Psychoacoustics of Soft Clipping and its Perception as a 'Clean' Type of Distortion

Mario Alberto Vallejo Reyes
School of Physics, Engineering and Technology
University of York
York, United Kingdom
ckb535@york.ac.uk

***Abstract*—This study examines the psychoacoustics of soft clipping and its perception as a clean type of distortion. It begins with an exploration of various types of distortions through background theory, followed by a review of listening tests from the study by Tan, Moore, and Zacharov[1]. MATLAB is then utilized to synthesize different types of distortions using custom functions created for this research, and time/frequency analyses are performed on these distortions. The findings from the frequency domain analysis, specifically waveform and spectrogram analyses, of the synthesized distortion types, objectively reinforce the conclusions of the reviewed listening tests. They demonstrate that soft clipping is the distortion type among those analyzed that most closely resembles the original signal. This supports the notion of soft clipping as a comparatively clean form of distortion, providing valuable insights for applications in audio processing and sound design.**

I. Introduction

This work investigated a specific psychoacoustic phenomenon: the perception of soft clipping as a form of 'clean' distortion compared to other types of audio distortion. In this context, 'clean' referred to the distortion that listeners perceived as closest to an original completely undistorted audio signal.

The approach began with a review of relevant literature, focusing on different types of audio distortions and presenting listening tests conducted in this area. Next, MATLAB was utilized to synthesize examples of various distortion types. Finally, this work presented time/frequency-domain analyses of these examples to illustrate how soft clipping retained its cleanliness, mainly when applied at higher intensities.

II. Background Theory

A. Some Types of Distortion

1) Hard Clipping: Hard clipping, as discussed by D. Reiss and McPherson[2], occurs when an audio signal surpasses the maximum processing capacity of either digital or analog systems, resulting in parts of the signal that exceed these limits being effectively 'capped', as seen in Equation 1:

$$f(x) = \begin{cases} -1 & \text{if } Gx \leq -1 \\ Gx & \text{if } -1 < Gx < 1 \\ 1 & \text{if } Gx \geq 1 \end{cases} \tag{1}$$

where x denotes the input signal and G represents the applied gain. The function outputs -1 when Gx is less than or equal to -1, and 1 when Gx is greater than or equal to 1. For Gx values between -1 and 1, the output is Gx, indicating the absence of clipping.

This phenomenon results in a sudden and stark transition in the signal's waveform from unclipped to clipped parts, as observed in Figure 1. Such an abrupt change contributes to a sharper, more intense sound character, often perceived as harsh.

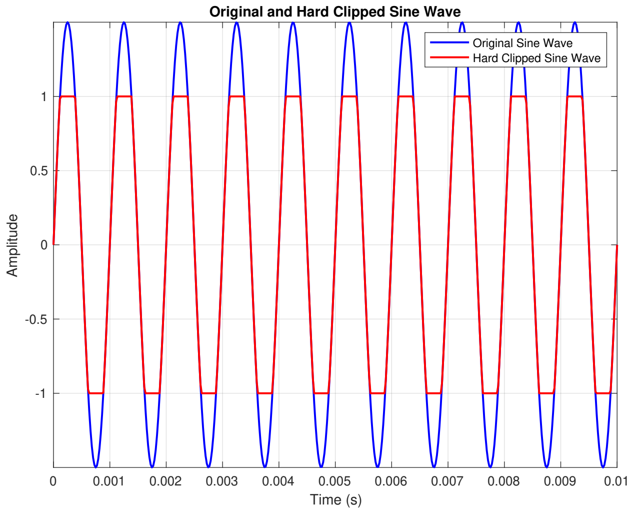

Fig. 1. Depiction of hard clipping. The blue line represents the original sine wave and the red line illustrates the sine wave subjected to hard clipping.

2) Soft Clipping: As explained by Creasey[3], unlike hard clipping, which introduces a sudden discontinuity in the signal once a certain threshold is exceeded, soft clipping gradually transitions the signal as it approaches and surpasses the threshold, resulting in a less abrupt distortion.

