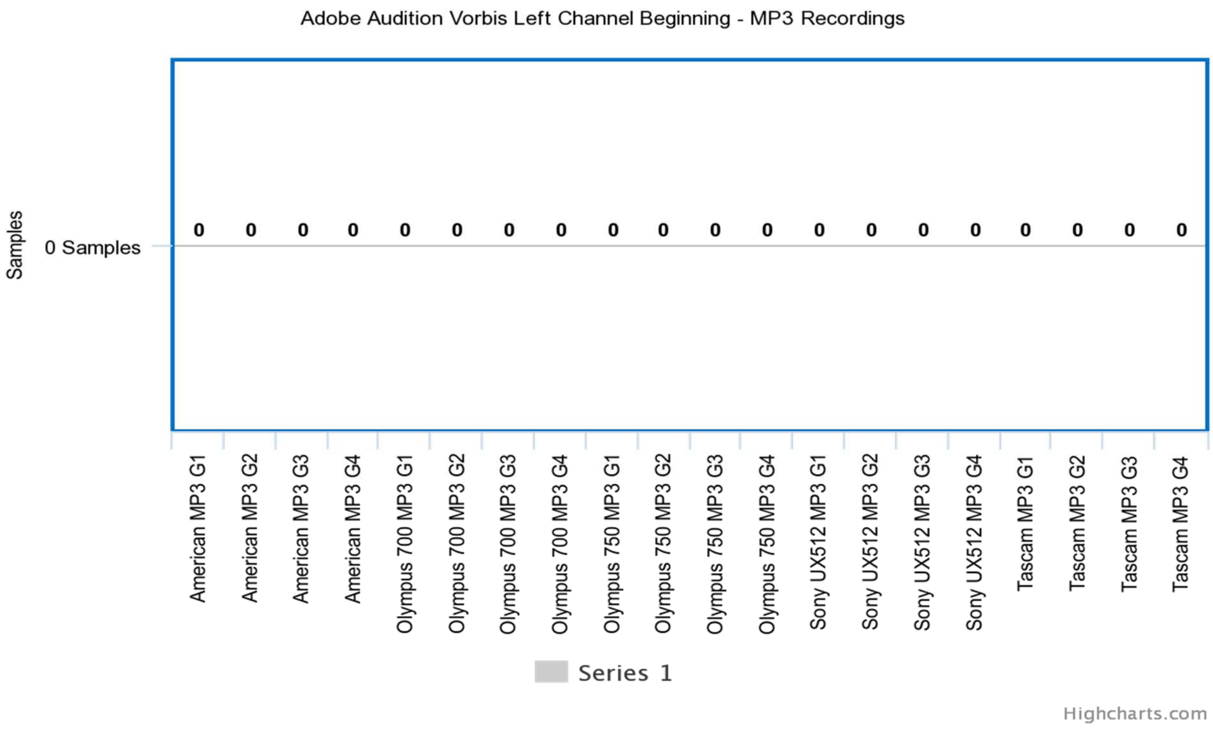

Figure 3.1: Adobe Audition MP3 Recordings—Left Channel Beginning

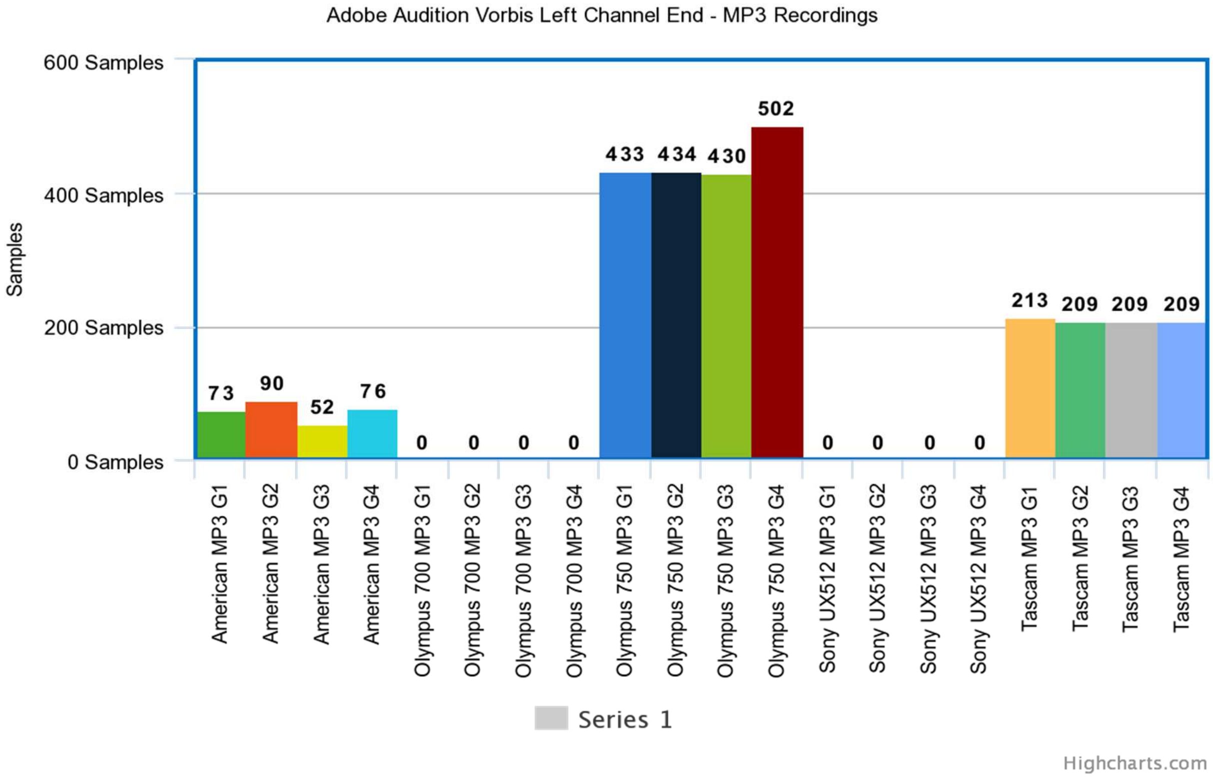

Figure 3.2: Adobe Audition MP3 Recordings—Left Channel End

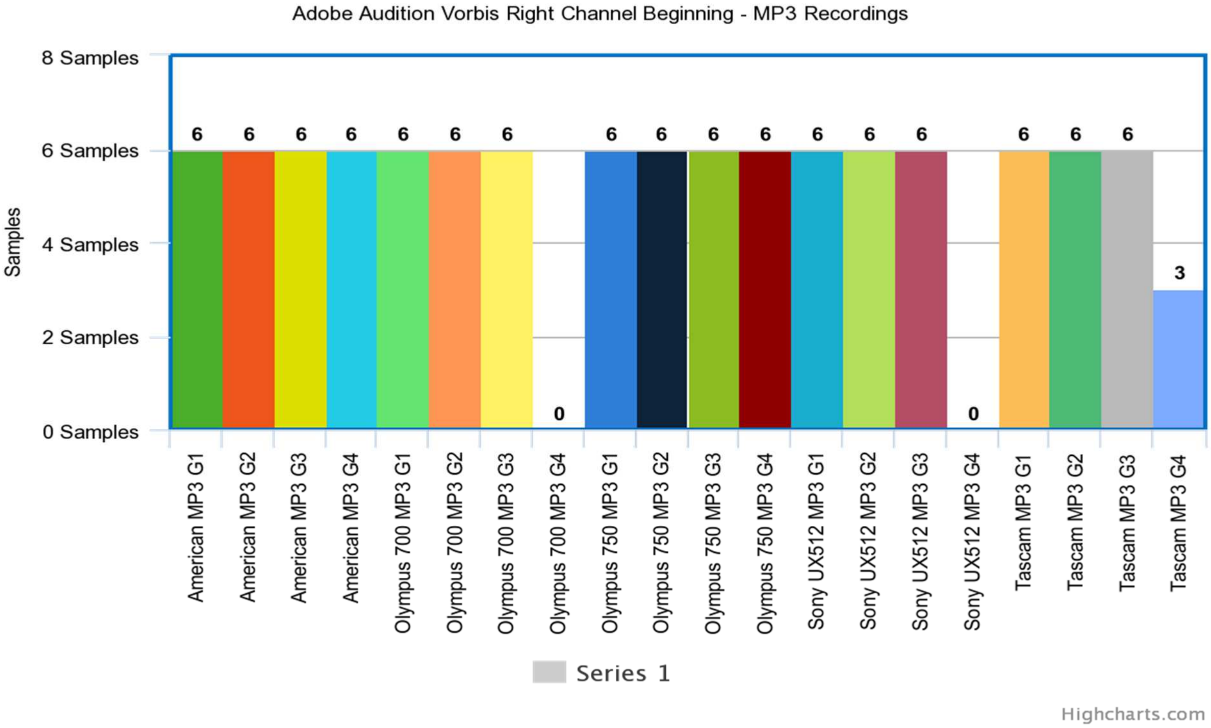

Figure 3.3: Adobe Audition MP3 Recordings—Right Channel Beginning

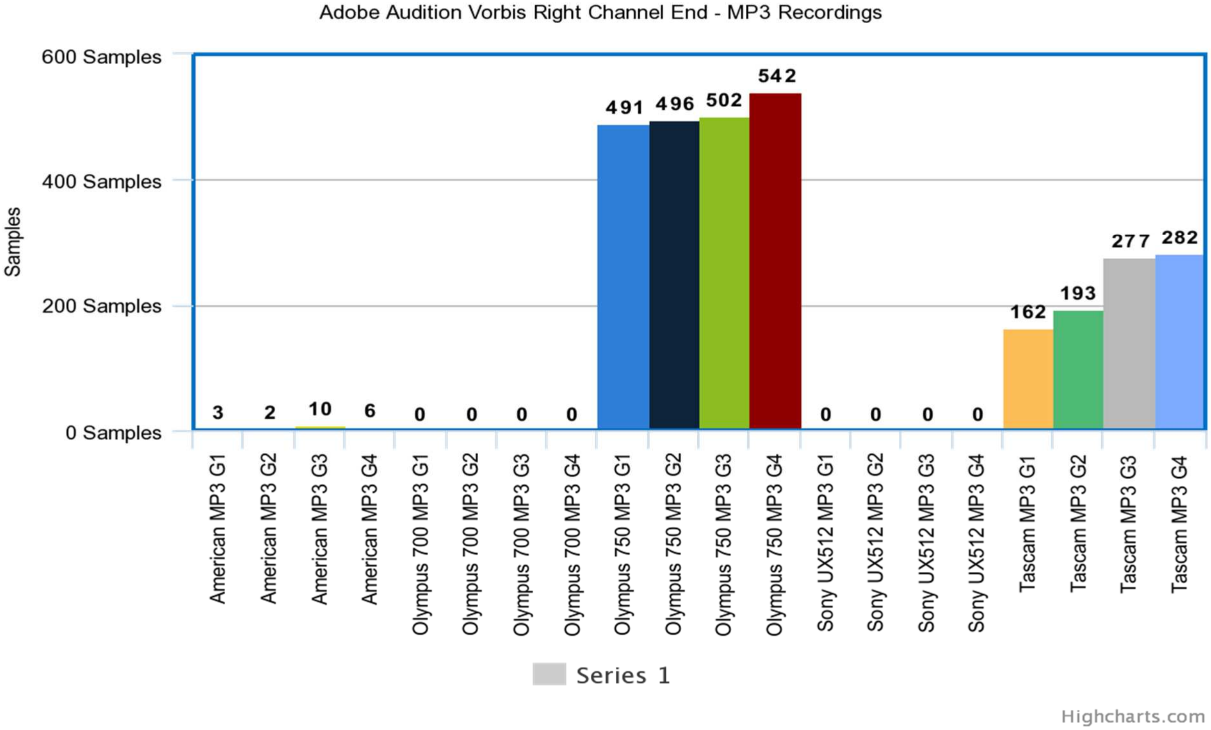

Figure 3.4: Adobe Audition MP3 Recordings—Right Channel End

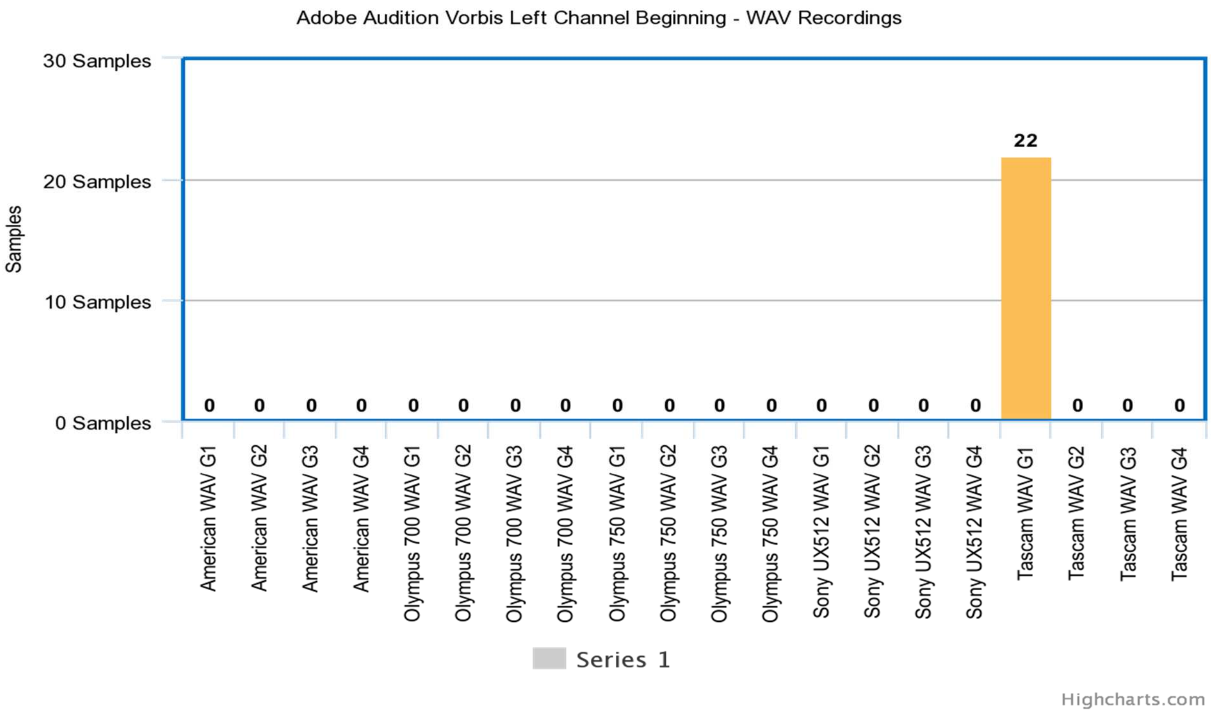

Figure 3.5: Adobe Audition WAV Recordings—Left Channel Beginning

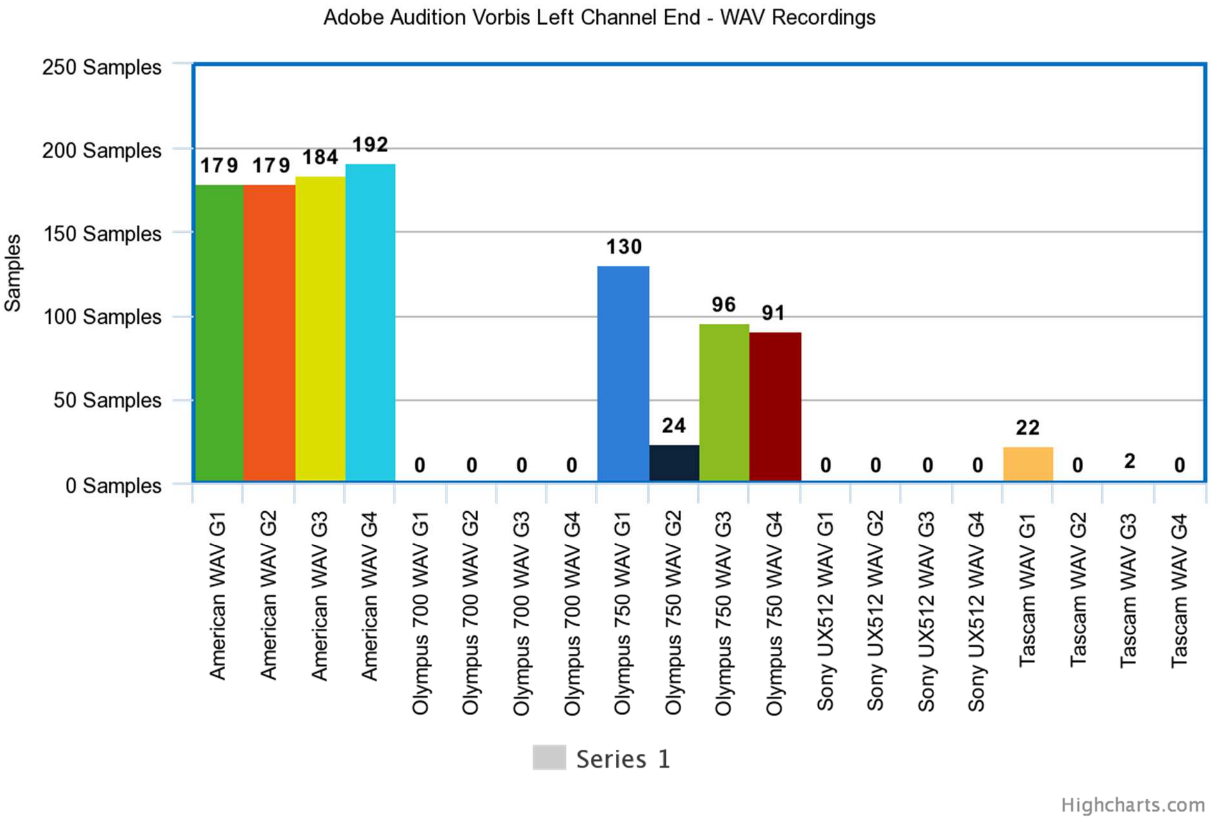

Figure 3.6: Adobe Audition WAV Recordings—Left Channel End

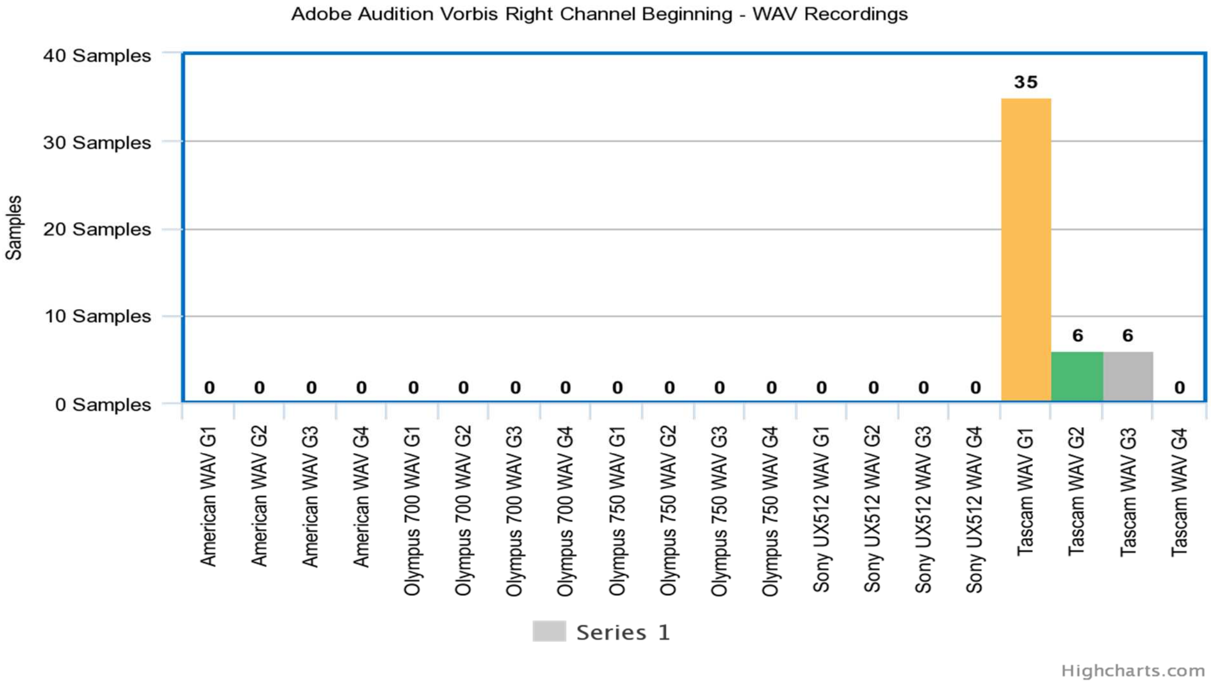

Figure 3.7: Adobe Audition WAV Recordings—Right Channel Beginning

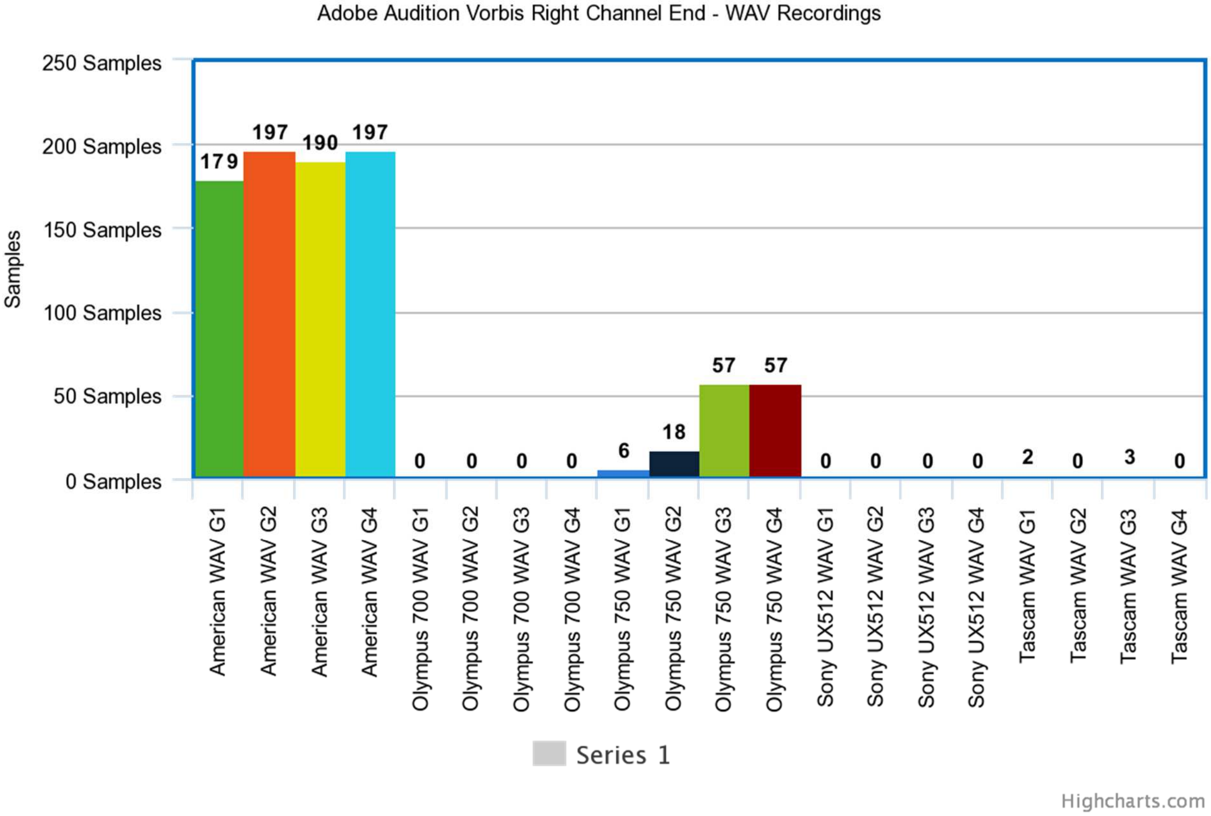

Figure 3.8: Adobe Audition WAV Recordings—Right Channel End

With Adobe Audition 2020 using Vorbis encoding, we can see that with the MP3 recordings, the beginning of the left channel had eliminated any ZLS with a significant reduction in the beginning of the right channel. Interesting to see that the beginning of the right channel had almost all MP3 recordings reduced to 6 ZLS as well. Both the left and right channels saw a significant rise in ZLS at the end of their recordings for the Tascam and Olympus 750 recorders.

