Psychoacoustic Noise-Shaping in digital signal audio processing

Robin Prillwitz (*00805291*)

Applied Computer Science, Deggendorf Institute of Technology

Englisch für Ingenieure 2: Presenting in English (*O7105*)

Dr. Jehona M. Murtezani

January 8, 2023

Contents

Introduction

This study assignment contains a transcript and analysis of my presentation titled "Psychoacoustic Noise-Shaping". It was held on the December 21, 2022 on the premises of the *Deggendorf Institute of Technology (DIT)*.

The presentation intended to introduce the audience to the topic of psychoacoustic noise-shaping. The main goal was to argue for the necessity of the technique, explain its concepts and mechanism and furthermore clarify all steps to a certain point. The goal was not to enable the listener to be able to implement such an algorithm but rather to introduce and demystify the topic in an approachable and understandable manner. The topic was chosen as it contains relatively simple operations of digital signal processing which are applied to a real, rather unknown, and nearly-impossible sounding application.

This paper does not focus on the topic of noise-shaping. It's simply reciting parts of a transcript and offers an analysis on the presentation itself – not its content.

Main Part

The following sections will deliver parts of a rough transcript of various parts of the presentation. The subsequent subsection will aim to provide an analysis in reference to the transcript.

Transcript. All technical sources from the presentation are from (Arar, 2017; Verhelst & De Koning, 2001; Shannon, 1949; Wannamaker et al., 2000; *Acoustics — Normal equal-loudness-level contours*, 2003; J. O. Smith, 2011; Peceli & Simon, 1996; S. W. Smith, 2002).

"Alright, so just as a preliminary remark: This presentation's going to be a *bit* more technical than the last few. So you're probably not going to understand everything right away – so I'm going to refer you to the handout on *iLearn*, where you can find additional information and sources on the topic — Anyway, today I'm going to talk to you about *psychoacoustic noise-shaping*. I'd usually start off by asking if anyone knows what this is, but quite frankly you probably don't, so we'll just skip that. Before jumping in, let's cover what we'll cover in this presentation …"

* * *

"So why do we need all this? Well, picture the following scenario: You have a digital audio source over here. That could be your smartphone, a computer, Bluetooth, whatever. And on the other side you have an analog audio sink: So headphones or loudspeakers. But now there's a slight problem: How do you go from the digital domain to the analog domain? Any ideas?

'How about a digital to analog converter?'

Correct we need a DAC (digital to analog converter). Now this setup is pretty ubiquitous. Same thing's in every phone, laptop, whatever plays back audio really. Now here's the catch however: DACs need data in a fixed-point format, often something like 16-Bits fixed. However, we like doing digital audio processing on the computer in floating point. Usually something like 32-Bits float. So now we need to convert from float to fixed. "

* * *

"Let's take a look at these signals discrete Fourier-Transform in the frequency domain[1]. On the X-Axis we have our frequency – not numbered, this is just an example. The Y-Axis, again, represents the amplitude. The first signal, the 32-Bit one, looks pretty nice: There are some artifacts of the transformation up here, but we can ignore those. We have this nice bump in here, which represents our fundamental frequency – the sine wave. Let's take a look at the quantized version[2]. It looks good up to the fundamental, but then – what is all this mess up here? This is the distortion which introduces harmonic overtones. "

* * *

`"…What if we take the noise and push it somewhere else? — "`
`"That idea might just be crazy enough…"` [3]

"…to actually work. So picture this, X-Axis, again, the frequency and Y the amplitude [4]. This green area shows the area of the signal we're interested in. The red one shows the noise. So what if instead of having this uniform noise floor, where there's equal amounts of noise at every frequency, we just *push* the noise somewhere else. Now note that we didn't add or remove

[1]This is in reference to Figure 2a in the Appendix.
[2]In reference to Figure 2b in the Appendix.
[3]Text in `monospace` is audio from a video interlude used to reveal the key idea of noise-shaping.
[4]This is in reference to Figure 1 in the Appendix

any of the noise; It's still the same. It's just that the area of noise intersecting our signal is greatly reduced, thus essentially decreasing the noise in our signal. This is the entire concept of noise-shaping. Now how does this actually work …"

Analysis. The presentation starts of in a very nonchalant manner. Addressing the audience to set up expectations of a technical and difficult topic and referencing the previously held presentations. There is no grand introduction or *Bang*. However, this is not in violation of what Powell says with "Don't waste time on long boring introductions."(Powell, 2011, Point 3 on Page 7). As the audience's attention is captured in two ways: Partially from the ominous subtitle of the presentation ("Exploiting biology in digital audio") which isn't addressed directly – An interesting and somewhat intriguing claim. The other part is simply the contrast of the casual appearance with virtually all other presenters up to that point. This is again used in reference to Powells Point 7: "speak naturally" (Powell, 2011, Page 7). It creates a more relaxed atmosphere at the start. Which nicely flows into the signposting, which relies heavily on the rule of three (Williams, 2008, Page 59, A). This is one of the few parts in the presentation where I'm relying on dreaded bullet points: "[…] 'bullets'–little black circles in front of phrases that were supposed to summarize things. There was one after another of these little goddamn bullets […] on the slides."(Feynman, 1988, Page 126–127). The slides feature very little text and bullet points to focus the attention to the presenter and only aid in displaying visuals.

