AI-TD: Analog transient shaping with digital envelope followers and AI-driven parameter setting

Kónya, Kristóf - kristof.konya@outlook.com
Tschavoll, Jakob - jakob@tschavoll.at

Concept

Prevalent throughout the art and industry of recording musical signals, compression is one of the most popular and necessary signal processing tools for the refinement of the recorded track's level and transients. Traditionally, an analog or digital compressor shapes a given signal's level based on an absolute threshold of volume. While this is useful for controlling tracks with great volume, it omits compression as a psychoacoustic parameter beneficial for transients which are smaller than the selected threshold but would nonetheless profit from its effects. To combat this issue, SPL electronics GmbH introduced the "Transient Designer" in 1998, a dynamic compressor which obtains its amount of compression from two envelope curves of the input signal with different time constants. If a slow adapting envelope is subtracted from a fast adapting one, the obtained signal acts a control voltage (CV) for the attack phase of a transient, see figure 1. With parametric gain applied to this signal, the attack of a transient can be shaped regardless of absolute level. The original device is fully analog and very complex to understand and design, due to the challenge of optimizing the time constants for the envelopes. For this reason, we propose a cheaper alternative which calculates and parametrizes the time constants digitally and shapes a given analog signal via a signal amplifier connected to a DAC of an STM32H7 microcontroller and propose a small deep neural network to extract optimal time constants from a given audio signal.