For the WAV recordings, we see that most of the recordings still have no ZLS at the beginning of both the left and right channels, with a significant number added at the end in the American Recorder and Olympus 750 recorders.

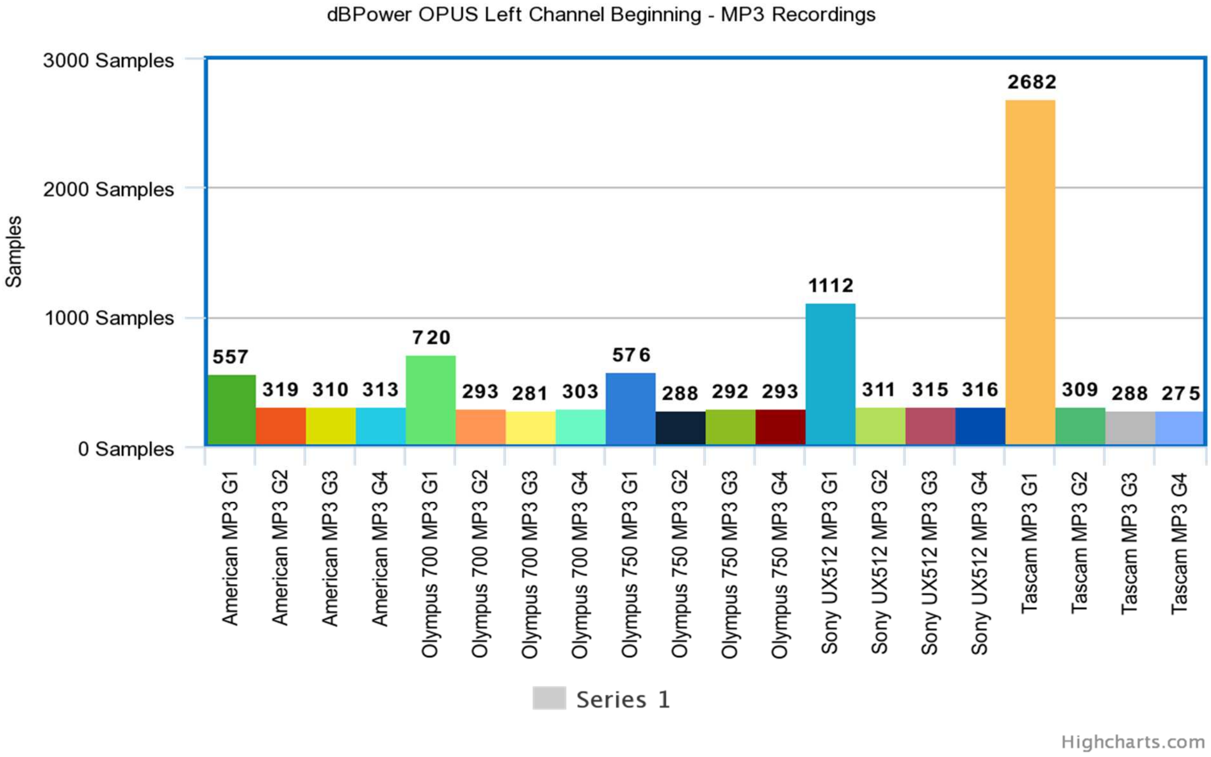

Figure 3.9: dBPower (Opus) MP3 Recordings—Left Channel Beginning

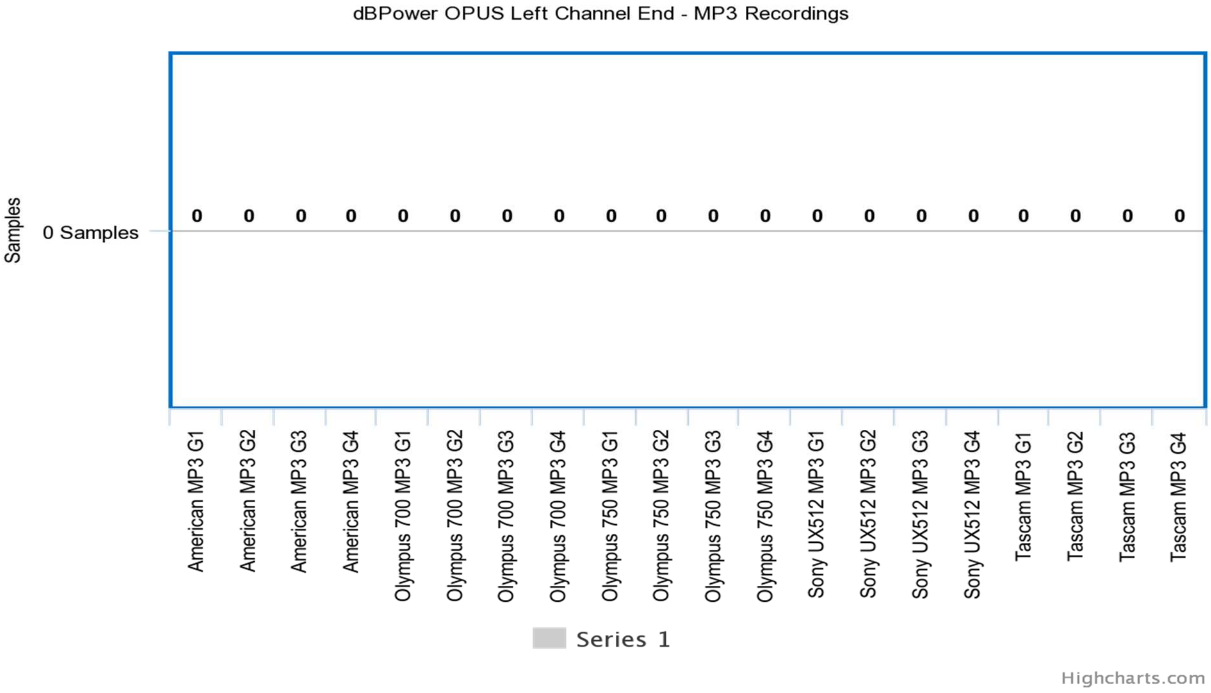

Figure 3.10: dBPower (Opus) MP3 Recordings—Left Channel End

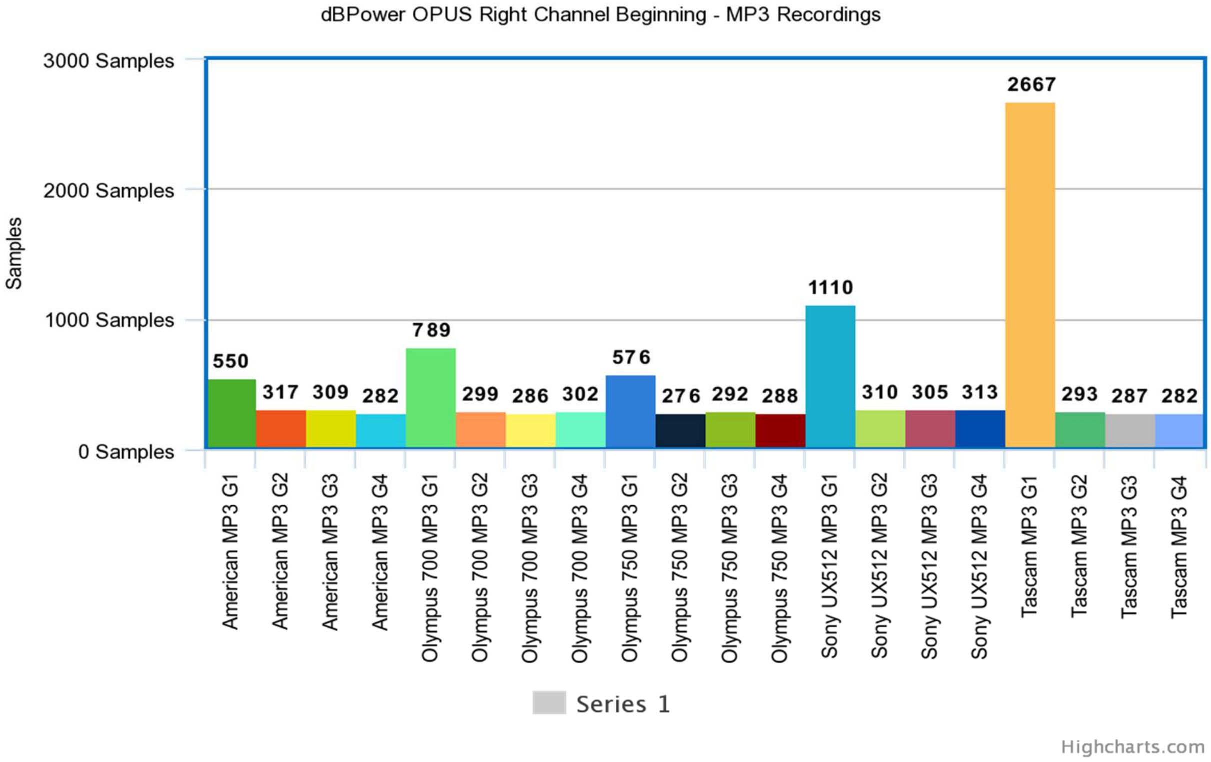

Figure 3.11: dBPower (Opus) MP3 Recordings—Right Channel Beginning

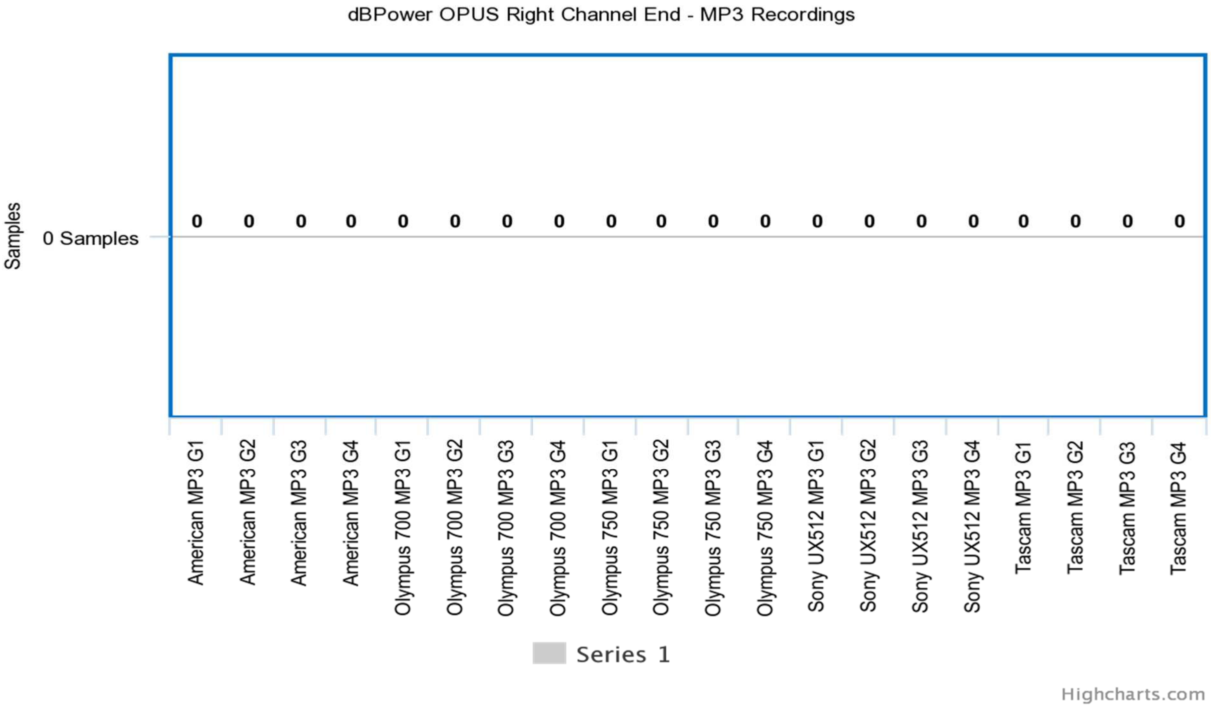

Figure 3.12: dBPower (Opus) MP3 Recordings—Right Channel End

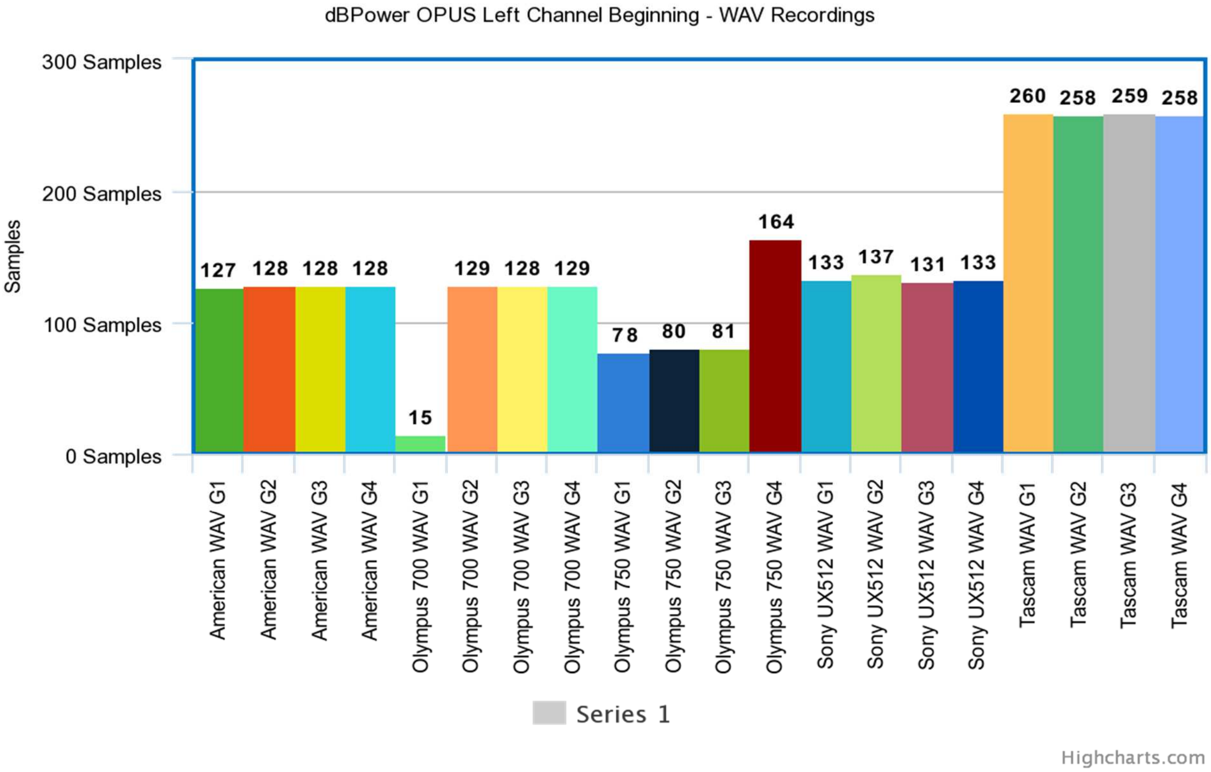

Figure 3.13: dBPower (Opus) WAV Recordings—Left Channel Beginning

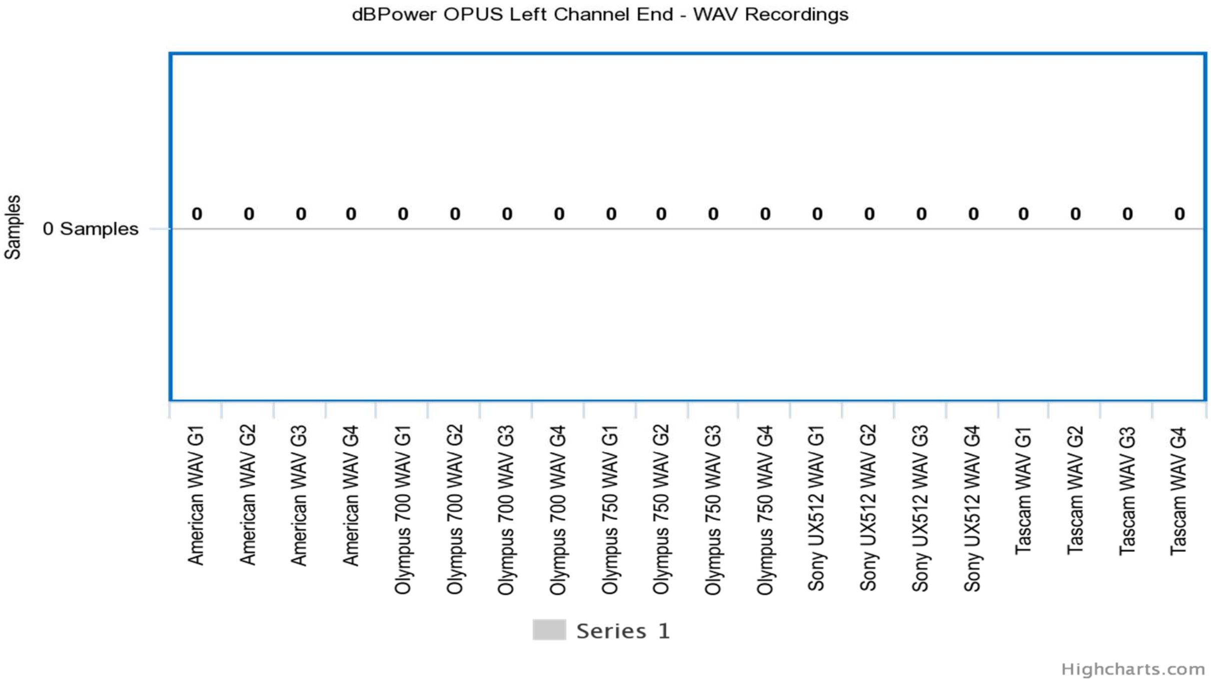

Figure 3.14: dBPower (Opus) WAV Recordings—Left Channel End

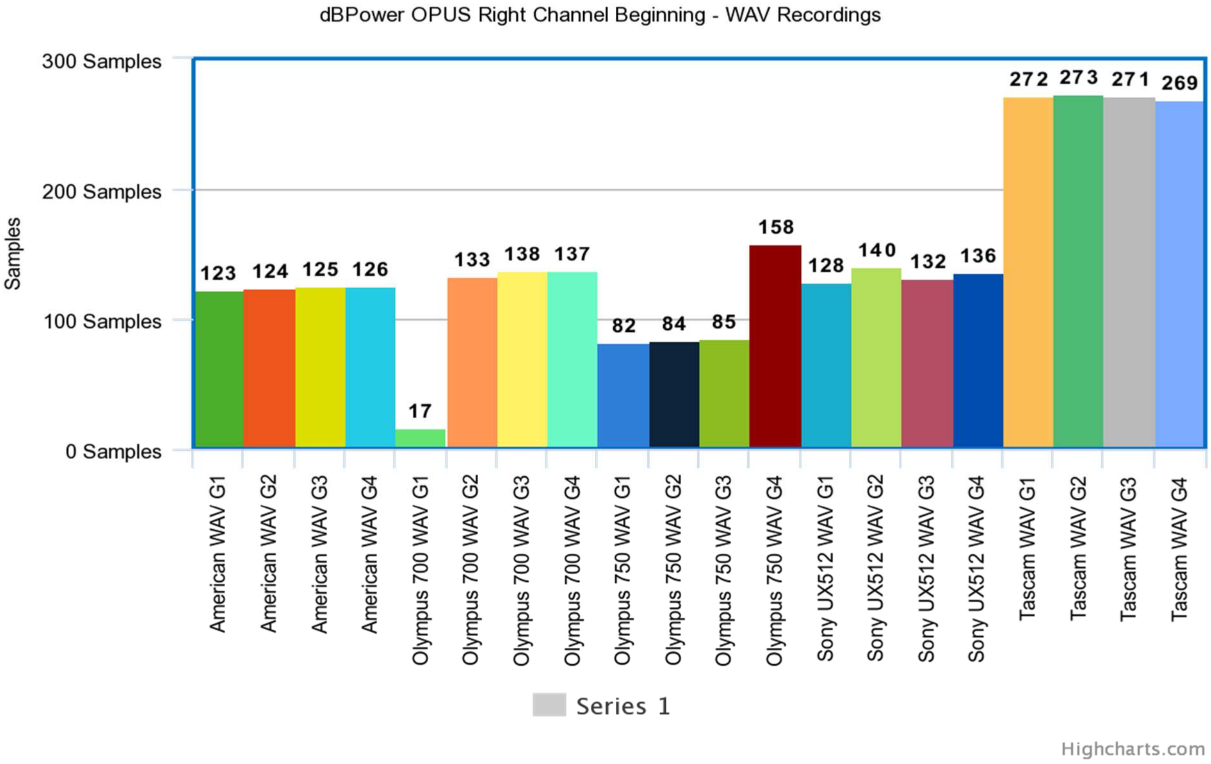

Figure 3.15: dBPower (Opus) WAV Recordings—Right Channel Beginning

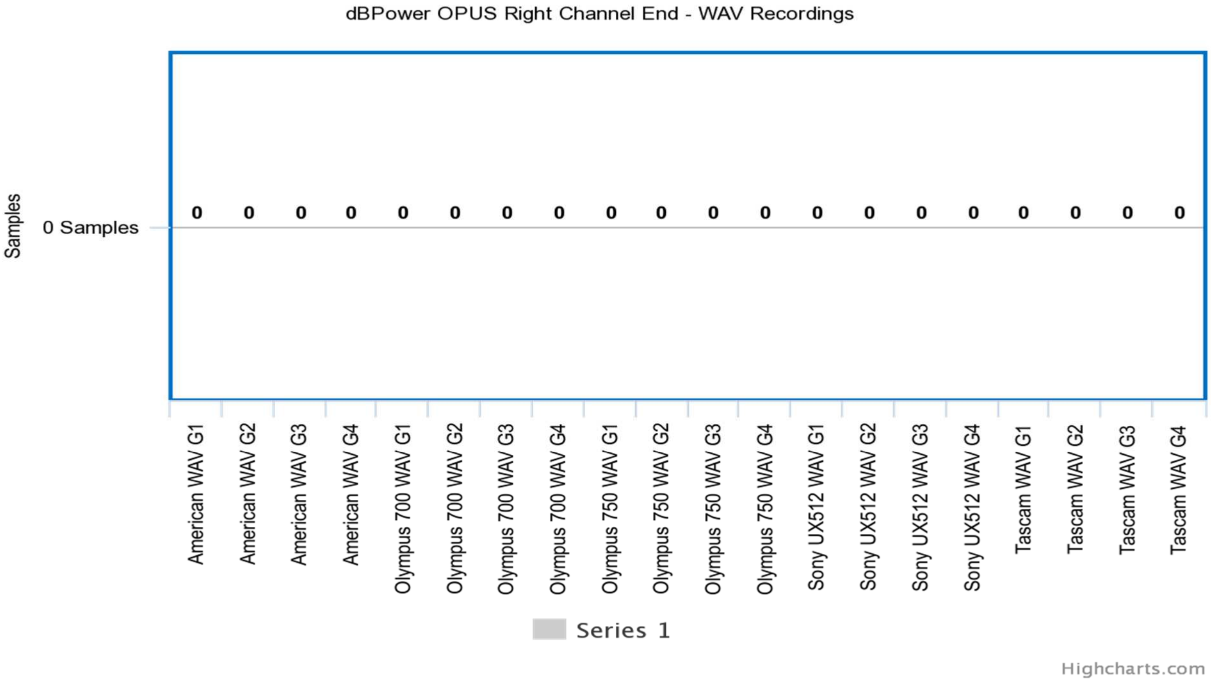

Figure 3.16: dBPower (Opus) WAV Recordings—Right Channel End

With dBPower using the Opus encoding, the MP3 recordings saw an initial increase of ZLS in the first generation of compression at the beginning of both the left and right channels. This quickly dropped to around 300 samples at the second generation and changed only slightly until the fourth generation. The WAV recordings did have an increase in samples from their original, however the changes between each generation tended to vary only slightly.

Both the MP3 and WAV recordings did not have any samples added to the ends of either of their channels.