Equation 2 can characterize a mathematical model for soft clipping:

$$y = \frac{\tanh(cx)}{\tanh(c)} \quad \text{where} \quad -1 \leq x \leq 1 \quad \text{and} \quad c > 0 \tag{2}$$

where y represents the output signal after the soft clipping has been applied to the input signal x. The hyperbolic tangent function, $\tanh$, is used to limit the amplitude of the output signal smoothly. For values of x within the range of $-1 \leq x \leq 1$, the function $\tanh(cx)$ produces an output y that follows the input closely. However, as x moves outside of this range, the hyperbolic tangent function ensures that y approaches 1 for $x > 1$ and approaches -1 for $x < -1$, thus avoiding the abrupt cutoff characteristic of hard clipping.

The constant c in $\tanh(cx)$ controls how quickly the output signal saturates. A larger value of c makes the transition from the unclipped to the clipped signal more abrupt, which can sound more aggressive, resembling hard clipping. Conversely, a smaller value of c yields a softer transition, preserving more of the signal's dynamic range before saturation.

Finally, the division by $\tanh(c)$ effectively scales the output. This ensures a more controlled and predictable response to varying x and c values,

In musical contexts, depending on the input gain and the specific shape of the transfer function used, this gradual quality of soft clipping is desirable because it allows for more detailed control over the distortion effect, as illustrated in Figure 2.

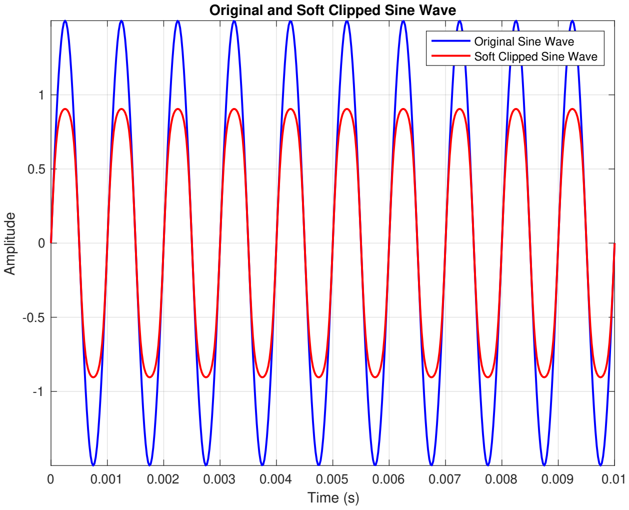

Fig. 2. Depiction soft clipping. The blue line represents the original sine wave and the red line demonstrates the effect of soft clipping.

3) Centre Clipping: Centre clipping, as described by Giannakopoulos and Pikrakis[4], is a signal processing technique where an audio signal is modified based on a predefined threshold. This technique involves analyzing and retaining only those samples whose absolute value meets or exceeds the threshold. Samples that fall below this threshold are set to zero. The application of centre clipping is mathematically represented in Equation 3:

$$x_c(n) = \begin{cases} x(n) - T_h, & \text{if } |x(n)| \geq T_h, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $x_c(n)$ represents the resulting signal after the application of centre clipping at the n-th sample. The function $x(n)$ denotes the value of the original signal at sample n, and T_h is the threshold which determines the level at which clipping is engaged. If the absolute value of $x(n)$, indicated by $|x(n)|$, is less than the threshold T_h, then the output $x_c(n)$ is set to zero. Conversely, if the magnitude of the input signal meets or exceeds T_h, the signal passes through unaltered, as seen in Figure 3.