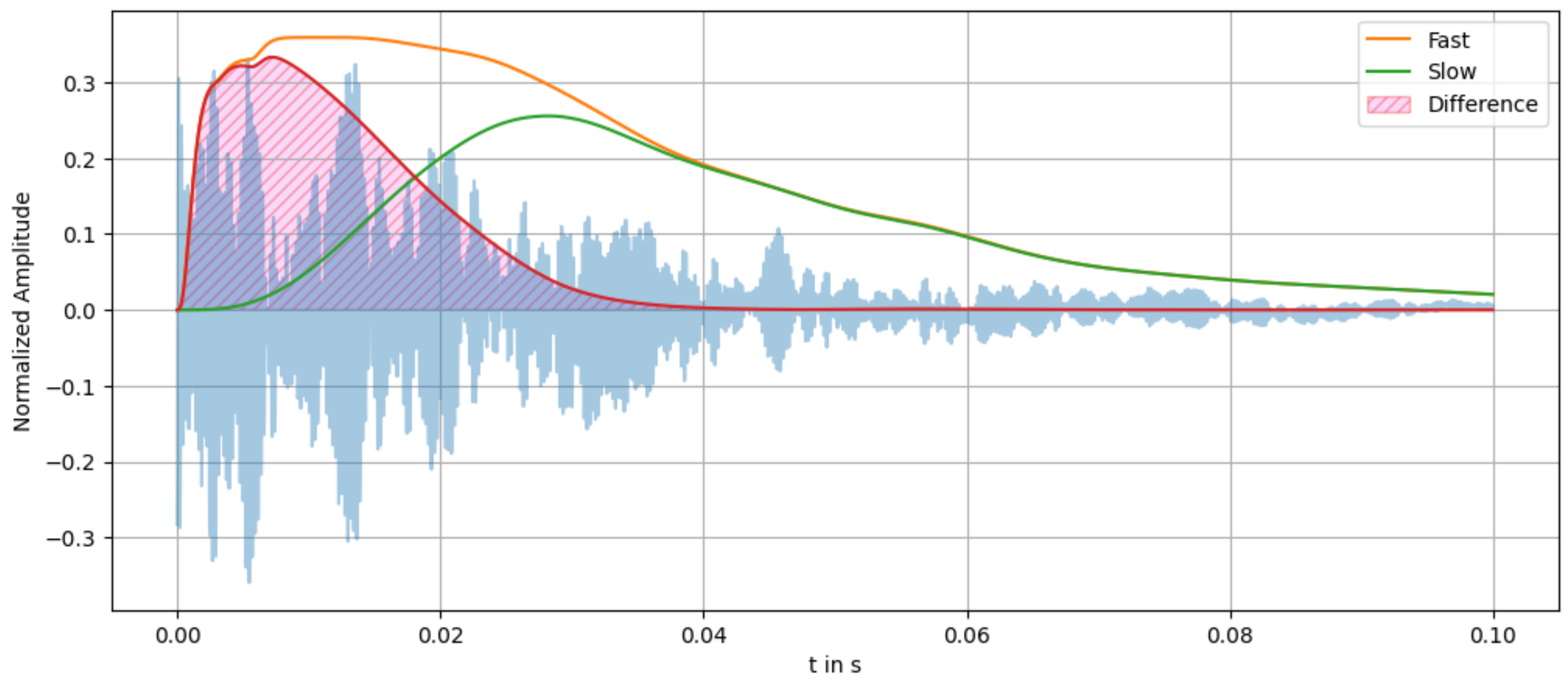


Fig. 1: Envelope difference generation.

Hardware

The device is encased in a guitar pedal format and powered by a typical 9V DC power supply. A stereo analog VCA design based on the 2164 is controlled by a Daisy Seed housing a STM32H7 running at 480MHz. The control voltage for the VCA is generated by the digitally implemented envelope followers. The digital output of the envelope followers is then converted via the STM32 onboard DAC into a analog voltage. Thus enable changes in the dynamics of the signal by changing the VCAs amplification factor. The incoming audio signal is sampled at 96k by the WM8731 codec.

Fig. 2: AI-TD Hardware in stylized guitar pedal format.

Software

Transient DSP

A necessary part of the software is the digital envelope creation, which is made possible by the combination of a peak-hold system and a cascade of exponential smoothers. A fixed set of staged exponential smoothers create envelope curves based on the setting of the time constants [1]. This process is done in two iterations for the attack and sustain phase each: A fast envelope follower estimates the "true" envelope of a given signal and a slow one is used to compute the difference between the two envelopes. By varying the time constants of the envelopes, the length of the CV can be changed for transient boost or attenuation. This gives the user the ability to apply compression to more abrupt or sustained signals, a feature which makes it possible to apply settings for a given instrument more effectively than traditional compressors or even the original "Transient Designer".

TauNet

Since we expose the parametrization of time constants, we also need to offer a default value for these parameters. For this we rely on automatic setting of the time constants based on a vector of audio features of the input signal. By playing a signal through the device while holding one of the stomp buttons, the signal's audio features (see section *Audio features*) are extracted from the signal and used to characterize it. To establish a connection between partially uncorrelated audio features and time constants for the envelope of attack and release, we propose a small, deep neural network called TauNet, optimized in size and simplicity of layers to be deployed on the MCU by usage of the RTNeural inference engine. It features five layers of dense layers of exponentially decreasing sizes with rectified linear units per layer.