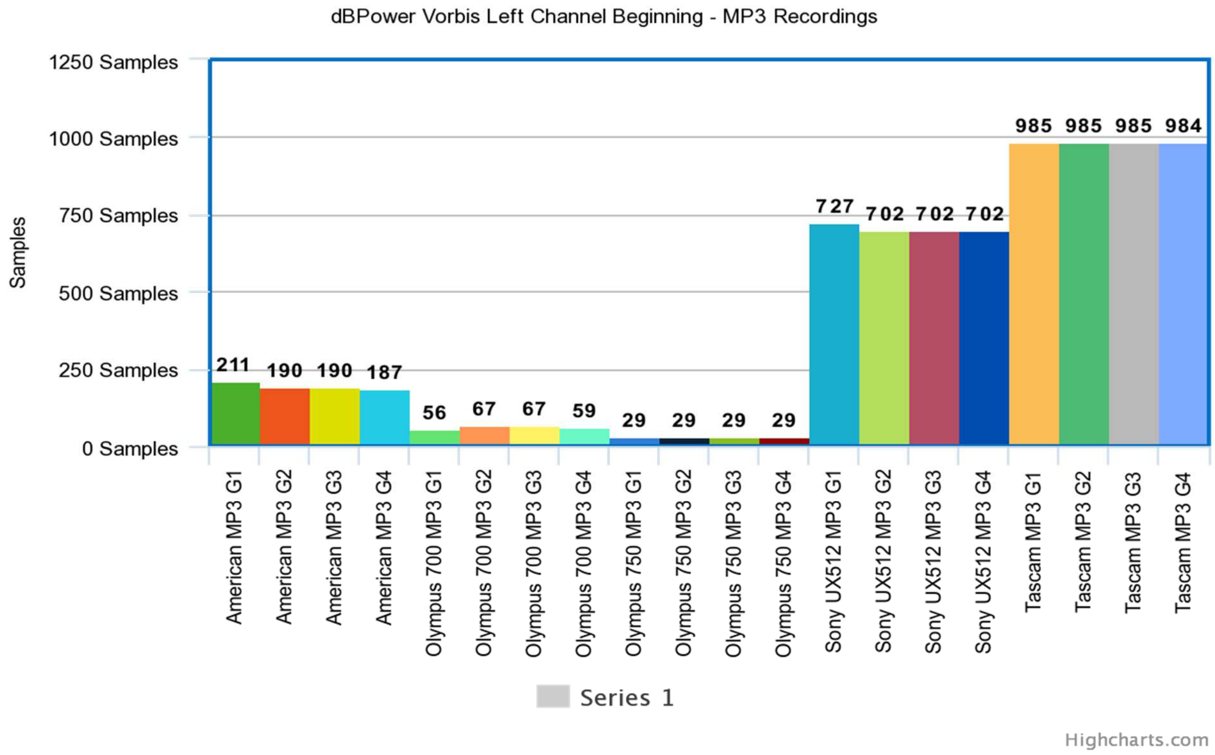

Figure 3.17: dBPower (Vorbis) MP3 Recordings—Left Channel Beginning

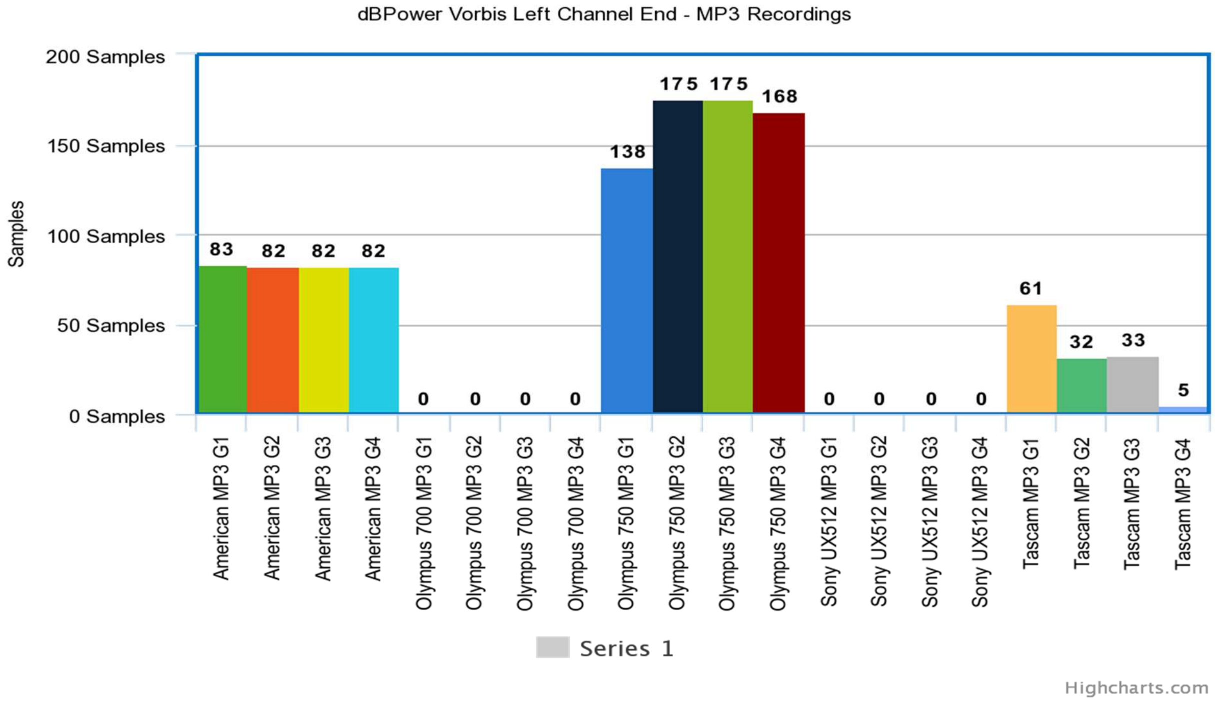

Figure 3.18: dBPower (Vorbis) MP3 Recordings—Left Channel End

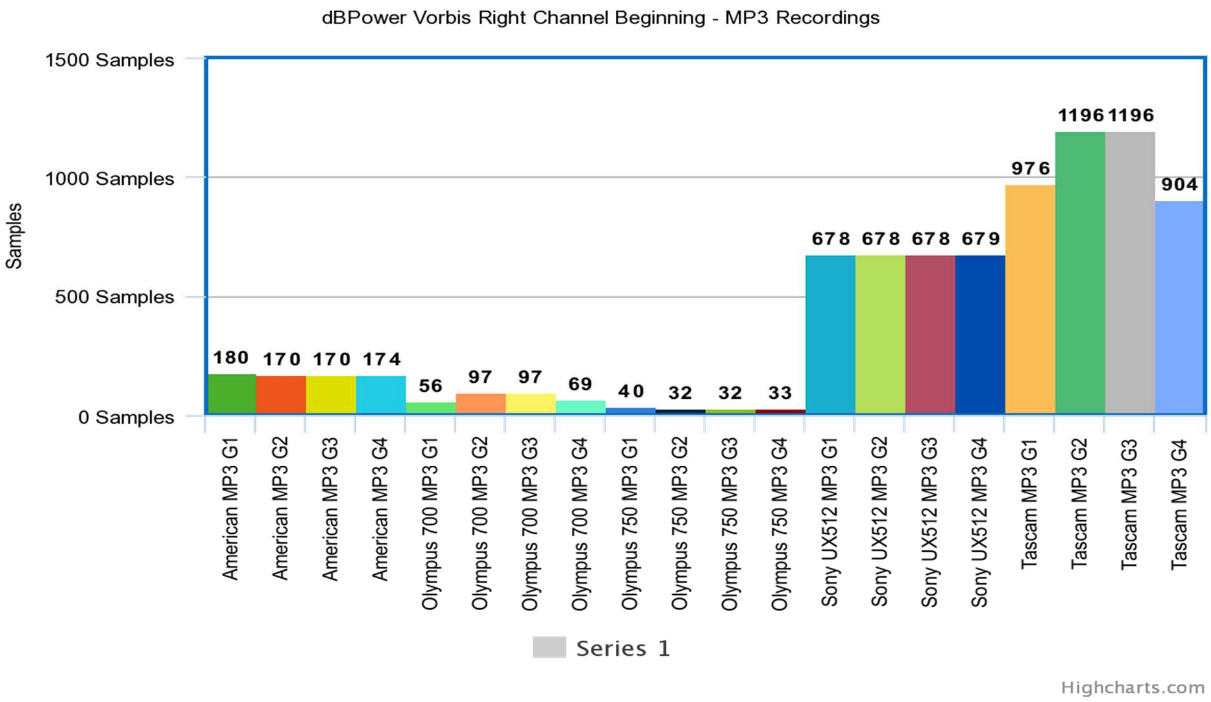

Figure 3.19: dBPower (Vorbis) MP3 Recordings—Right Channel Beginning

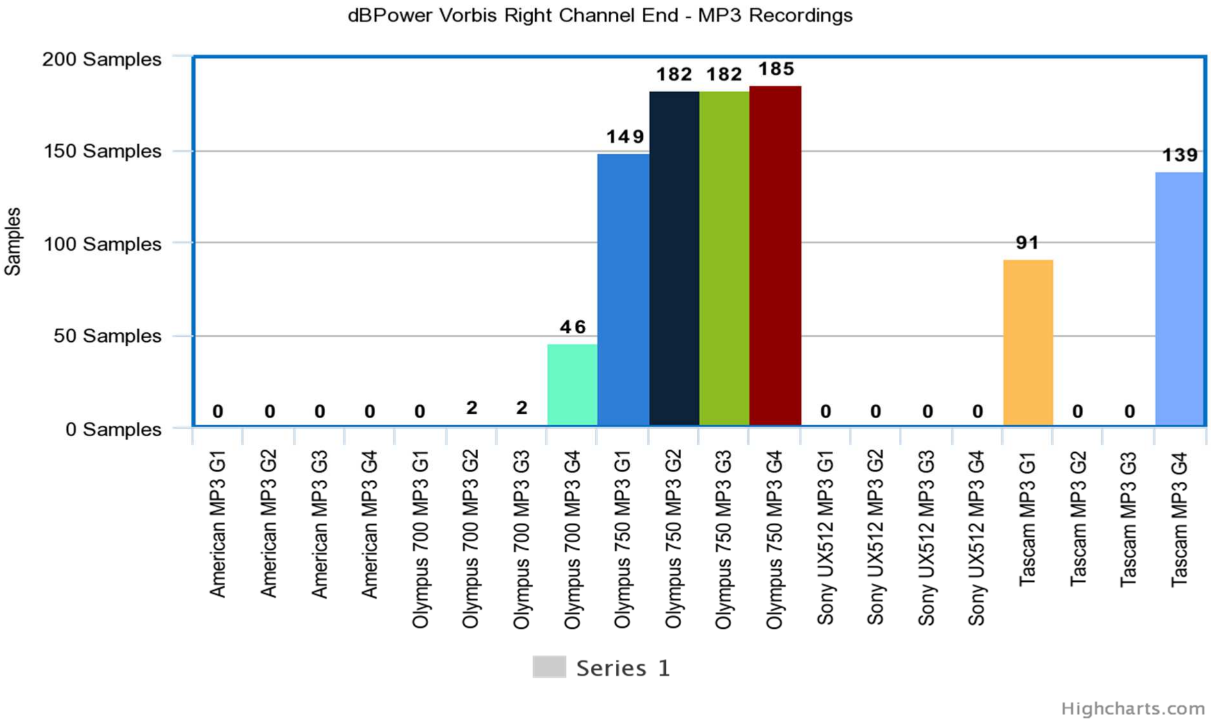

Figure 3.20: dBPower (Vorbis) MP3 Recordings—Right Channel End

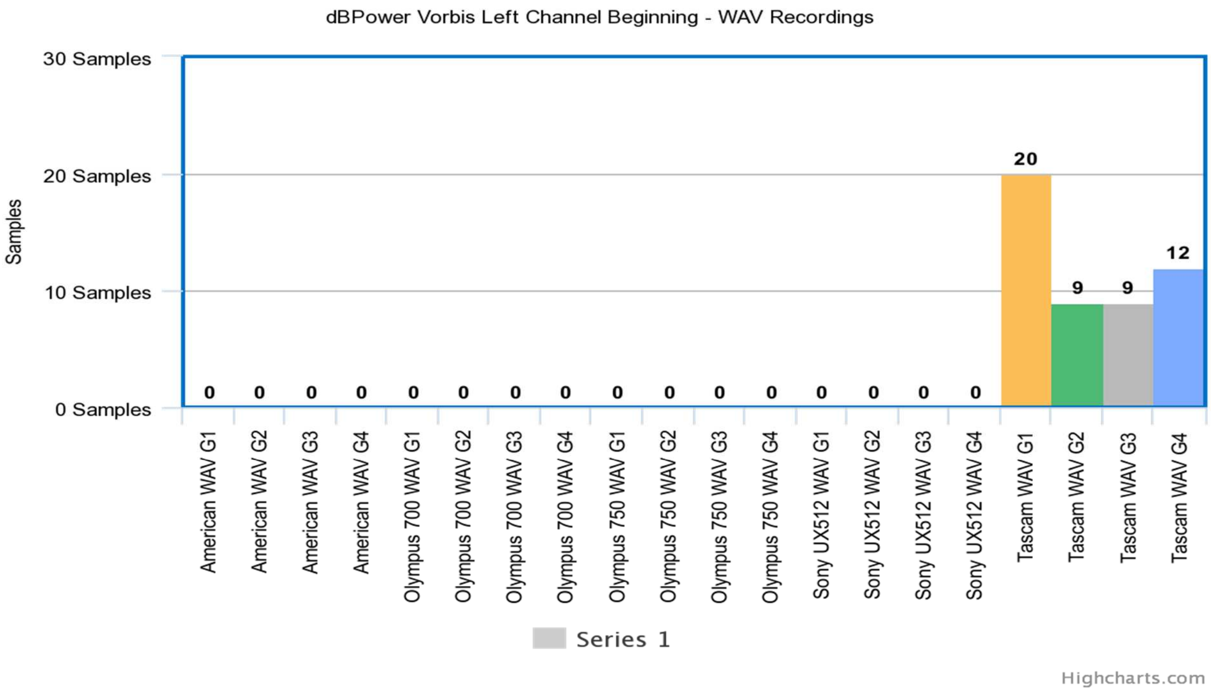

Figure 3.21: dBPower (Vorbis) WAV Recordings—Left Channel Beginning

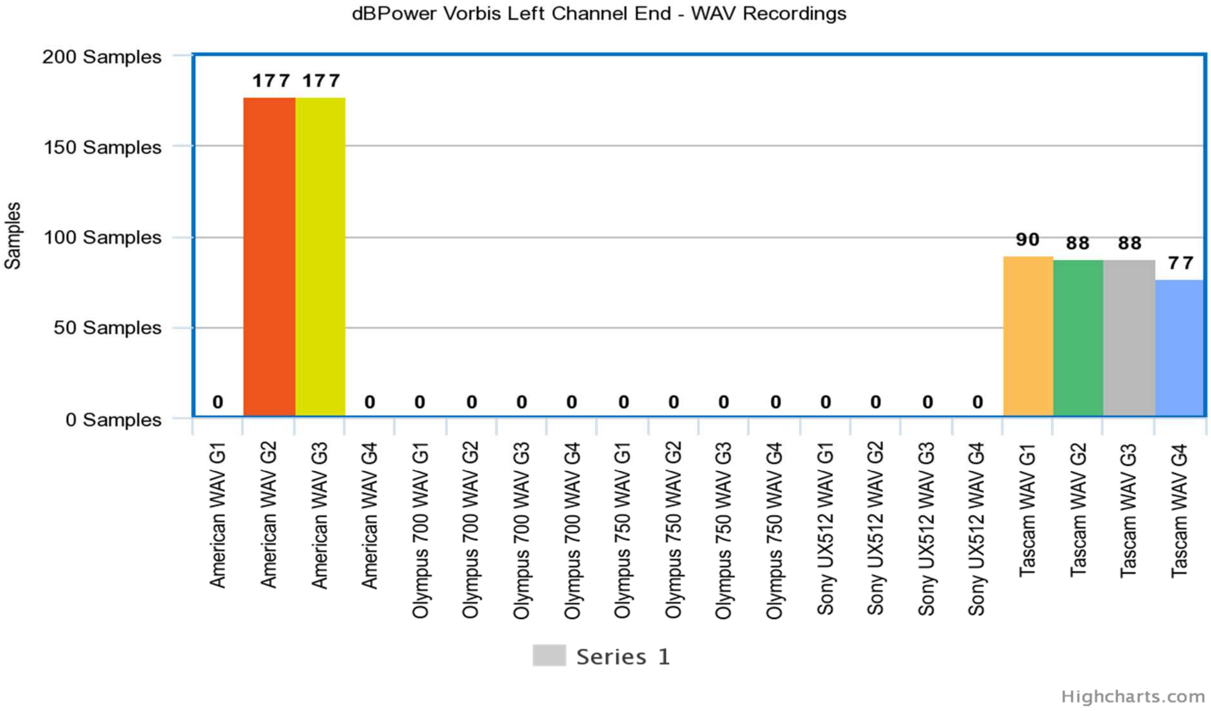

Figure 3.22: dBPower (Vorbis) WAV Recordings—Left Channel End

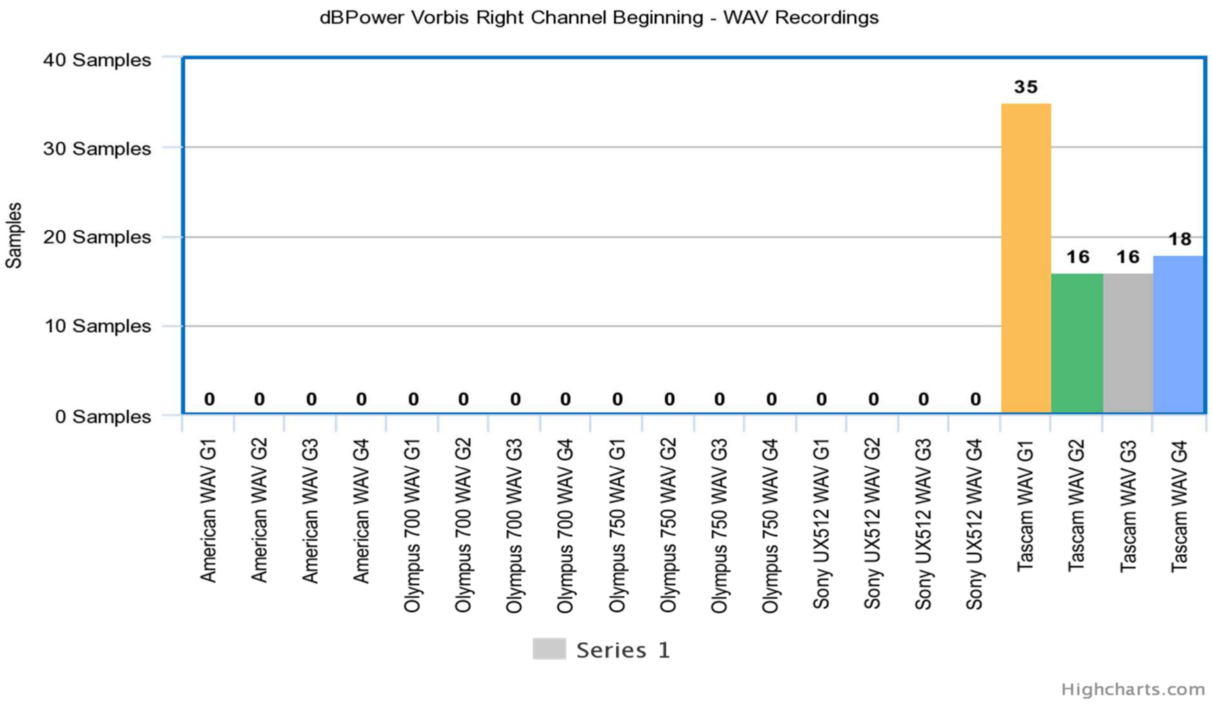

Figure 3.23: dBPower (Vorbis) WAV Recordings—Right Channel Beginning

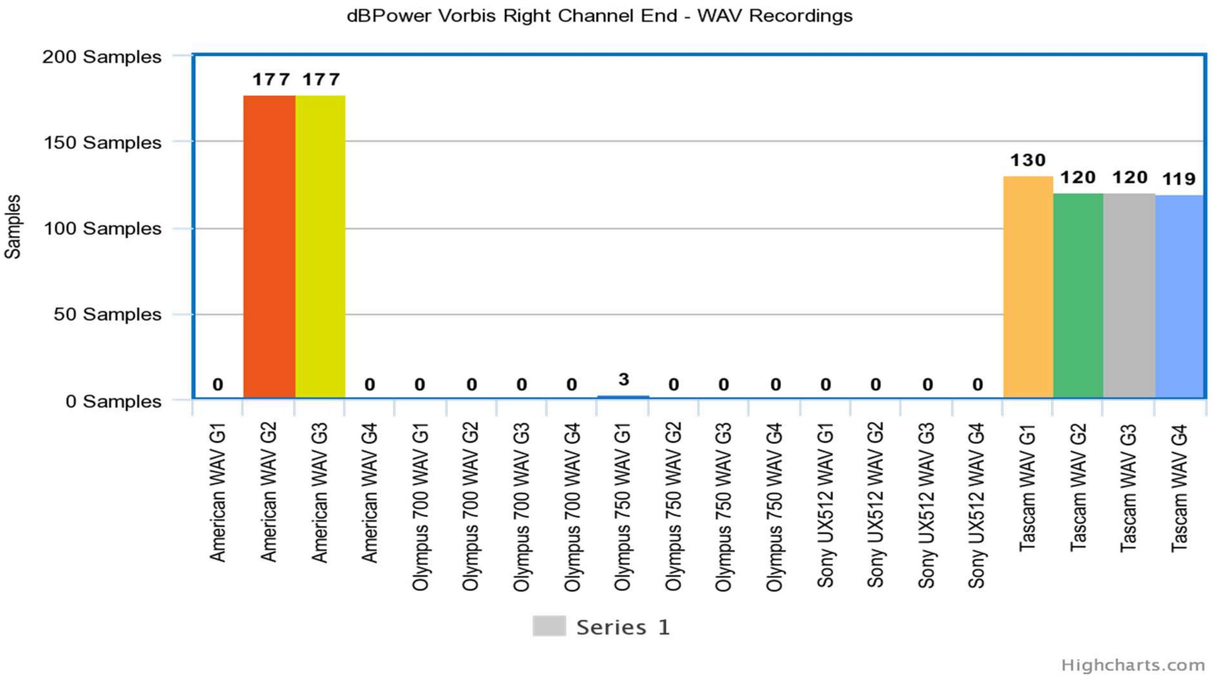

Figure 3.24: dBPower (Vorbis) WAV Recordings—Right Channel End

With dBPower using Vorbis the number of ZLS in the MP3 recordings for the first generation in the beginning of both the left and right channels was reduced across all recordings. The Olympus 700/750 and Tascam recorders saw a significant reduction in their ZLS from their original recordings. The number of samples varied only slightly in both channels as well throughout the four generations of recordings. We can also see that these recordings acted similarly to those encoded by the Adobe Audition Vorbis encoder at the end of the left and right channels. Both encoders added samples for the American Recorder and Olympus 750 at the end of the left channel, as well as only added samples to the Olympus 750 at the end of the right channel.

WAV recordings saw little change, with the Tascam being the outlier in that it was the only WAV recording to start with ZLS initially. Another outlier is the end of the left and right channels, where generations 2 and 3 of the American Recorder saw an increase of 177 samples and then dropped to zero again for the fourth generation.