This relaxed atmosphere is then put into stark contrast once the introduction and signposting is over. Demanding the audiences' attention as I introduce a fairly common scenario vividly: Converting a digital signal to an analog signal. By moving around and guiding the attention I can introduce the depicted situation in an understandable and controlled way without relying on animations, but still keeping it interesting. As this is probably the least technical aspect I felt comfortable to include a question to the audience: "How do you go from the digital domain to the analog domain? Any ideas?" This introduces the main problem of the topic: Conversion from a high Bit-Depth to a lower one – Quantization. This problem is relevant in the real world as this scenario is truly ubiquitous. The relevance is again underlined by mentioning how the application for this is found "…in every phone, laptop, whatever plays back audio really". At this point the audience knows what the problem and the context is: Quantization and signal processing respectively.

The presentation goes on with explaining the effects this quantization creates. As visually explaining the differences is immensely more simple, the presentation relies heavily on graphs and diagrams. The third excerpt is in reference to a set of graphs. These are not derived from real world data but rather exaggerated to clearly demonstrate the point I want to make. As such the focus is on readability and ease of understanding. Again referencing Powell with Point 13: "Let your visual speak for themselves"(Powell, 2011, Page 8). However, I didn't want to leave the audience guessing for themselves what a given chart is trying to say. For every graph I gave a quick overview to orient the listener: "On the X-Axis we have our frequency – not numbered, this is just an example. The Y-Axis, again, represents the amplitude."

The midpoint of the presentation starts after explaining the dilemma: We either have noise or distortion, which both are undesirable. The question arises if we can't just do something else to avoid this situation. To which the answer is no, we can't break the laws of physics. The revolutionary revelation of not removing the noise, but simply "push[*ing*] the noise somewhere else" is then revealed though a cartoon interlude. The characters in the altered video each offer ridiculous solutions: A comedic break from the technical topics, which makes strong uses of the visual and auditory senses as defined by the *VAKOG* system in (Williams, 2008, Page 77). This simple concept is then re-stated in a more formal way in the shape of a graph, again using *VAKOG*. For easy reference it's again depicted in Figure 1. Again a quick introduction is given with: "This green area shows the area of the signal we're interested in. The red one shows the noise." An interesting detail here is the use of color. Throughout the presentation the same colors have been used to refer to conceptually similar ideas (Red always being noise and green always being the quantized signal[5]).

The ending of the presentation consists of a quick re-cap by going over some key figures. This re-iterates the journey and all logical steps taken to get to the end result. The slides then end on the obligatory sources. However, as time was still plenty, I decided not to end the presentation here and instead show a concrete example of the topic in action. Showing the source code and running it to see the effect in action on real data further fundaments the understanding as well as demystifies the complex concepts. It wasn't originally planned, but worked out nonetheless, as it acts as a closing *Bang*, as mentioned by Williams on Page 30. This, again, creates a more natural and relaxed environment: It's less of a lecture and more of a conversation.

[5]The colors are taken directly from the *DIT* branding guidelines and are used by the faculties *BIW*, *AWW* and *MB-MK* giving the presentation a homogenous and harmonious look.

Conclusion

I'm quite passionate about the topic of this presentation, as I've had trouble understanding the concept myself for close to year now. Being able to not just understand it myself, but to be able to convey the information in a hopefully clear and concise manner, represents a somewhat important achievement to me. The need for this study assignment gave the motivation to finally finish up on this topic.

The presentation itself was somewhat different as to what I would have done normally. As the presentation is very technical, I would usually simplify the concepts massively. However, as my audience was knowledgeable in the field, I felt I could get away with diving into the interesting but complex technical details. The added structure provided by the signposting, *VACOG* system, etc. was certainly helpful. Although the concepts themselves weren't particularly new to me, they provide a more formal insight into not just how to make a presentation interesting but into why these techniques work. I've definitely enjoyed the process of creating and holding this presentation.

* * *

Bibliography

Acoustics — normal equal-loudness-level contours (Norm No. ISO 226.2003). (2003). International Organization for Standardization. Retrieved November 21, 2022, from `https://www.iso.org/standard/34222.html`

Arar, S. (2017, September 29). *Design of fir filters using the frequency sampling method* (tech. rep.). All About Circuits and EETech Media, LLC. Retrieved November 20, 2022, from `https://www.allaboutcircuits.com/technical-articles/design-of-fir-filters-using-frequency-sampling-method/`

Feynman, R. P. (1988). *What do you care what other people think? Further adventures of a curious character* (R. Leighton, Ed.). W. W. Norton.