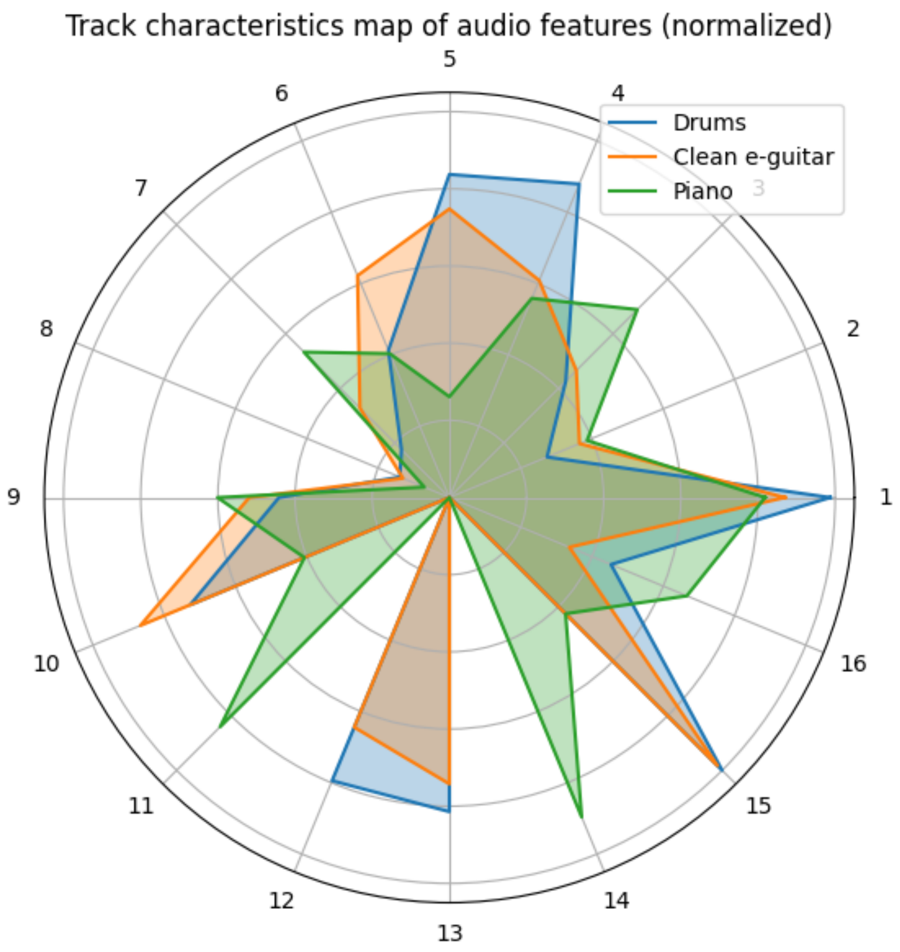
The regression output of the model obtained a RMSE value of 0.14, which lies within the parameter's unhearable tolerance range when de-normalized, as such an error represents tiny wiggles when compared to the physical UI.

Audio features

The data used to train and use TauNet are basic audio features from traditional signal processing. To remain in a one-dimensional feature space, all features are scalar values and represent the entire signal. If a feature also represents a change in time, the mean value of that feature is computed. Figure 3 demonstrates the audio features and their different normalized values per track type.

Dataset generation

Since we set parameters for compression in a musical context, the ground truth is inevitably based on opinion. To obtain optimal settings for the time constants nonetheless, we gave the hardware to experienced producers in order to record the settings they applied to a fixed set of five second recordings of varying release years and genres. The derived "human" features, attack gain and sustain gain were used as additional inputs, while attack time constant and sustain time constant were used as outputs for training. The values themselves were obtained by probing the device during the usage by the producer and printing them to a serial terminal, from which they were recorded into a table.



1	Tempo
2	Average signal rise time
3	Average signal fall time
4	Spectral centroid
5	Bass band energy
6	Mid-bass band energy
7	Mid-treble energy
8	Treble energy
9	Crest factor
10	Spectral flux
11	Attack attenuation (human input)
12	Attack amplification (human input)
13	Sustain attenuation (human input)
14	Sustain amplification (human input)
15	Attack time constant (human output)
16	Sustain time constant (human output)

Fig. 3: Spiderweb plot of audio features for different track types.