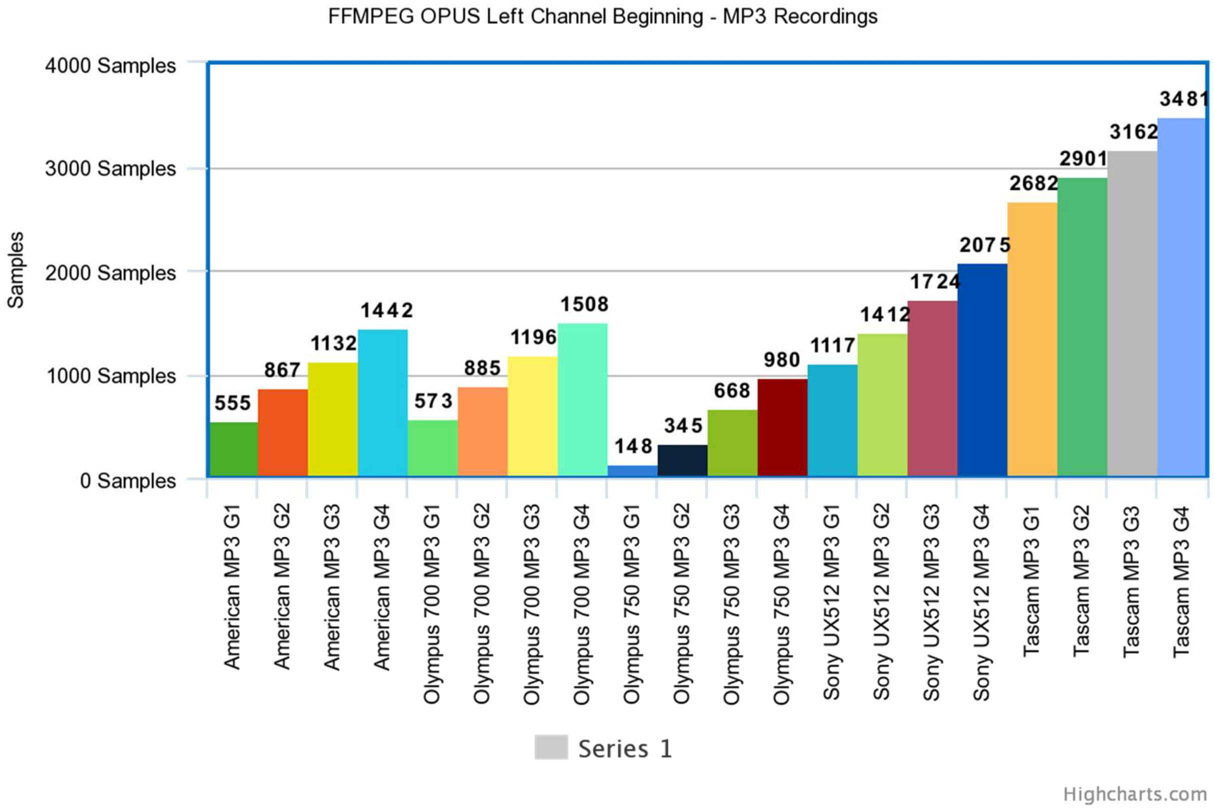

Figure 3.25: FFMPEG MP3 Recordings—Left Channel Beginning

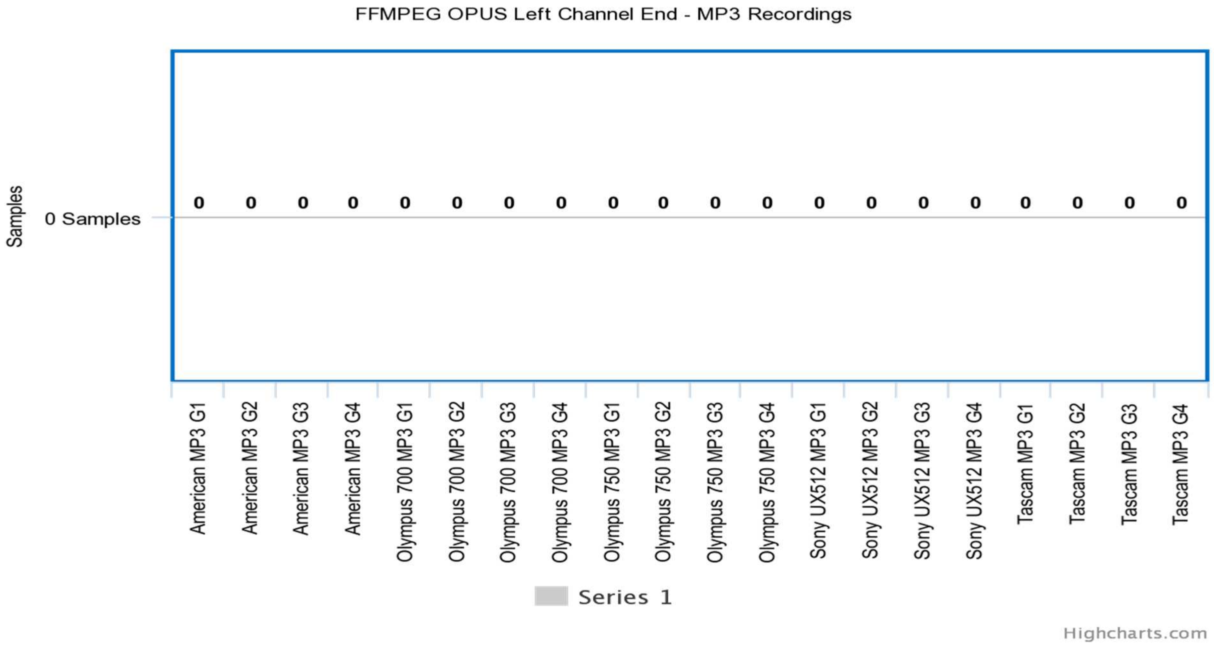

Figure 3.26: FFMPEG MP3 Recordings—Left Channel End

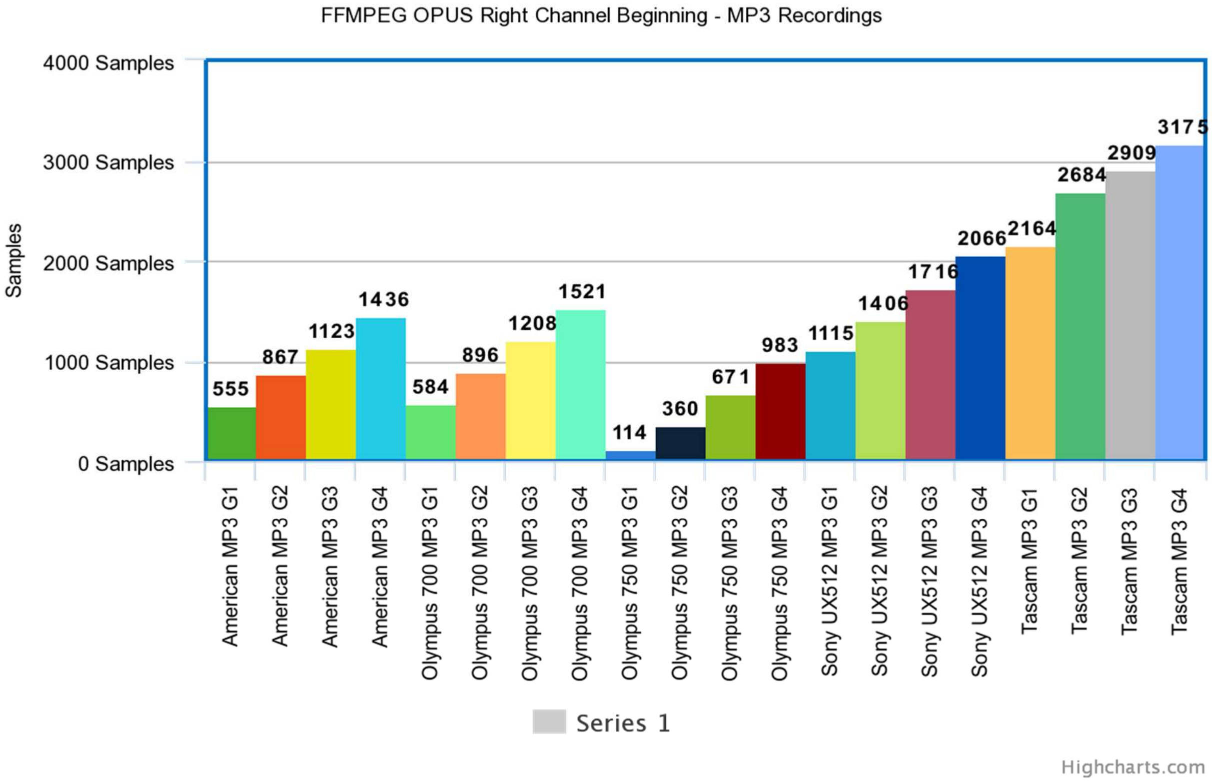

Figure 3.27: FFMPEG MP3 Recordings—Right Channel Beginning

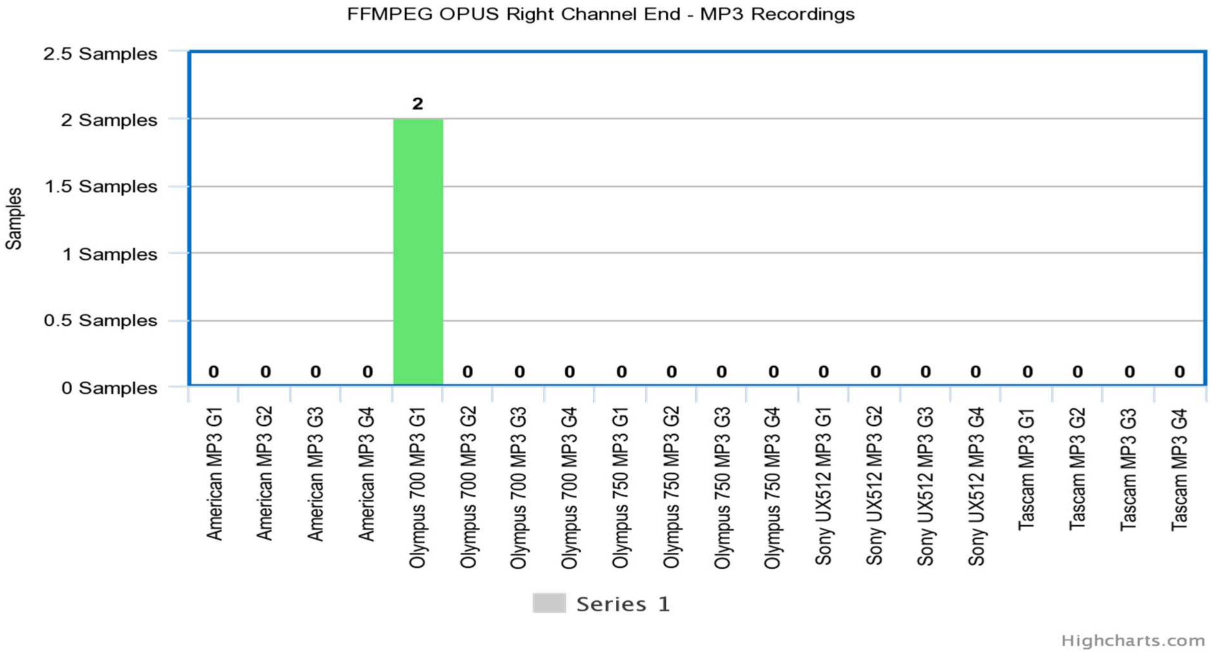

Figure 3.28: FFMPEG MP3 Recordings—Right Channel End

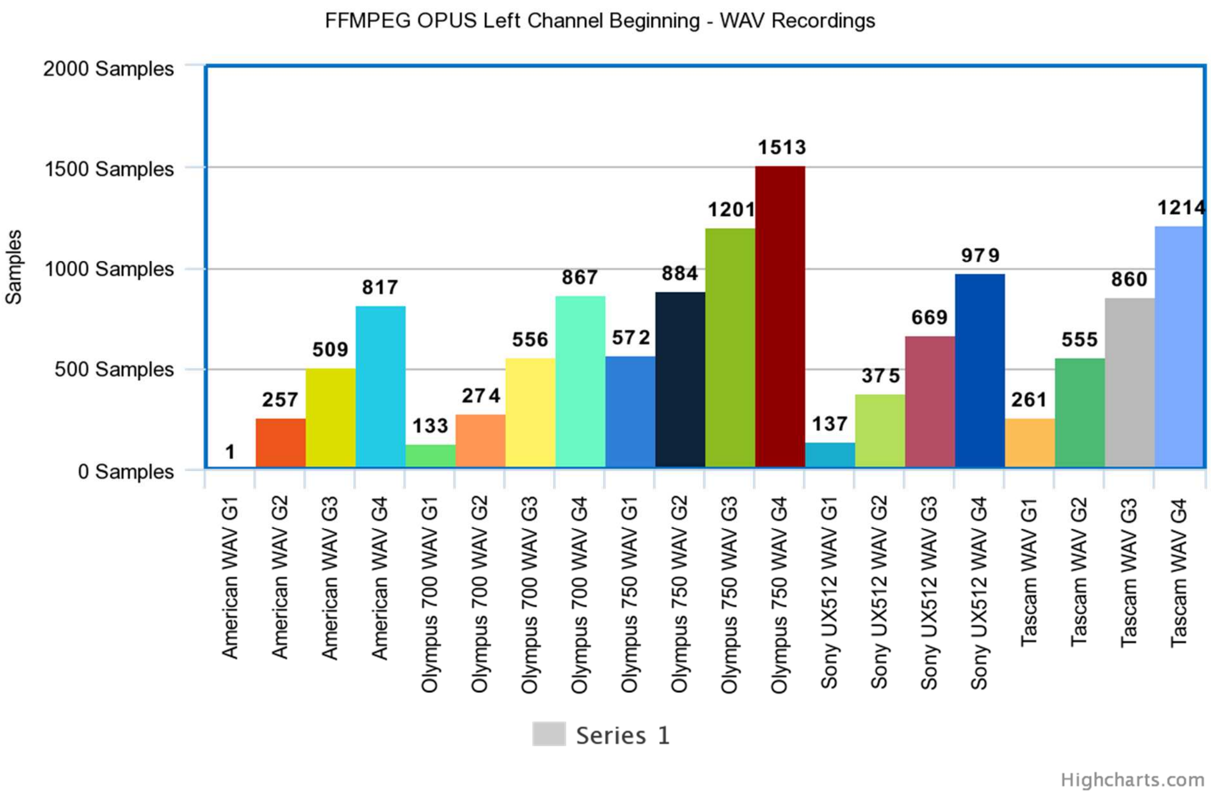

Figure 3.29: FFMPEG WAV Recordings—Left Channel Beginning

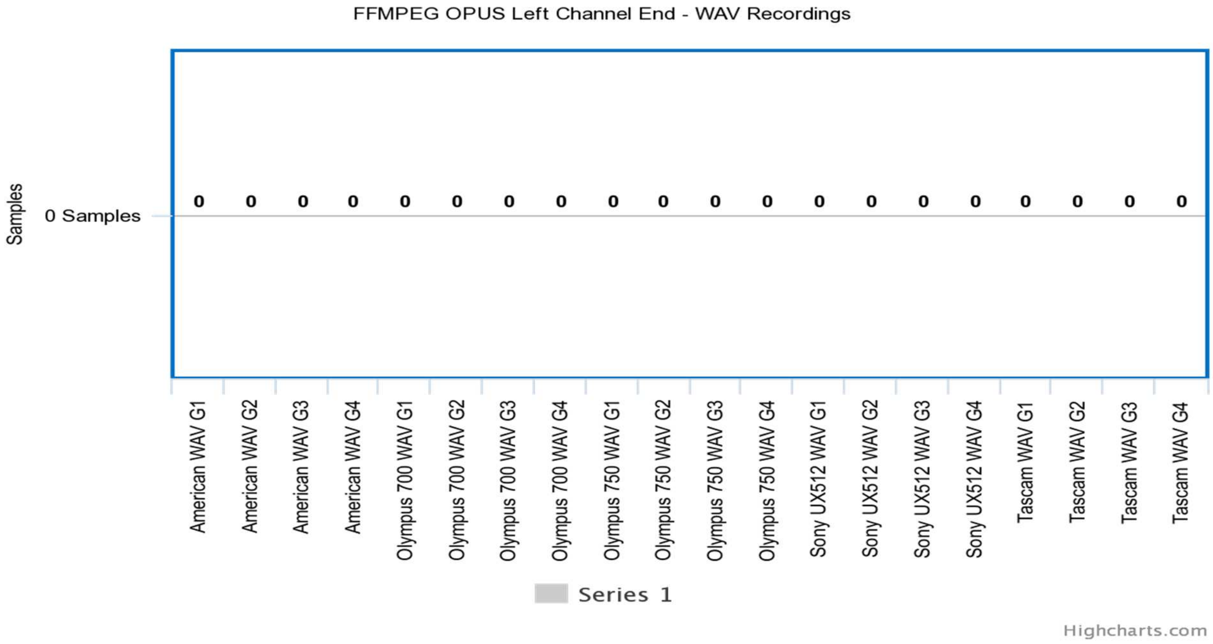

Figure 3.30: FFMPEG WAV Recordings—Left Channel End

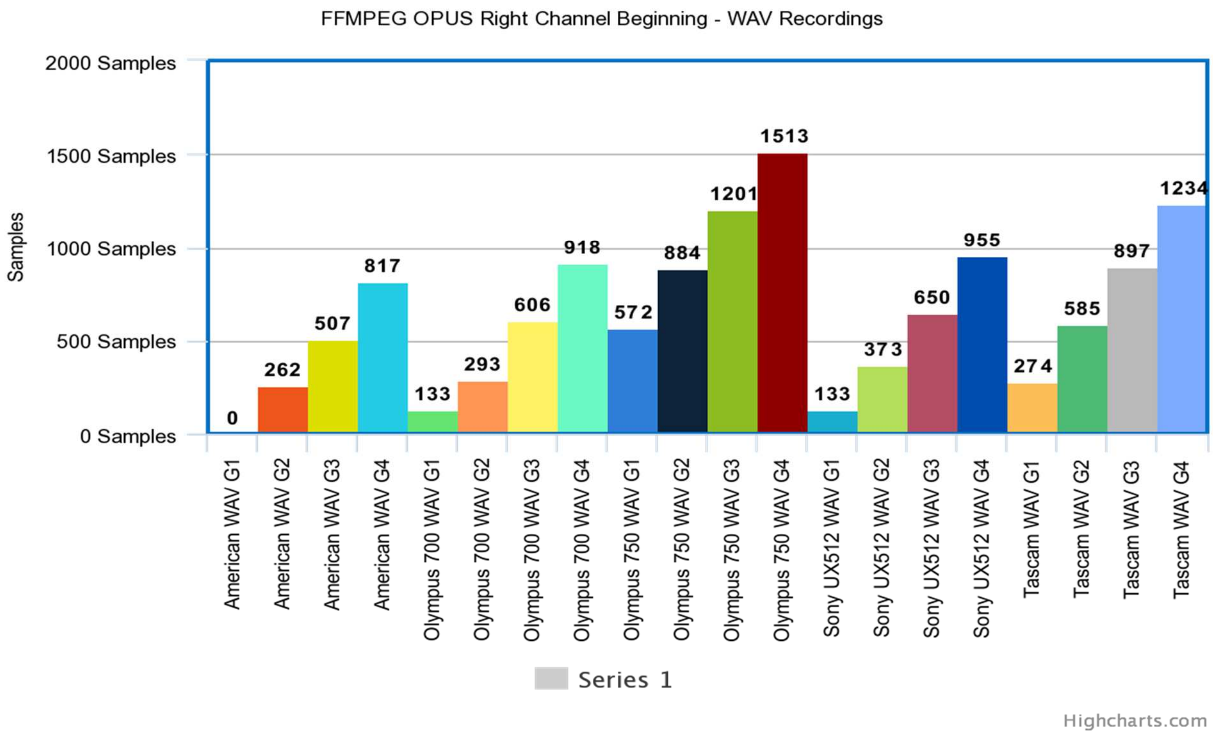

Figure 3.31: FFMPEG WAV Recordings—Right Channel Beginning

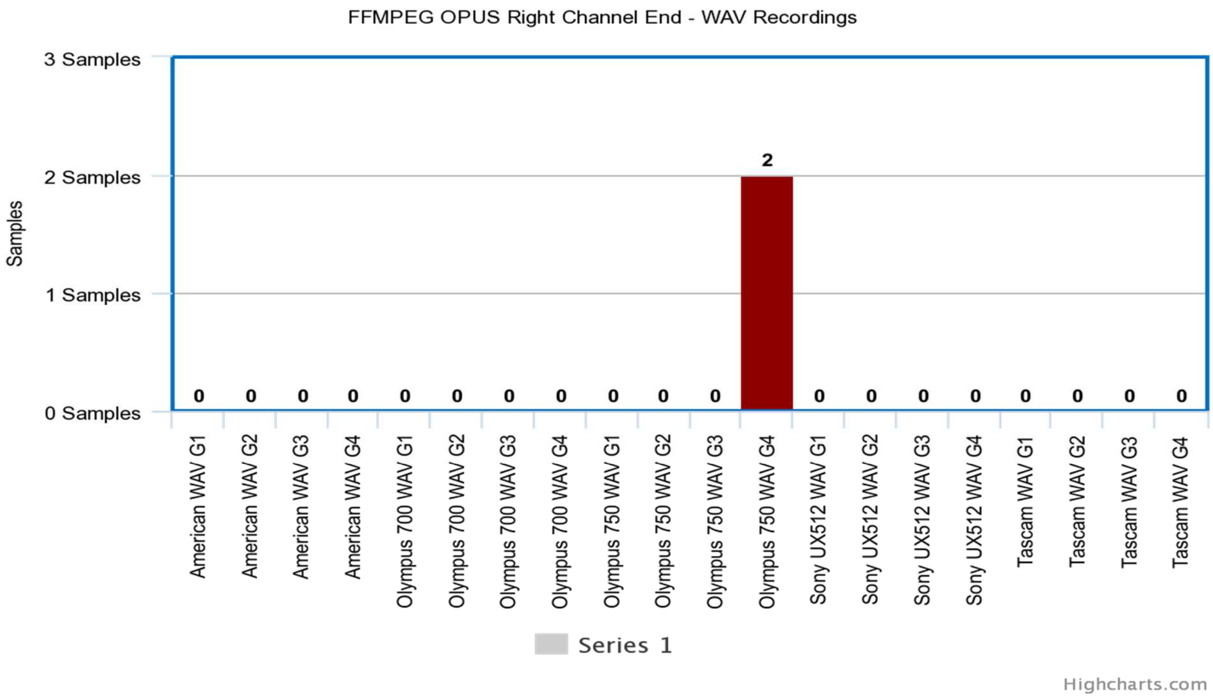

Figure 3.32: FFMPEG WAV Recordings—Right Channel End

FFMPEG using the Opus encoding was the only encoder type to behave as was expected. Each generation of each recording saw an increase in ZLS. This increase, however, did not seem to follow a pattern that could be used as part of a form of audio authentication. We also see that FFMPEG tended to front load the samples in both channels for both the MP3 and WAV recordings, with two outliers seeing an increase of two samples at the end of the right channel.