4) Wave Shaping and Full Range Distortion: Wave shapers, as explored by Roey Izhaki[5], are fundamental audio processing tools that apply a transfer curve to a signal according to a specific predefined function, offering immediate amplitude transformation. One application of wave shaping is creating full-range distortion. The following Equation 4 represents a common form of this transfer function:

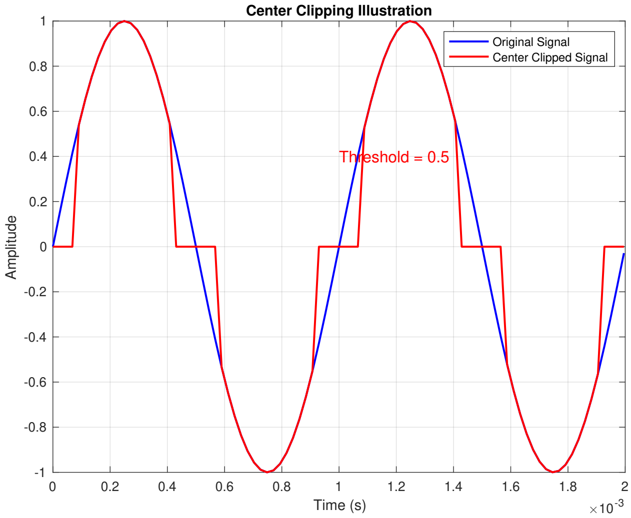

Fig. 3. Audio signal before and after centre clipping, where amplitudes below 0.5 are zeroed.

$$y(t) = \alpha x(t)^{\beta} \tag{4}$$

where $x(t)$ is the input signal, $y(t)$ is the output signal, α is a scaling factor, and β is the exponent determining the degree of nonlinearity. This equation showcases how input amplitudes are non-linearly mapped to output amplitudes, as illustrated in Figure 4.

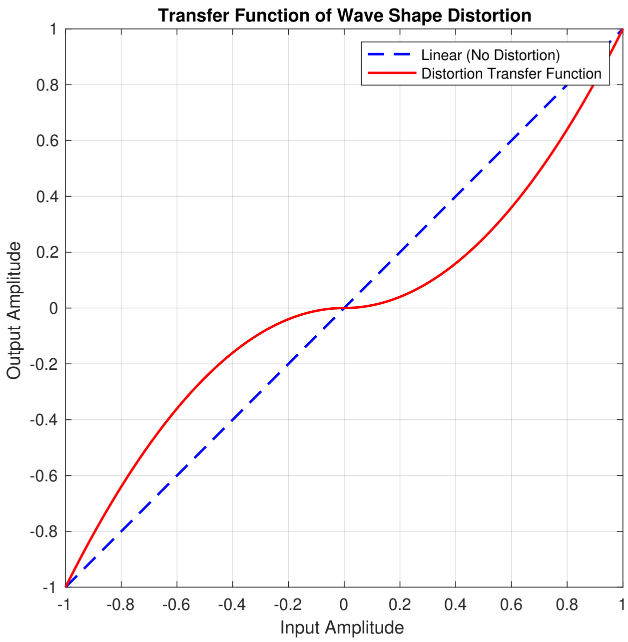

Fig. 4. Transfer function of a full-range distortion wave shaper. This graph illustrates the nonlinear mapping of input amplitudes to output amplitudes, with the red line depicting the distortion transfer function and the blue dashed line representing the linear undistorted signal.

Wave shapers can be employed subtly for minor enhancements to the audio, adding a touch of complexity or warmth, or used more aggressively to create dramatic and creative effects. This range of applications makes them a valuable tool for audio engineers and producers looking to experiment with the texture and character of sound.

B. Symmetrical vs Asymmetrical Clipping

The concepts of symmetrical and asymmetrical clipping, key forms of signal modification in audio distortion, are studied in Kevin Robinson's analysis[6]. Symmetrical clipping uniformly reduces the peaks of an audio signal. On the other hand, asymmetrical clipping applies uneven clipping to the signal waveform, clipping one side more heavily than the other. This results in a mixture of odd and even harmonics, giving the sound a distinct character. These techniques open up various experimental possibilities for sound manipulation, each with unique sonic qualities.