Peceli, G., & Simon, G. (1996). Generalization of the frequency sampling method. *Quality Measurement: The Indispensable Bridge between Theory and Reality (No Measurements? No Science!) Joint Conference - 1996: IEEE Instrumentation and Measurement Technology Conference and IMEKO Tec*, *1*, 339–343. `https://doi.org/10.1109/IMTC.1996.507403`

Powell, M. (2011). *Presenting in english: How to give successful presentations*. Heinle Cengage Learning. Retrieved December 29, 2022, from `https://books.google.de/books?id=PrpKmgEACAAJ`

Shannon, C. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10–21. `https://doi.org/10.1109/JRPROC.1949.232969`

Smith, J. O. (2011). *Spectral audio signal processing* [online book, 2011 edition]. W3K Publishing. Retrieved November 21, 2022, from `https://ccrma.stanford.edu/~jos/sasp/`

Smith, S. W. (2002). *Digital signal processing: A practical guide for engineers and scientists*. California Technical Publishing. Retrieved December 27, 2022, from `http://www.dspguide.com/pdfbook.htm`

Verhelst, W., & De Koning, D. (2001). Noise shaping filter design for minimally audible signal requantization. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, 147–150. `https://doi.org/10.1109/ASPAA.2001.969564`

Wannamaker, R., Lipshitz, S., Vanderkooy, J., & Wright, J. (2000). A theory of nonsubtractive dither. *IEEE Transactions on Signal Processing*, *48*(2), 499–516. `https://doi.org/10.1109/78.823976`

Williams, E. (2008). *Presentations in english: Find your voice as a presenter*. Macmillan. Retrieved December 29, 2022, from `https://books.google.de/books?id=eCDpGAAACAAJ`

Appendix

Figure 1: Noise-spectrum with and without noise-shaping.

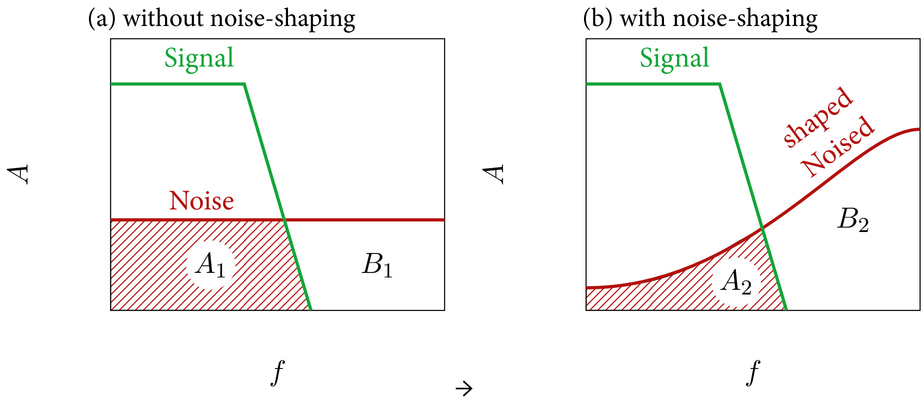
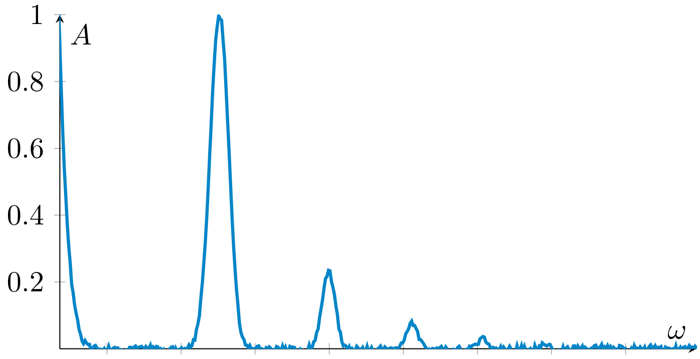
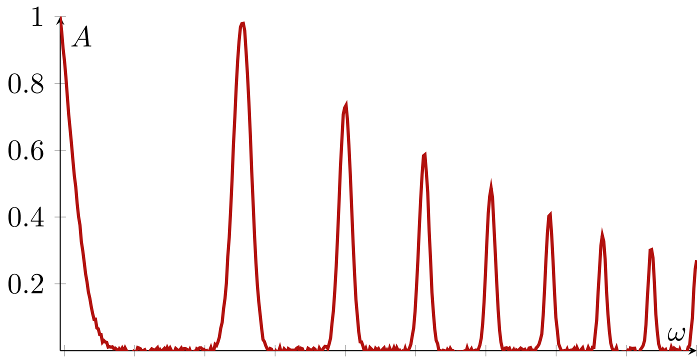


Figure 2: Fourier Transforms of different signals.

(a) DFT of the original signal.



(b) DFT of the quantized signal.



(c) DFT of the dithered signal.