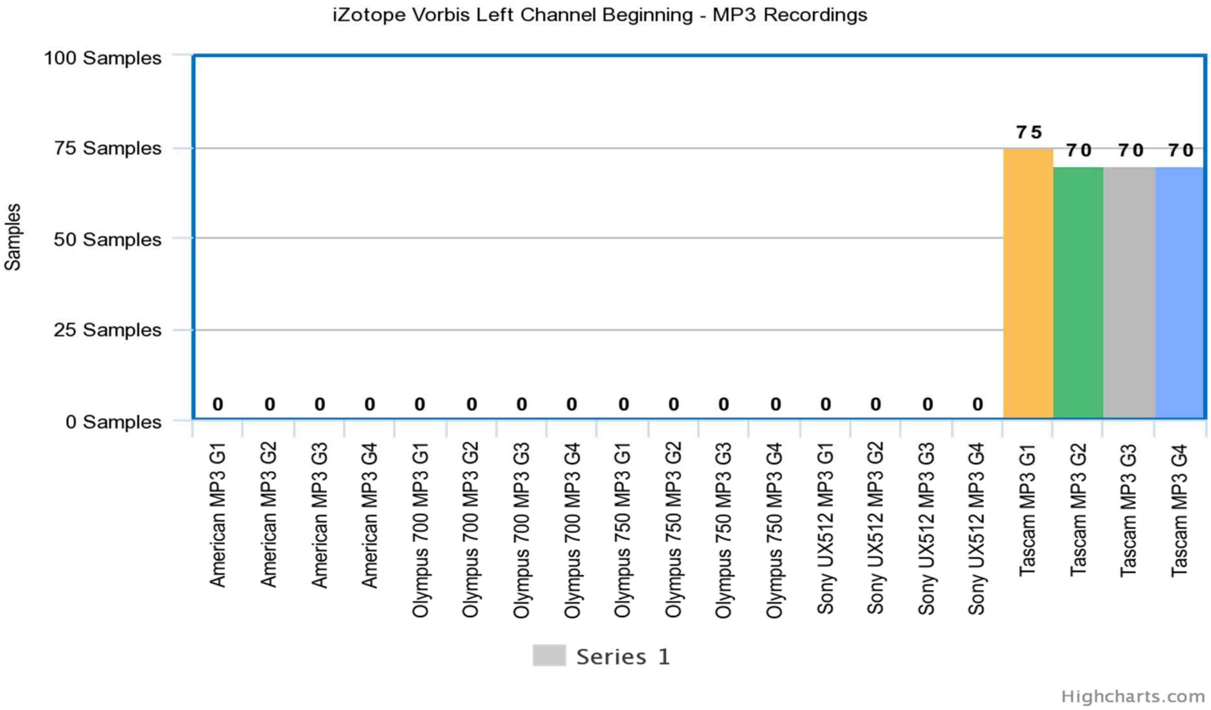

Figure 3.33: iZotope MP3 Recordings—Left Channel Beginning

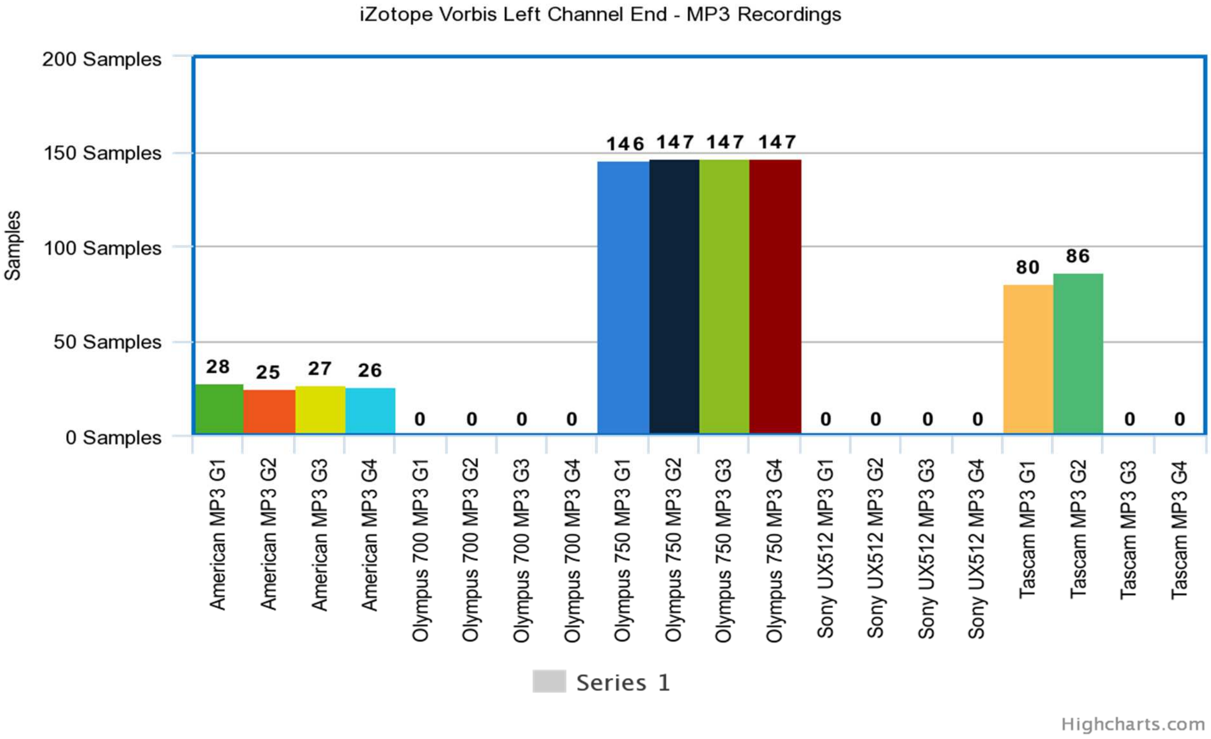

Figure 3.34: iZotope MP3 Recordings—Left Channel End

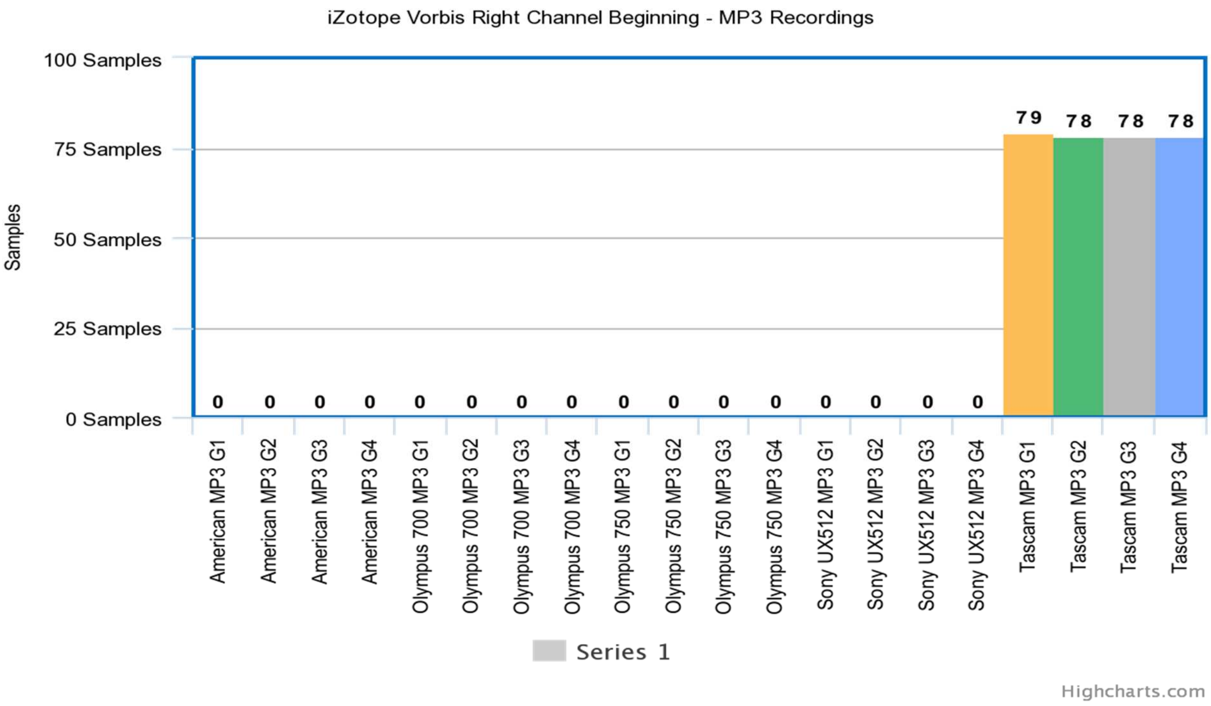

Figure 3.35: iZotope MP3 Recordings—Right Channel Beginning

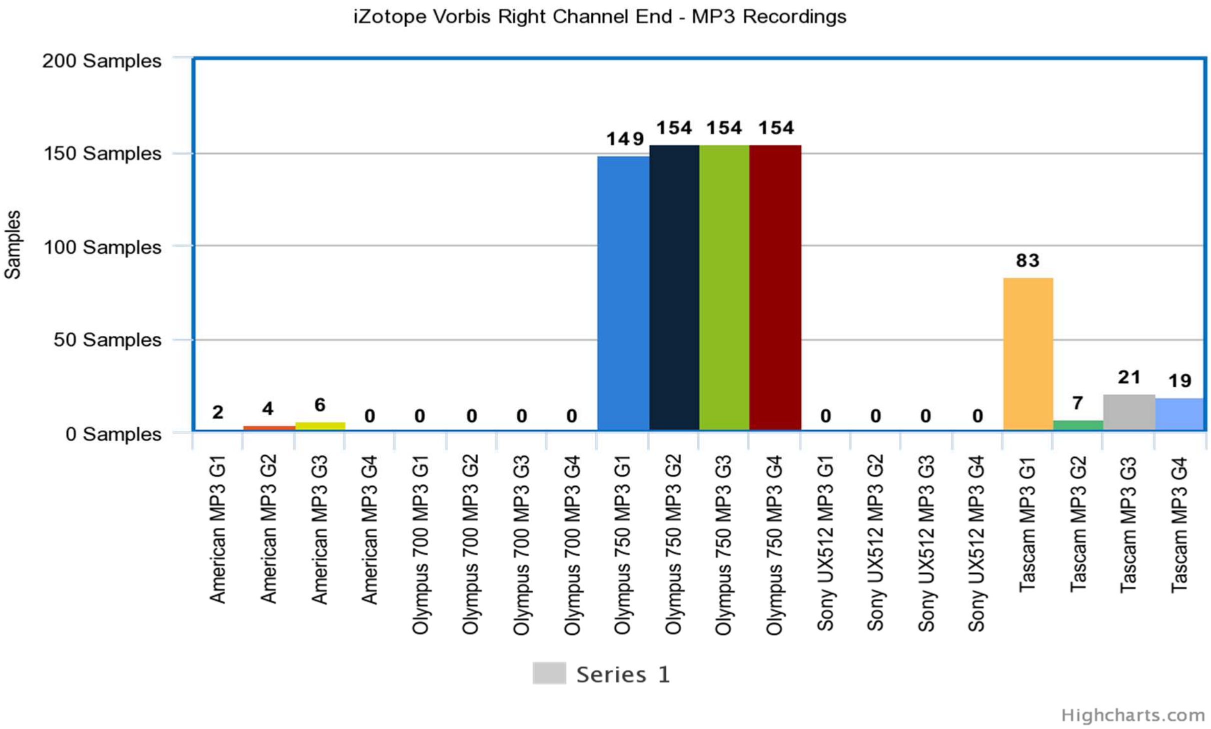

Figure 3.36: iZotope MP3 Recordings—Right Channel End

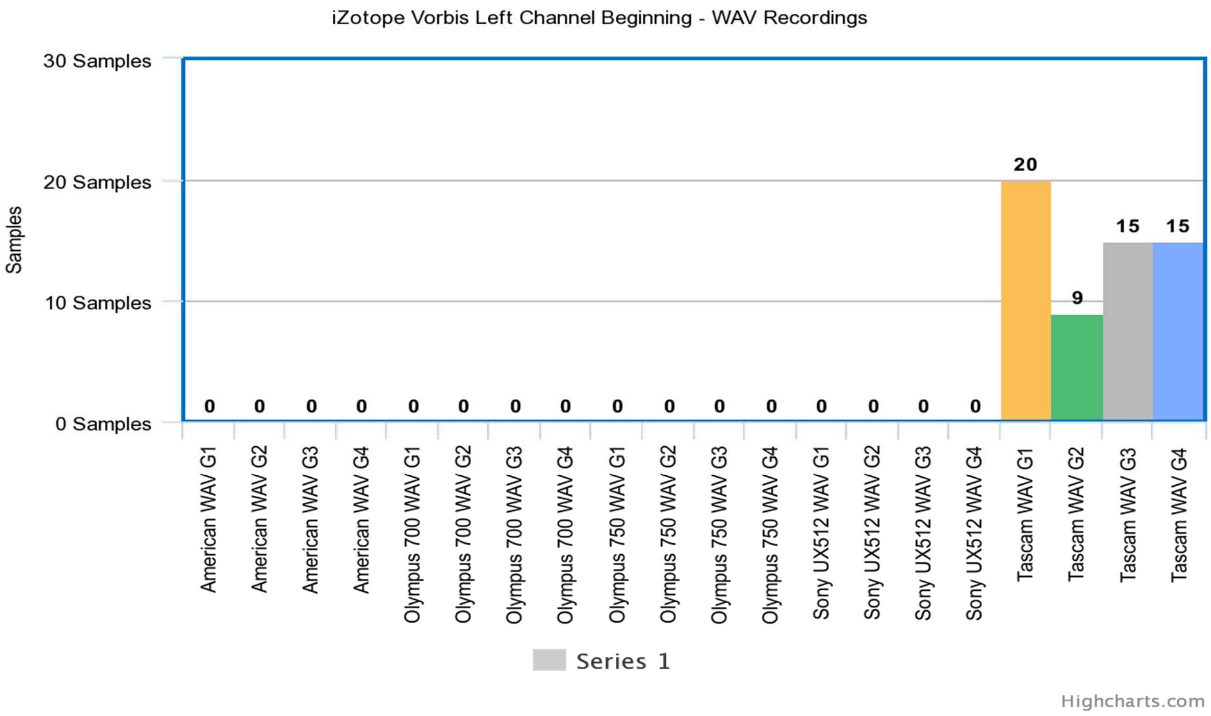

Figure 3.37: iZotope WAV Recordings—Left Channel Beginning

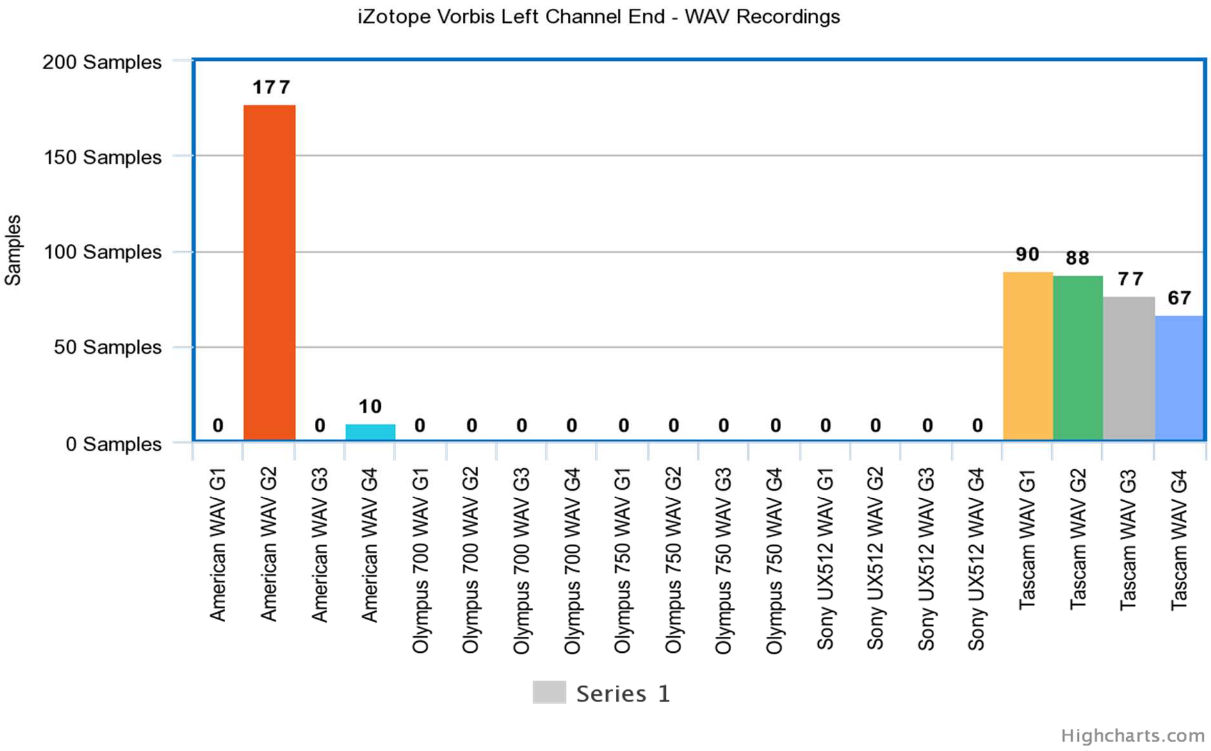

Figure 3.38: iZotope WAV Recordings—Left Channel End

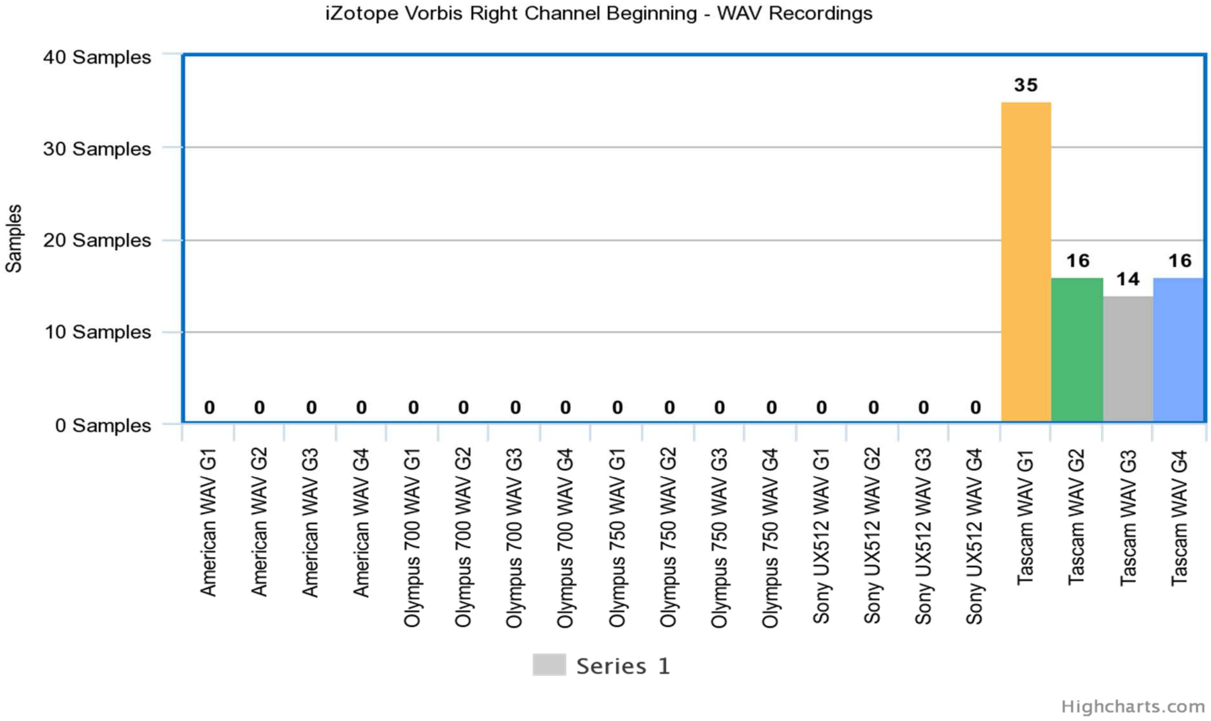

Figure 3.39: iZotope WAV Recordings—Right Channel Beginning

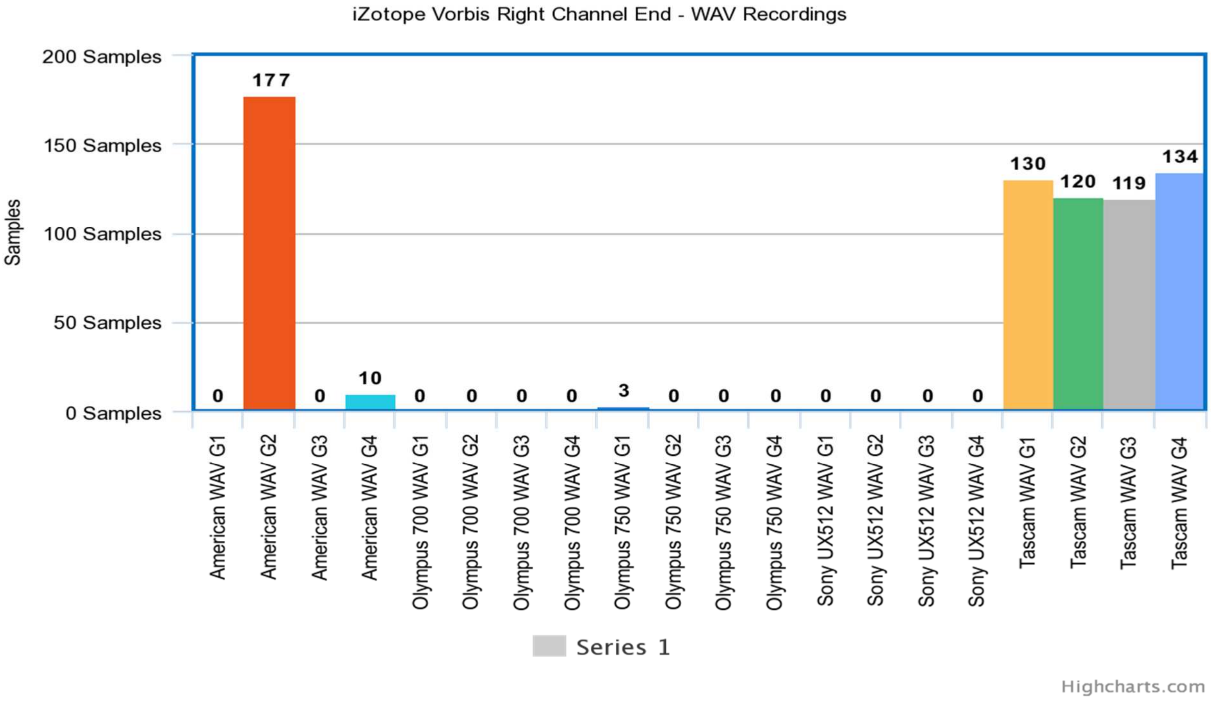

Figure 3.40: iZotope WAV Recordings—Right Channel End

With iZotope using Vorbis encoding, we can see that it acted similarly to the other Vorbis encoders at the end of the right and left channels of the MP3 recordings. Again, the American Recorder and Olympus 750 had ZLS added at the end of the left channel, while the Olympus 750 saw this again at the end of the right channel. We can also see the WAV recordings were affected similarly to the dBPoweramp Vorbis encoding, with zero samples added at the beginning with the Tascam again being the outlier, as well as 177 samples added to the end of the left and right channels of the second generation of the American Recorder WAV file.