C. Perception of different types of distortions

In their study, Tan, Moore, and Zacharov[1] explored the impact of different types of nonlinear distortion on how listeners perceive the quality of speech and music signals. The researchers utilized various types of distortions, including hard and soft symmetrical and asymmetrical clipping, centre clipping, and full-range waveform distortion. This full-range distortion involved altering the waveform by raising its instantaneous absolute value to a power while preserving the waveform's sign.

Participants in the study were asked to rate the level of perceived distortion on a scale from 1 to 10, with 1 indicating the most distortion and 10 the least. The research was conducted in two phases: In the first, distortions were applied to broadband signals, while in the second, they were applied to signal subbands.

The study involved applying various distortions over the audio signals at different intensities. For this work on the psychoacoustics of soft clipping, the focus was on the particular condition where distortions were applied at their maximum intensity for this study. Under these conditions, as seen in Figure 5, the findings revealed that among the different types of distortion, soft clipping was perceived as the cleanest, causing only minimal alterations in the ratings. This contrasted with full-range distortion, which was perceived as the most severe.

Interestingly, the subjective assessments made by the listeners aligned well with objective distortion measurements. These objective measures, denoted as DS, were based on the output spectrum of each nonlinear system in response to a multitone signal. The study found a high negative correlation between these objective measures and the subjective ratings, indicating that larger values of DS (denoting more distortion) corresponded to lower subjective ratings (indicating a perception of more distortion).

An additional experiment further confirmed the relationship between subjective perceptions and objective measurements. In this phase, the stimuli with nonlinear distortion were produced by recording the outputs of real transducers, which were then digitally filtered to minimize amplitude-frequency response irregularities. The results showed a moderately strong negative correlation between subjective ratings and the objective DS measure.

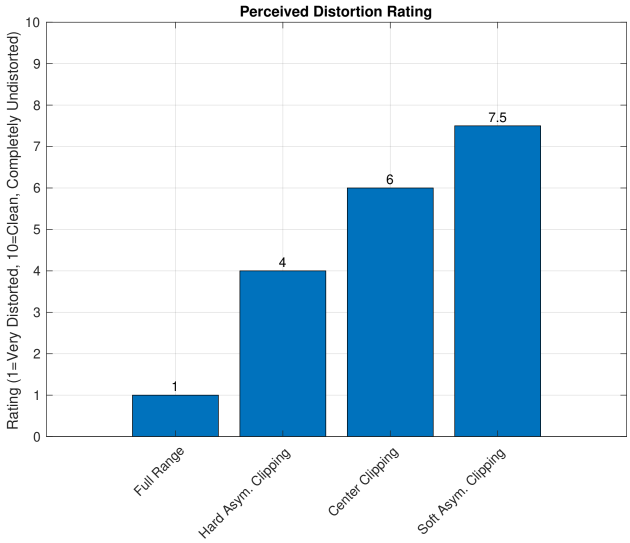

Fig. 5. Perceived distortion ratings. This bar chart ranks various distortions with ratings ranging from 1 (very distorted) to 10 (clean, completely undistorted).

III. Design and Implementation

The custom function `applySoftAsymmetricalClipping` was developed to synthesise soft clipping. This function processes an audio signal by employing the arctangent function, which ensures a smooth and gradual distortion characteristic of soft clipping. It operates asymmetrically, affecting only the positive peaks that exceed a calculated threshold. The function maintains a consistent peak amplitude throughout the process by normalising the signal before and after applying the clipping.

In the specific example synthesised for this work, music material (specifically the song 'Sakura[7]' by Artist Ruddi Nizz) was processed through the `applySoftAsymmetricalClipping` function with a threshold set to 0.9 and an intensity affecting the signal 10% of the time. The selection of '10 percent of the time' as the threshold frequency was chosen to replicate the conditions under which soft clipping was perceived as the cleanest form of distortion in the previously reviewed listening tests, thereby allowing for a practical examination and comparison in line with those observations.