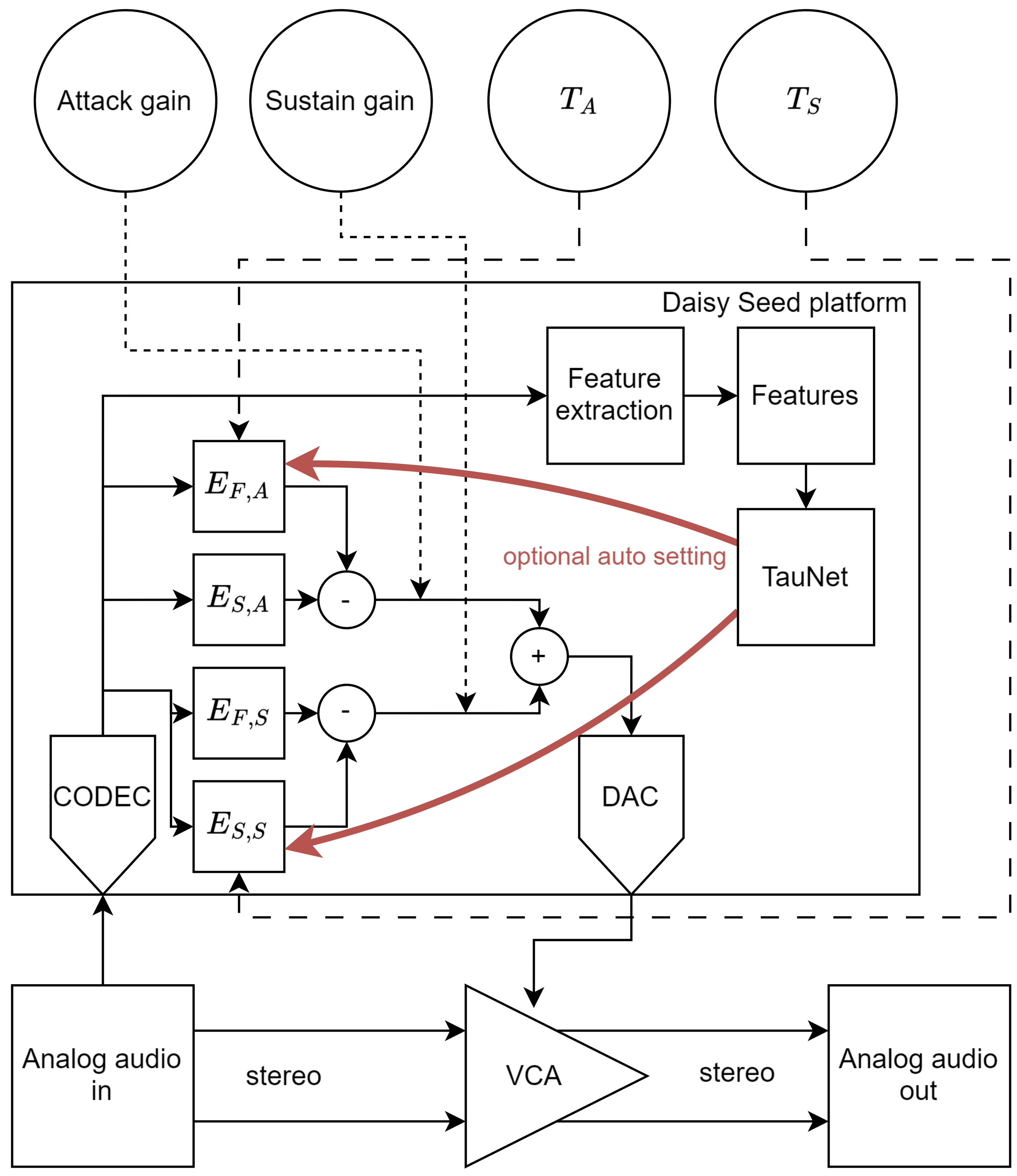


Fig. 4: Functional diagram for the AI-TD program flow. Envelopes E are computed from the signal and time constants T by either the user or TauNet.

Conclusion

The presented work proposes hardware and software necessary to create a transient shaping device which is cheaper than its fully analog counterpart and with automatic parameter control derived from a deep neural network. The work shows how to apply dynamic compression by calculating envelopes digitally while still keeping a fully analog audio path. The additional automatic parameters setting works by analyzing audio features and running inference directly on chip, limiting the barrier some users might feel when using the additional parameters. With this work we challenge what can be done by combining analog audio, digital signal processing and machine learning in a musical context while also remaining cheap and expandable and entirely within an MCU ecosystem.

[1] Sanfilippo, Dario: *Envelope Following via Cascaded Exponential Smoothers For Low-Distortion Peak Limiting and Maximisation*, 2022. DOI: 10.5281/zenodo.6577136

https://kristof-konya.dev/projects/AITD/paper