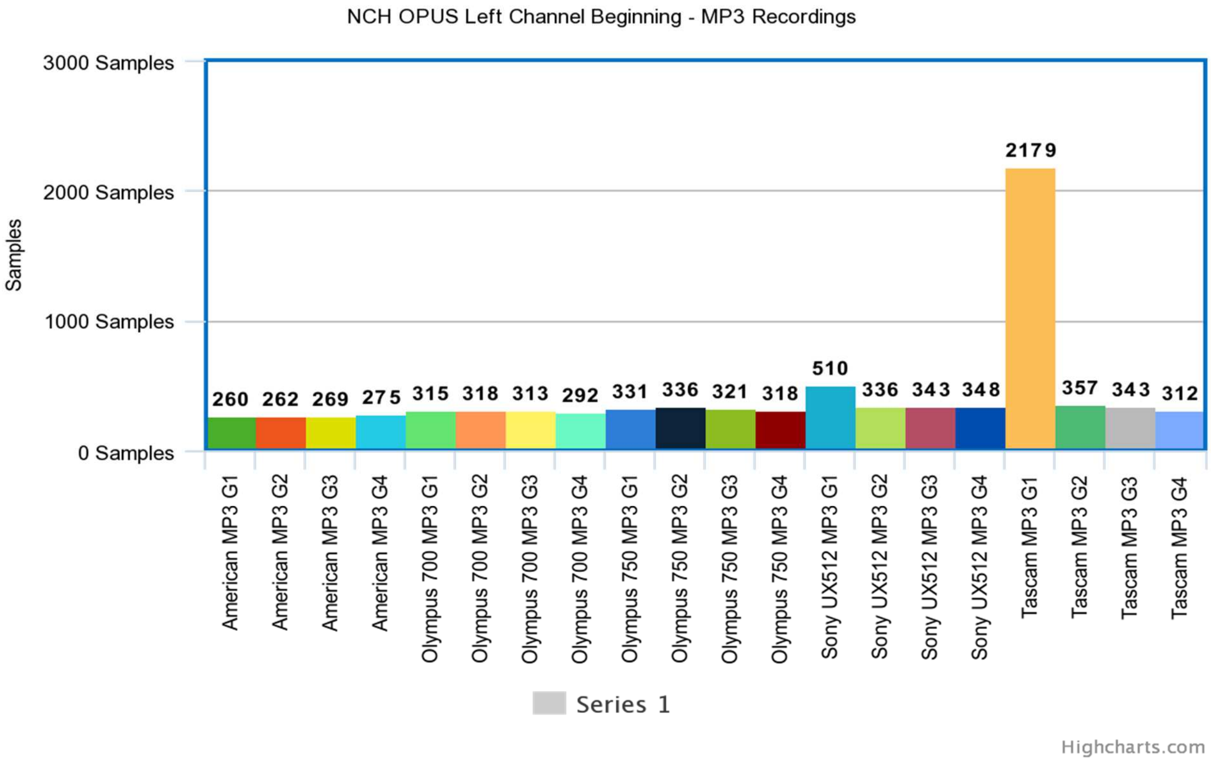

Figure 3.41: NCH Switch (Opus) MP3 Recordings—Left Channel Beginning

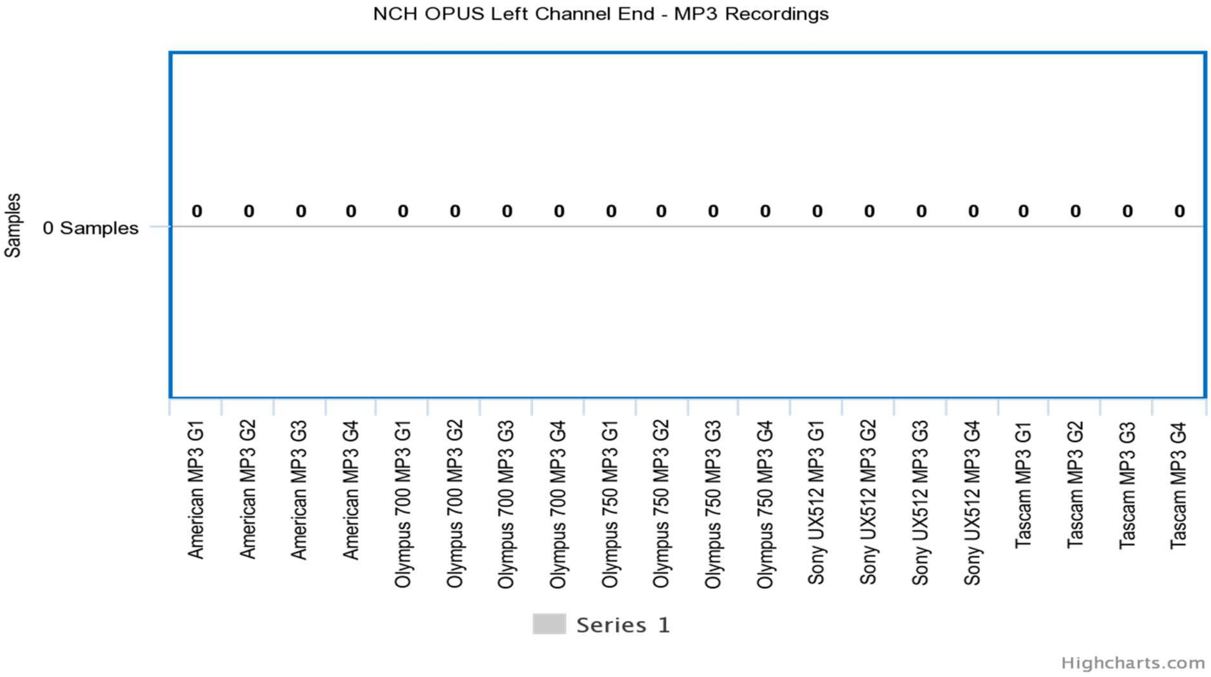

Figure 3.42: NCH Switch (Opus) MP3 Recordings—Left Channel End

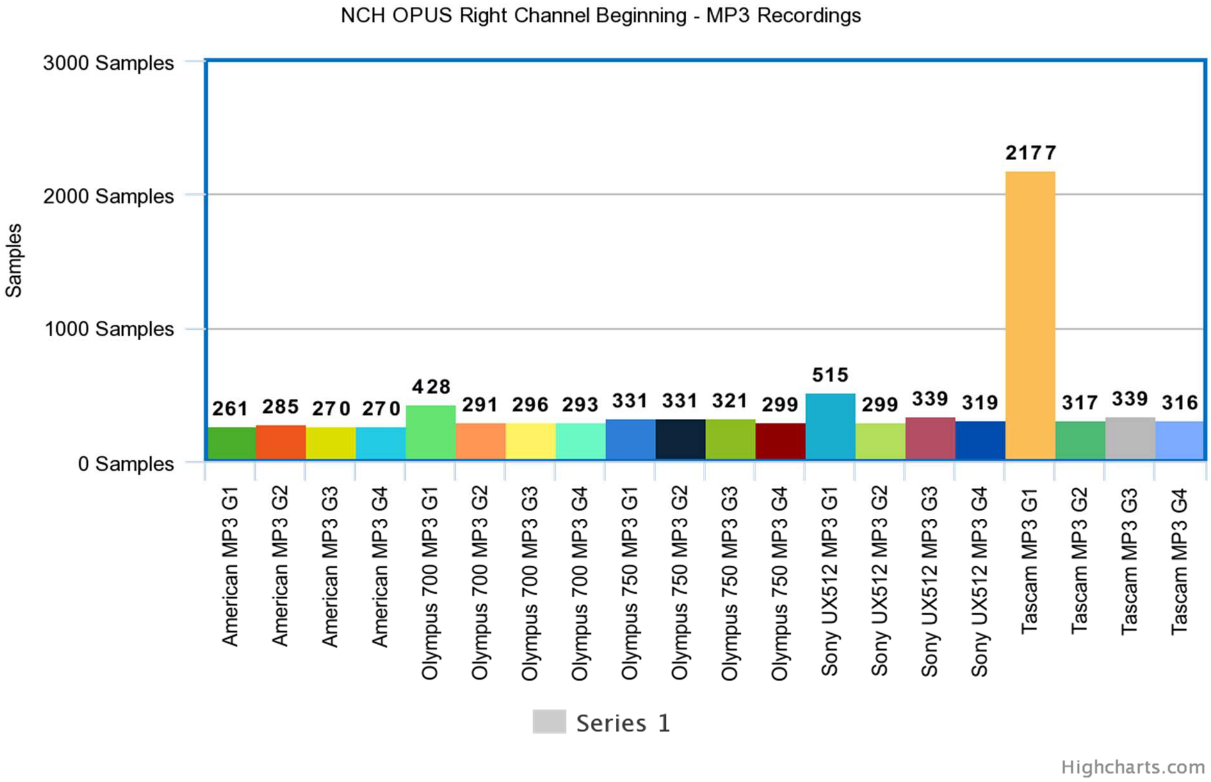

Figure 3.43: NCH Switch (Opus) MP3 Recordings—Right Channel Beginning

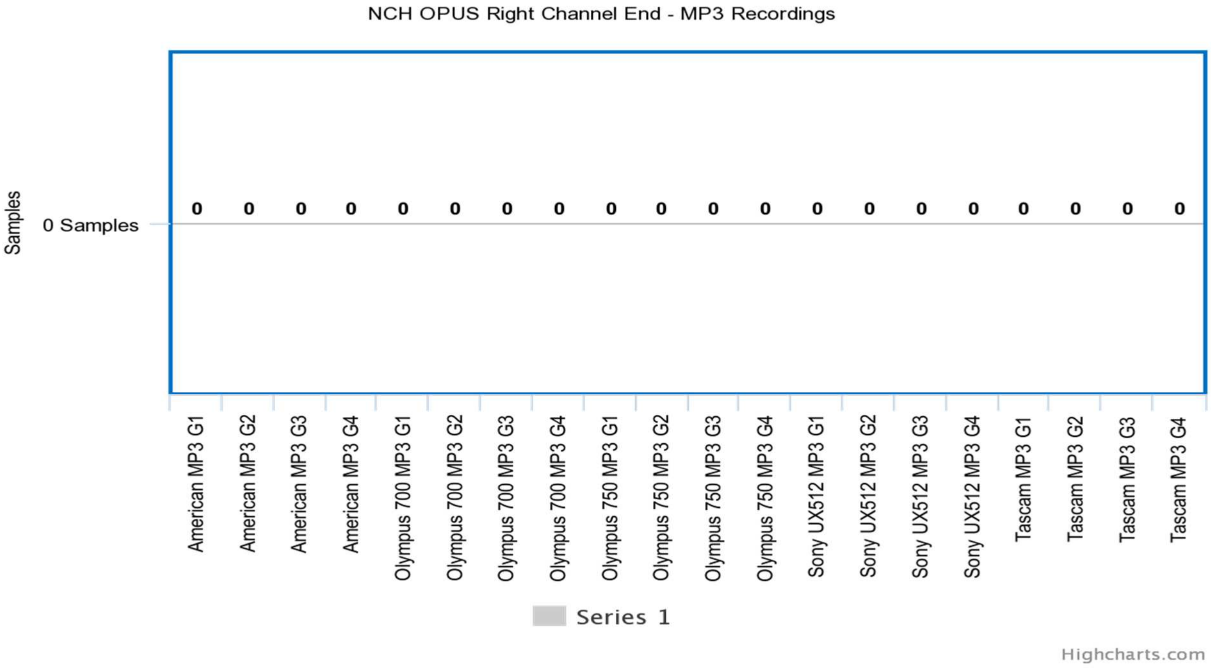

Figure 3.44: NCH Switch (Opus) MP3 Recordings—Right Channel End

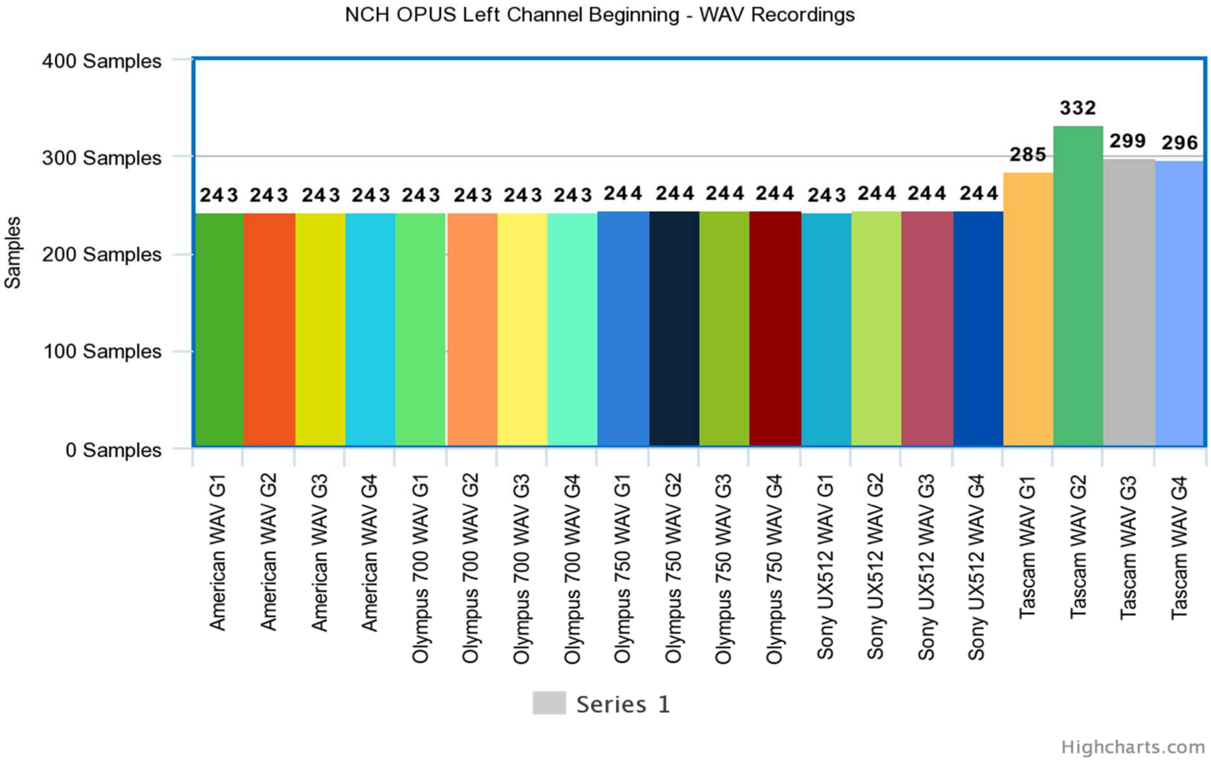

Figure 3.45: NCH Switch (Opus) WAV Recordings—Left Channel Beginning

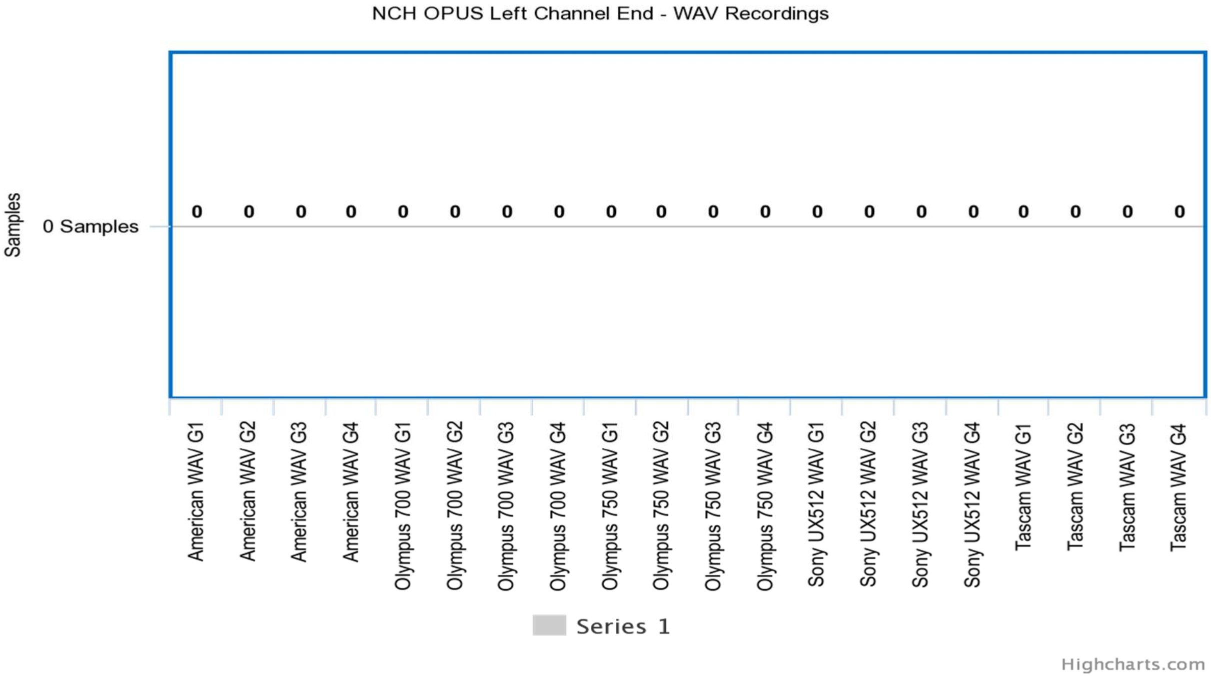

Figure 3.46: NCH Switch (Opus) WAV Recordings—Left Channel End

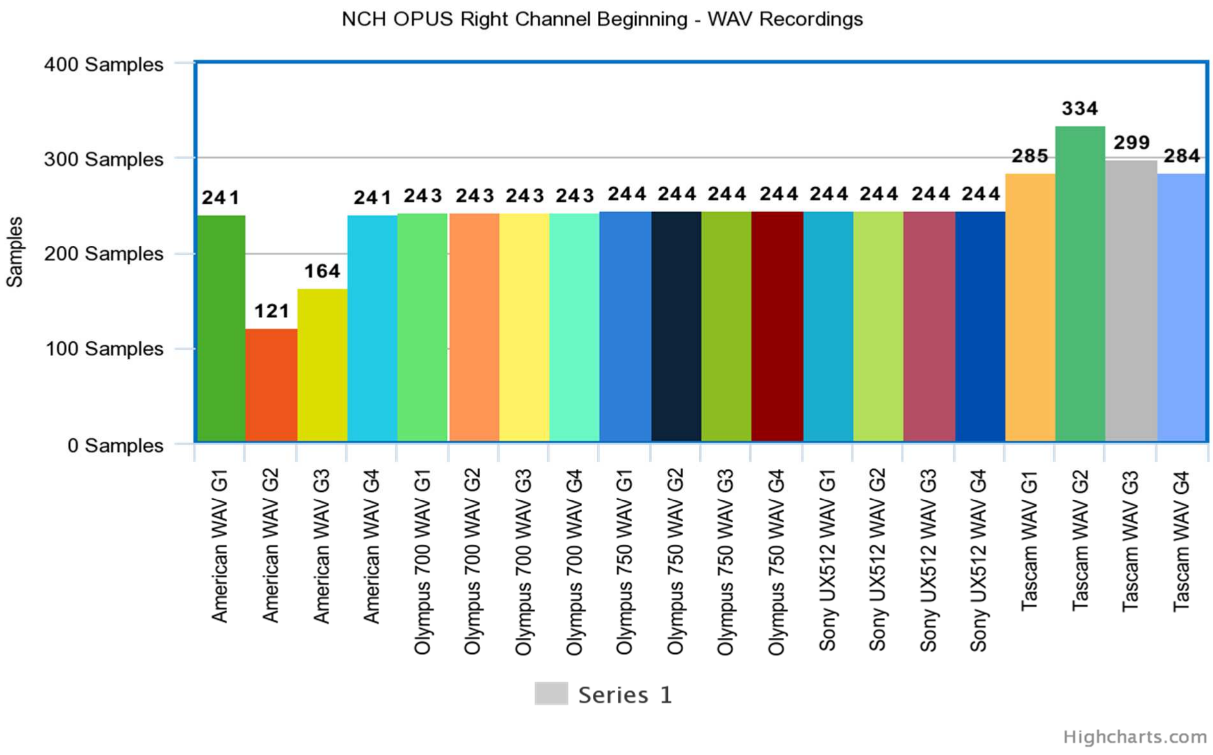

Figure 3.47: NCH Switch (Opus) WAV Recordings—Right Channel Beginning

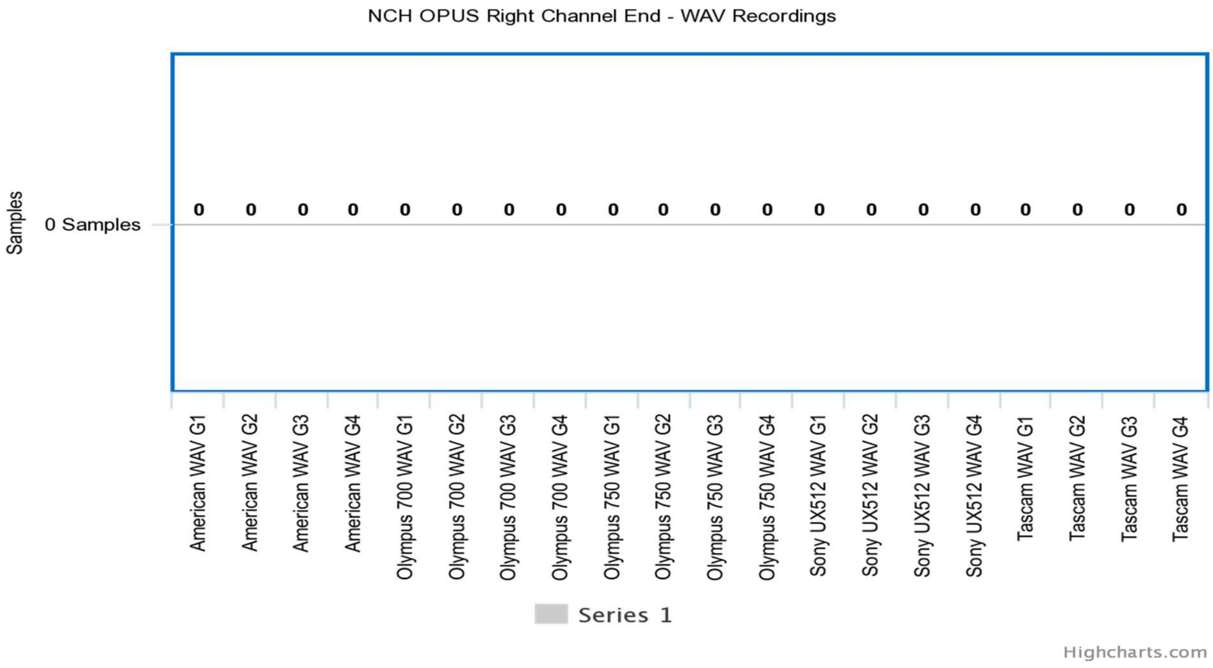

Figure 3.48: NCH Switch (Opus) WAV Recordings—Right Channel End

NCH Switch using Opus did not perform similarly to FFMPEG using Opus in that the number of samples increased with each generation of recordings. The ZLS of the MP3 files were only minorly affected and had small variation between generations. The WAV recordings see a rise in samples from their initial recordings but changed little between generations, like the MP3 recordings. The only similarity between NCH Switch using Opus and FFMPEG was that, again, the samples were front loaded to the beginning of the file, with no samples added to the end of both channels.