Bespoke functions were developed to apply hard asymmetrical clipping, full-range distortion, and centre clipping to the same music material to conduct a comparative analysis with soft symmetrical clipping. The arguments in these functions were inputted to replicate the specific conditions in the research by Tan, Moore and Zacharo, where distortions were applied at the maximum intensities. Specifically, hard asymmetrical clipping was configured to affect 10 percent of the signal's duration, full-range distortion was implemented with an alpha parameter set to 2 and centre clipping was adjusted to impact 10 percent of the signal's RMS value.

In order to visually illustrate the psychoacoustic phenomenon where soft clipping is perceived as relatively clean, Figure 6 presents four subplots. Each subplot displays the resulting spectral differences obtained by summing the original audio mix with its distorted counterpart. Before this summation, both the original and the distorted versions were normalised in terms of their root mean square (RMS) values, and the phase of one signal was inverted. This inversion is crucial as it emphasises frequency discrepancies over amplitude disparities. The subplot contrasting the original mix with the soft-clipped version (top left) reveals a minimal spectral difference, particularly at transient peaks. This observation aligns with the characteristic impact of soft clipping, which, despite affecting the entire signal, predominantly influences the transients. The psychoacoustic effect of soft clipping being perceived as relatively clean is further reinforced by comparison with other subplots. These additional plots demonstrate the differences between the original mix and versions subjected to hard clipping, full-range distortion, and centre clipping. In these cases, the spectral deviations from the original mix are more pronounced, highlighting how these distortions are more readily perceived as distorted.

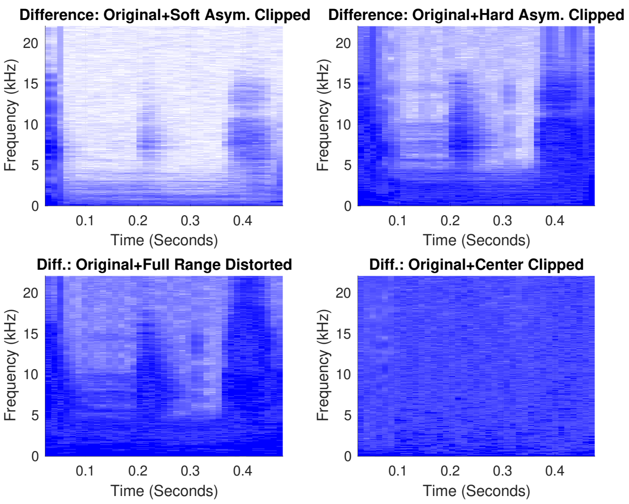

Fig. 6. This figure shows the results of summing the original audio mix with its inverted-phase, distorted counterparts, revealing the distinct spectral changes induced by each type of distortion.

An alternative approach to demonstrate the relatively unobtrusive nature of soft asymmetrical clipping involves analysing the waveform alterations through various distortion processes. In preparing the comparative analysis, all signals were first normalised to their root mean square (RMS) values. Notably, despite normalisation, the signal subjected to full-range distortion exhibited a reduced amplitude compared to the original. For comparative clarity, its amplitude was subsequently doubled post-normalisation. Figure 7 showcases four plots, each representing a segment of the original mix's waveform juxtaposed with its counterpart processed through different distortion techniques.

The comparison of the clean mix versus the soft-clipped mix is illustrated in the top-left plot. Here, the waveforms are nearly indistinguishable at a glance, with minor discrepancies appearing only upon very close inspection, particularly at the peaks where they do not perfectly overlap. The top-right plot contrasts the original mix with the hard-clipped version. Although the waveforms appear similar overall, the hard-clipped waveform exhibits notable variations, especially at points distant from zero amplitude.