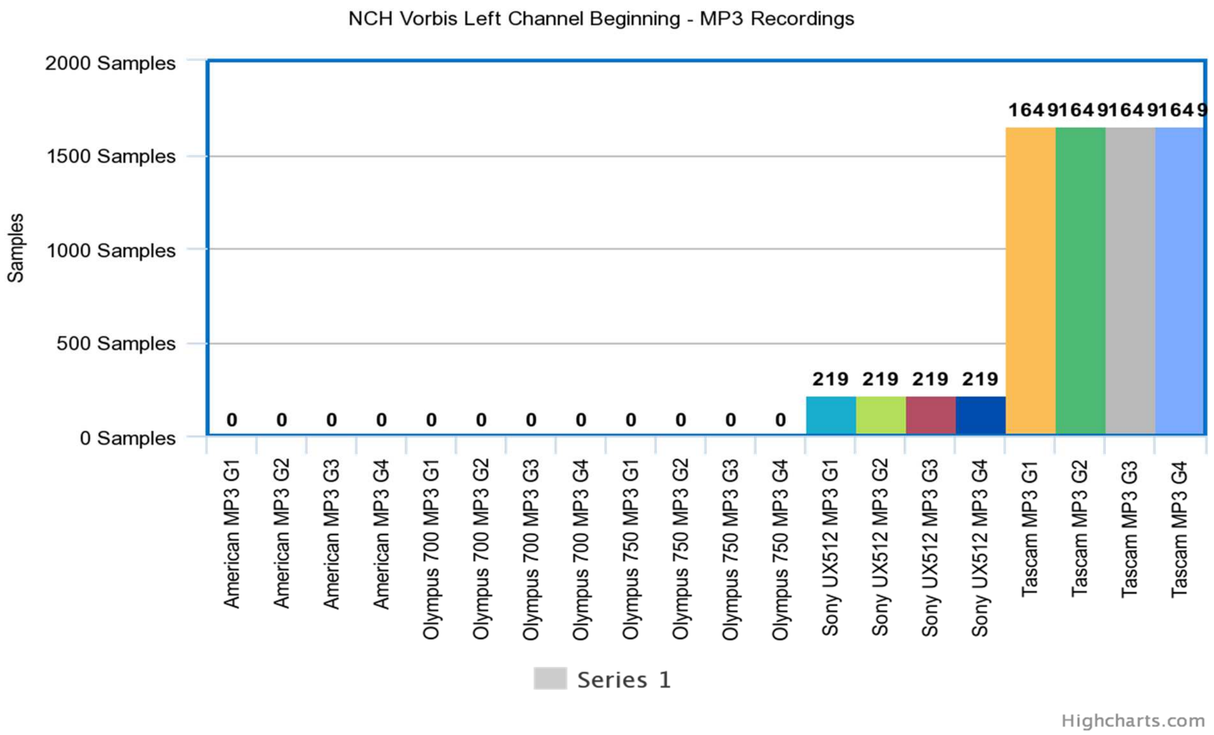

Figure 3.49: NCH Switch (Vorbis) MP3 Recordings—Left Channel Beginning

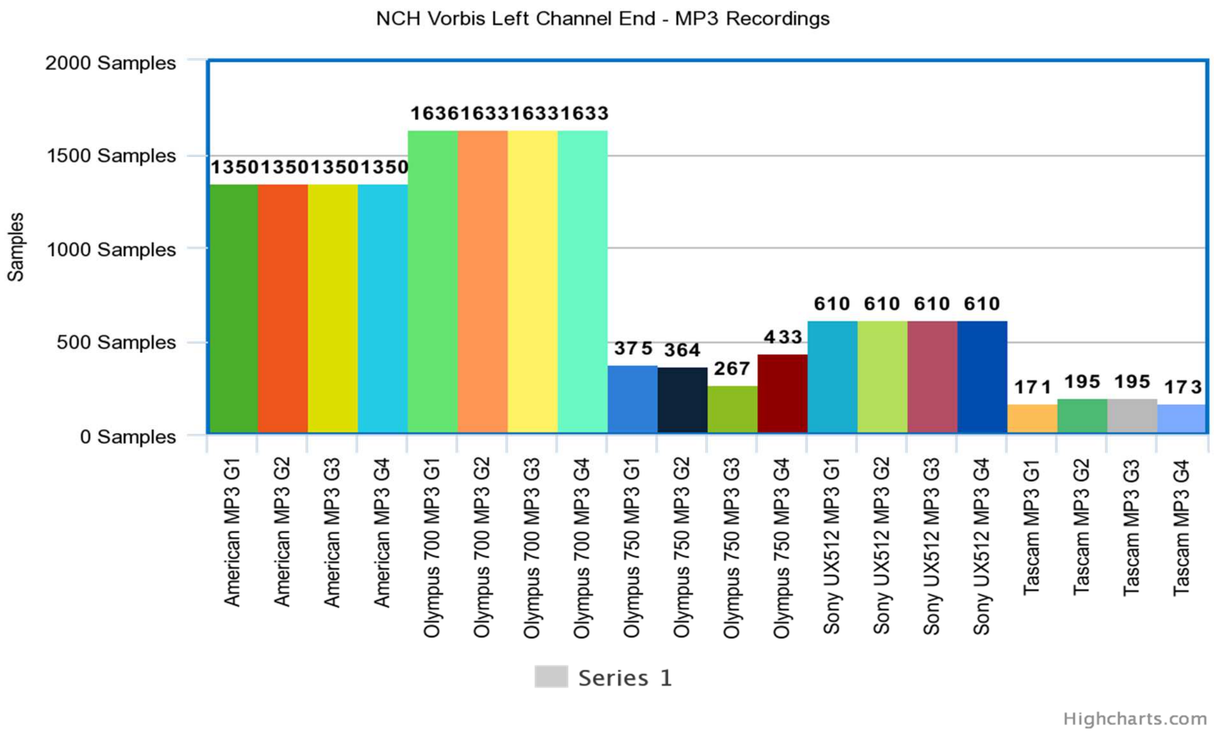

Figure 3.50: NCH Switch (Vorbis) MP3 Recordings—Left Channel End

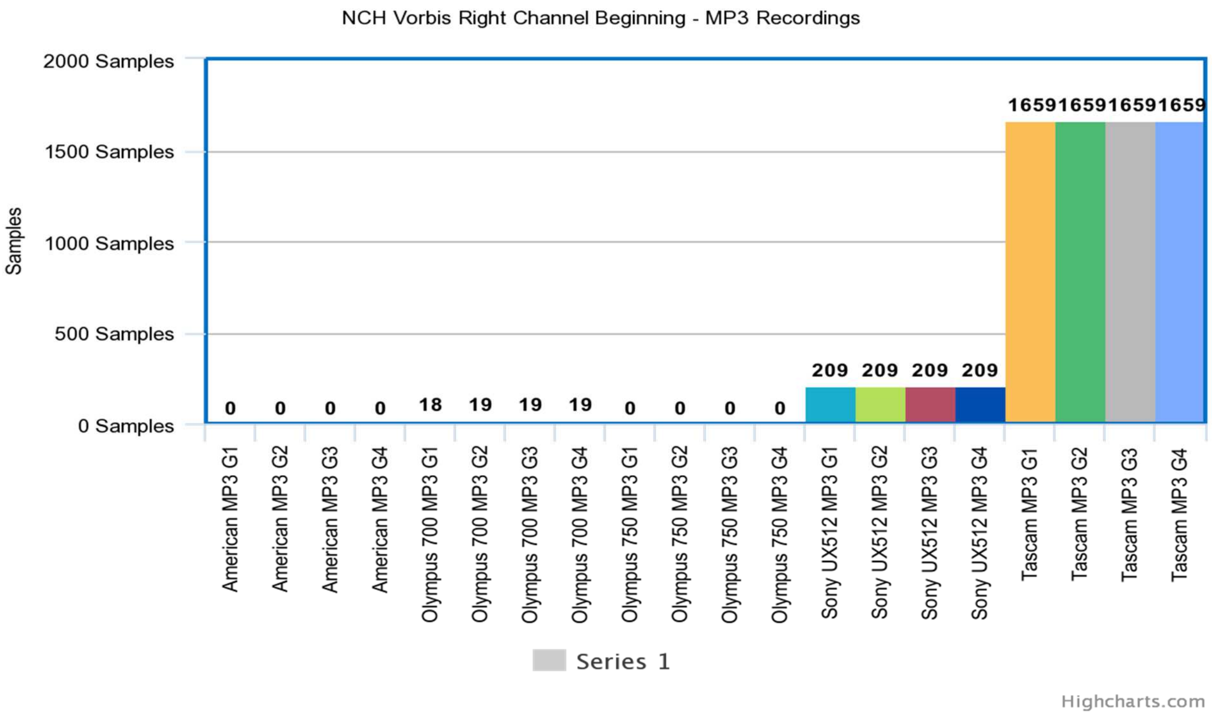

Figure 3.51: NCH Switch (Vorbis) MP3 Recordings—Right Channel Beginning

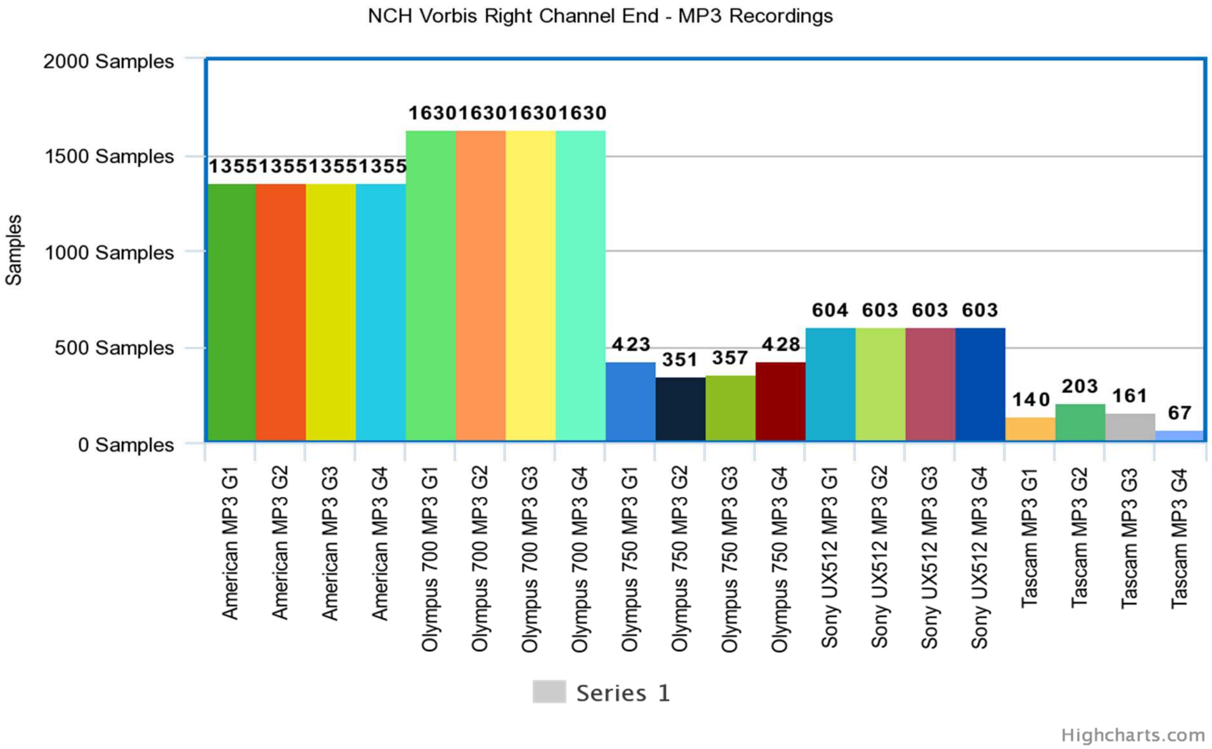

Figure 3.52: NCH Switch (Vorbis) MP3 Recordings—Right Channel End

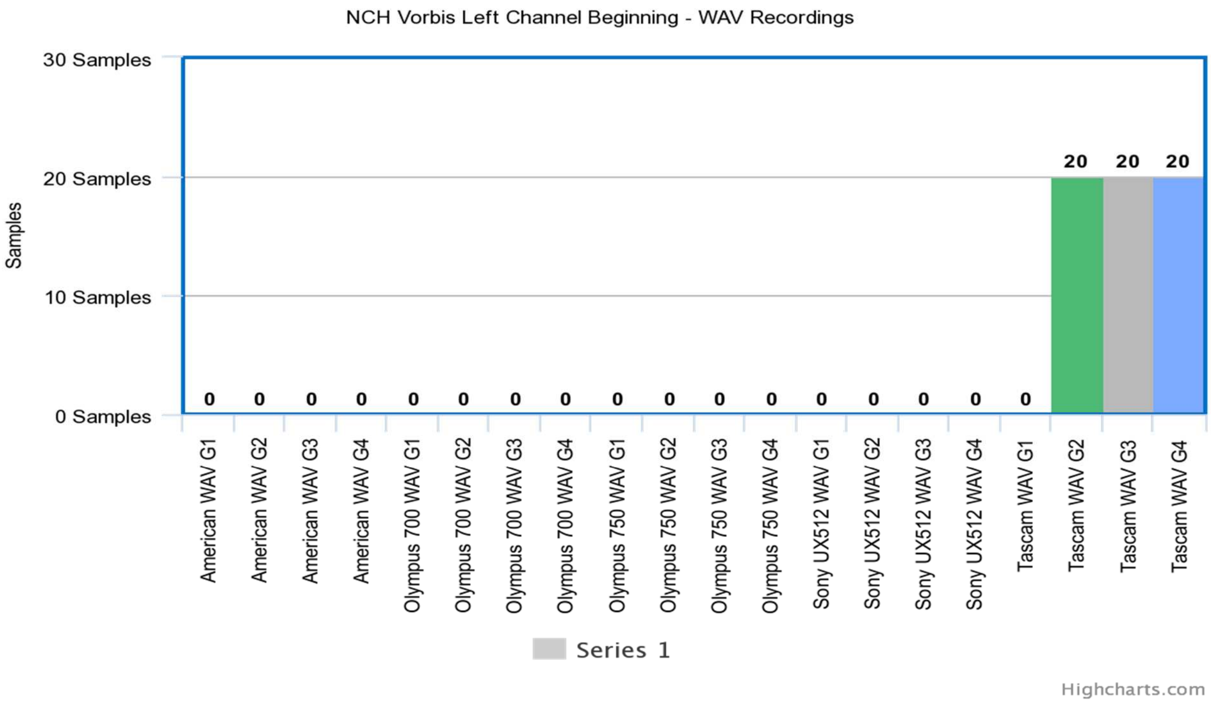

Figure 3.53: NCH Switch (Vorbis) WAV Recordings—Left Channel Beginning

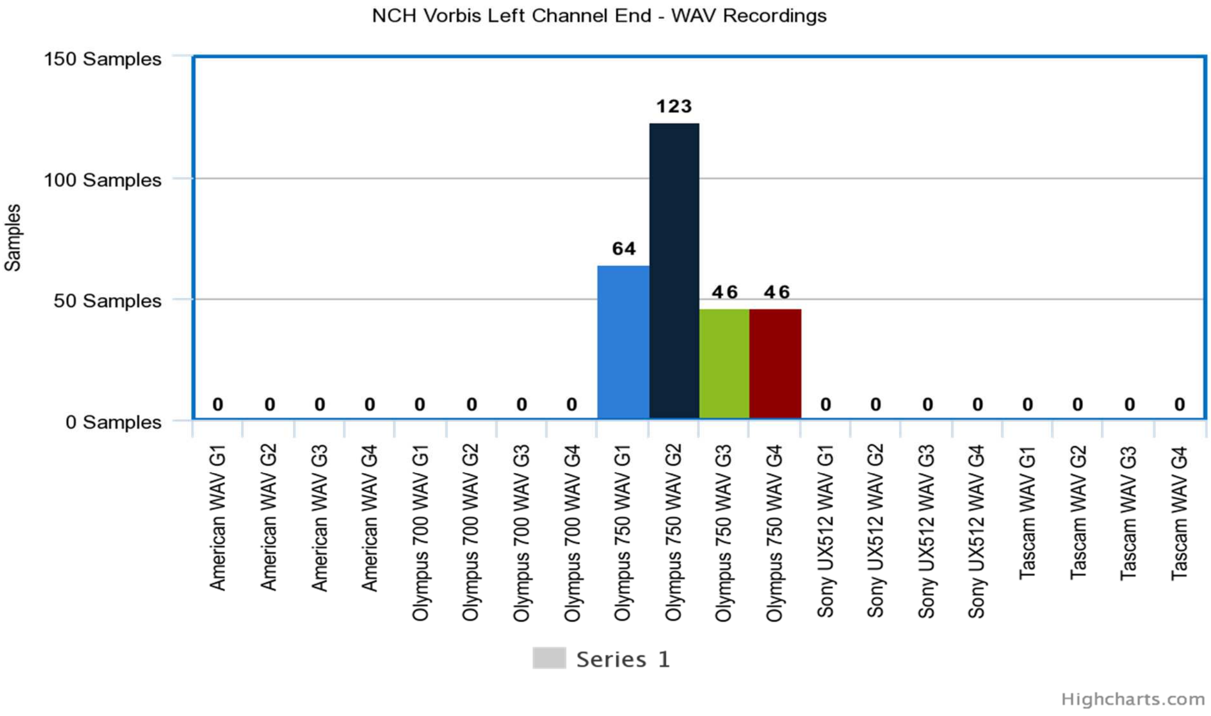

Figure 5.54: NCH Switch (Vorbis) WAV Recordings—Left Channel End

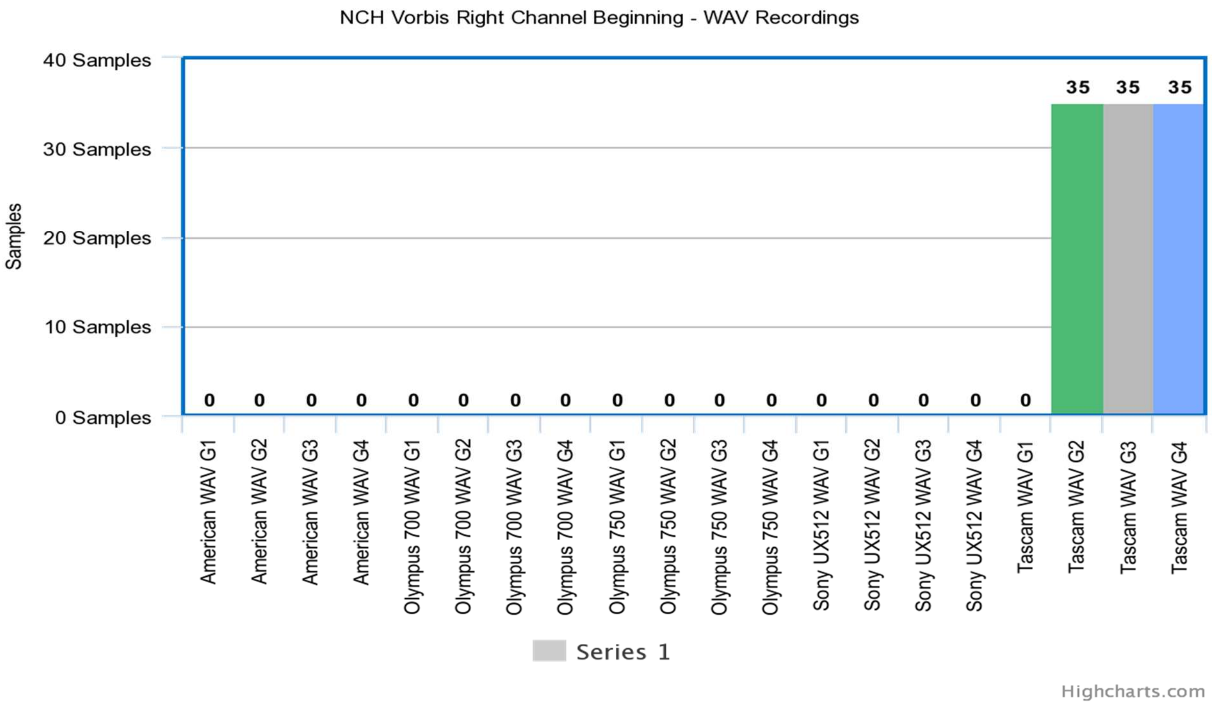

Figure 5.55: NCH Switch WAV Recordings—Right Channel Beginning

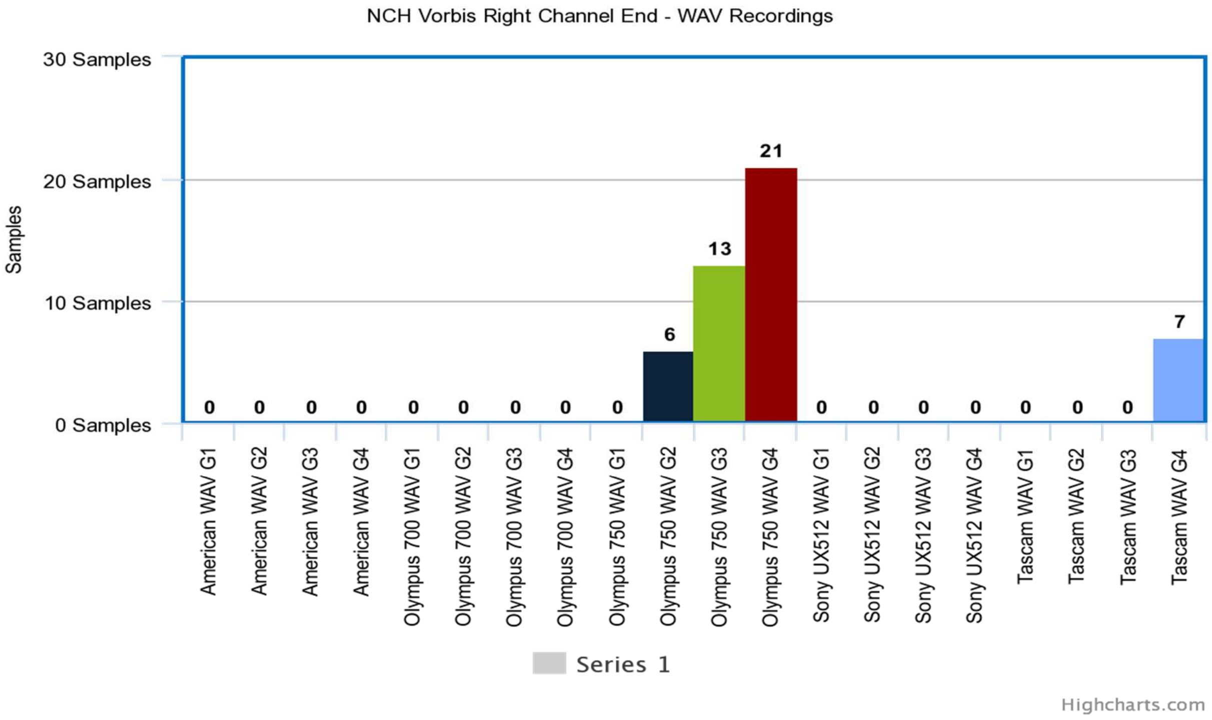

Figure 5.56: NCH Switch WAV Recordings—Right Channel End

With NCH Switch using Vorbis encoding, we see the biggest variation of how ZLS were distributed across the different recorder types. The Sony UX512 and Tascam recorders had the same number of samples for each of their four generations at the beginning of both channels of their MP3 recordings, with difference of only 10 samples between them. The other recorders had little, if any, samples added to the beginning of their MP3 recordings.

The American Recorder and Olympus 700 saw a significant increase in samples at the end of both of their channels. Again, the number of samples does not change across all four generations of recording, with a small difference between both left and right channels. The Olympus 750 also saw samples added to the end of both of its channels, again with little difference between channels and no change between generations of recordings.

The WAV recordings saw no samples added to the beginning of their recordings, with the outlier again being the Tascam recorder. The Tascam did have samples added to the second through fourth generations of recordings, but the number of samples between generations did not change for either the left or right side.

Only the Olympus 750 and the Tascam (fourth generation recording) saw any samples added to the end of either channel.

CHAPTER IV

DISCUSSION

Originally it was said that with each subsequent generation of compression, it would be expected for the number of zero-level samples to increase[3]. However, as we can see with the results, only FFmpeg using the Opus encoding behaved as expected with each generation for all recordings and recording devices. Even then, the number of zeros added between each generation did not increase in a way that would indicate a pattern between generations. Berman and Yancy discussed this in their research on ZLS analysis for various other codecs[1,2]. They also found that not all audio programs and codecs changed the number of zero-level samples through multiple generations of compression.

One interesting observation that could be made with these findings is that, in every file except one (Olympus 750M WAV Generation 4), the programs that used the Opus encoding put all zero-level samples at the beginning of the file and none at the end. Even when using the same program for Vorbis and Opus encoding, you can see that both dBPoweramp and NCH Switch both front loaded the zeroes when using Opus encoding, and varied when using Vorbis encoding.

iZotope, NCH Switch, and dBPoweramp behaved the most consistently across all generations with all recordings and recording devices. Often the changes were very slight and in many cases there were no changes in zeroes between generations.

Perhaps unsurprisingly, the original WAV recordings for each device (except the Tascam GTR1 and American Recorder) had no original zero-level samples. The American Recorder only had one zero level sample in the beginning of the left channel, this may be an outlier. However, the WAVs did not always increase in ZLS between generations either, as many stayed at zero throughout each generation, even using different encoders and audio programs.

CHAPTER V

CONCLUSION

It was stated in the purpose of this study that the information found could be used to create new ways of authenticating audio recordings. Unfortunately, the number of ZLS between recordings did not indicate a measurable pattern between generations of compression and did not always behave as expected depending on which audio program or encoder was used to compress the audio. As a result, we could not say for certain what generation of compression, if any, has been done to an audio recording based off of the ZLS analysis alone. A ZLS analysis may be part of a larger set of tests of audio authentication if the original recorder and settings on an audio file in question are known, but even with that it would be difficult to say for sure if recompression has occurred due to some programs showing little if any variance between generations of recordings.

That being said, if the audio file in question is claimed to be an original WAV file, and has a significant number of ZLS, we can see with this analysis that there is a good possibility that it has been recompressed at least once. While some recorders and programs did not add any ZLS for each generation that came from a WAV file, many did add at least some. Keeping that in mind it is possible that again, a ZLS could be part of a bigger authentication analysis for a WAV file.

CHAPTER VI

FUTURE RESEARCH

Future research could include how one recorder with the capability of recording in a multitude of different formats, such as MP3, OGG, WMA, WAV, behaved across multiple different audio programs, recompressing using the same audio encoder. This could help build a database of ZLS analysis where the controller and encoder were controlled, but the audio programs were different.

A ZLS analysis of surround sound encoders, such as AC3 is also a possible avenue for future research, as those recording types become more common.

A meta-analysis of this research combined with previous research by Josh Berman[1] and Jeremy Yancy[2] could be performed as well. This could help narrow down which codecs and programs could be used to build a ZLS analysis database that could be used for future audio authentication purposes. Because of the variance shown between programs and codecs, we have concluded that it is not reliable to perform only a ZLS analysis for authentication purposes, but it would be good to establish which codecs behave as one would expect with multiple generations of compression.

REFERENCES

1. Berman, Josh *Analysis of Zero-Level Sample Padding of Various MP3 Codecs*, M.S. Thesis. University of Colorado, 2015

2. Yancey, Jeremy *Analysis of Zero-Level Sample Padding of AAC and WMA Encoders*, M.S. Thesis. University of Colorado, 2019

3. Schroeder, Ernst F., and Johannes Boehm. "Original File Length (OFL) for mp3, mp3PRO and Other Audio Codecs." Audio Engineering Society, 22 Mar. 2003.

4. "Vorbis Audio Compression." *Xiph.org,* 2016, xiph.org/vorbis/.

5. "Ogg Bitstream Overview." *Xiph.org,* 2010, https://xiph.org/ogg/doc/oggstream.html

6. "About Xiph." *Xiph.org,* 2016, https://xiph.org/about/

7. Allamanche, Eric, Ralf Gieger, Jürgen Herre, and Thoma Sporer. "MPEG-4 Low Delay Audio Coding Based on the AAC Codec." Audio Engineering Society (1999). Web. 14 Oct. 2015