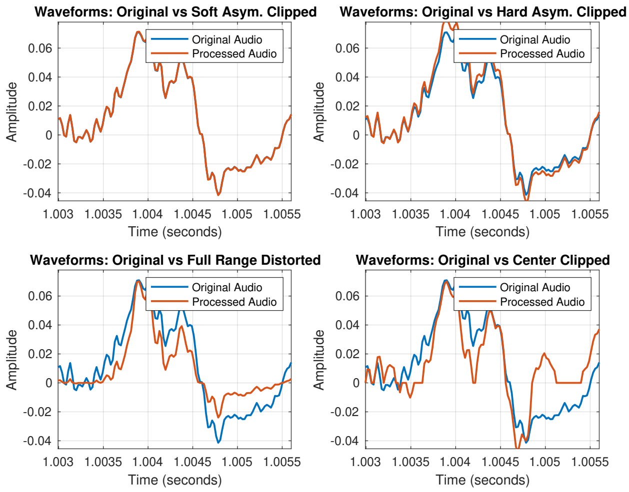

Fig. 7. Waveform comparisons of original audio with various distortion techniques.

The bottom-left plot displays the original mix alongside the full-range distorted version. This comparison reveals a stark contrast between the two waveforms, with many segments being compressed towards the zero line, resulting in a marked loss of dynamic range. Finally, the bottom-right plot compares the original mix with the center-clipped signal. This plot demonstrates the most significant divergence between the original and distorted signals. Extensive portions of the waveform are altered, with segments consistently flatlining at zero amplitude and the emergence of new forms, thereby significantly altering the signal's characteristics.

IV. Evaluation

The primary objective of this work was to investigate the psychoacoustics of soft asymmetrical clipping, particularly its interpretation as a 'clean' form of distortion in contrast to other distortion types. The review of pertinent literature provided essential insights before studying listening tests and performing time-frequency analyses.

The examination of existing listening test research provided insights into listener perceptions of audio distortions. This analysis highlighted the importance of integrating subjective experiences with objective metrics, offering a detailed perspective on audio evaluation.

The development of MATLAB scripts for synthesizing soft clipping, hard clipping, central clipping, and wave shaping full-range distortion was a critical technical aspect of this work. These scripts successfully generated the intended types of distortions, each distinct and representative of their respective categories in audio processing. The accuracy of these distortions was crucial, as they served as foundational elements for the subsequent time-domain and frequency-domain analyses.

The effectiveness of using both time-domain (waveform) and time-frequency-domain (spectrogram) analyses in this work was instrumental in highlighting the clean nature of asymmetrical soft clipping. Waveform analysis was a key step, revealing a striking resemblance between the waveforms of the original mix and the soft-clipped version.

Simultaneously, spectrogram analysis provided a complementary perspective, offering a detailed view of the frequency content changes induced by soft clipping. This analysis further supported the findings from the waveform analysis, showing only slight deviations in the frequency domain when the original signal was soft-clipped. The combination of waveform and spectrogram analyses served as a validation of each method's findings.

Additional listening tests and time-frequency analyses involving a broader range of distortions are recommended to expand upon the psychoacoustics of distortion. It is also imperative to extend these tests to various sound types beyond vocals and music, such as ambient sounds, foley effects, immersive audio material and others, to fully understand the impact and perception of different distortions across diverse audio content.

References

[1] C.-T. Tan, B. Moore, and N. Zacharov, "The effect of nonlinear distortion on perceived quality of music and speech signals," *Journal of the Audio Engineering Society*, vol. 51, pp. 1012–1031, 11 2003.

[2] J. D. Reiss and A. McPherson, *Audio effects: Theory, implementation and application*. Boca Raton, FL: CRC Press, 2014.

[3] D. Creasey, *Audio processes: Musical analysis, modification, synthesis, and control*. London, England: Routledge, 2016.

[4] T. Giannakopoulos and A. Pikrakis, *Introduction to Audio Analysis: A MATLAB (R) Approach*. London, England: Academic Press Inc. (London), 2014.

[5] R. Izhaki, *Mixing audio: Concepts, practices, and tools*, 4th ed. Oxford, England: Focal Press, 2023.

[6] K. Robinson, *Practical audio electronics*. London, England: CRC Press, 2020.

[7] RuddiNizz, "Ruddi nizz- sakura (video oficial)," Aug. 2023.