FEATURE DESIGN FOR THE CLASSIFICATION OF AUDIO EFFECT UNITS BY INPUT / OUTPUT MEASUREMENTS

Felix Eichas, Marco Fink, Udo Zölzer

Department of Signal Processing and Communications,
Helmut-Schmidt-Universität
Hamburg, Germany
felix.eichas@hsu-hh.de

ABSTRACT

Virtual analog modeling is an important field of digital audio signal processing. It allows to recreate the tonal characteristics of real-world sound sources or to impress the specific sound of a certain analog device upon a digital signal on a software basis. Automatic virtual analog modeling using black-box system identification based on input/output (I/O) measurements is an emerging approach, which can be greatly enhanced by specific pre-processing methods suggesting the best-fitting model to be optimized in the actual identification process. In this work, several features based on specific test signals are presented allowing to categorize instrument effect units into classes of effects, like distortion, compression, modulations and similar categories. The categorization of analog effect units is especially challenging due to the wide variety of these effects. For each device, I/O measurements are performed and a set of features is calculated to allow the classification. The features are computed for several effect units to evaluate their applicability using a basic classifier based on pattern matching.

1. INTRODUCTION

Audio effect units are used by musicians or sound engineers to transform the signal of an (electrical) instrument in order to modify it in a certain way. Many of these effect units already have been emulated and digital models have been derived to apply these effects while, for example, mixing a recording [1]. Many musicians value *vintage* music equipment because of its specific tonal characteristic and want to recreate this characteristic without spending a large amount of money on rare vintage equipment. With the aid of system identification, the unique properties of an effects unit can be captured and emulated as a *VST plugin* or with a DSP-based effects unit in a live-setup.

A lot of effect units have been emulated in very different ways. Modeling via circuit analysis, as done by [2–5], is realized by transforming the schematic of the device into a digital model. This procedure is very precise and can describe effect units accurately. But it also has several drawbacks. The computational load is very high, because one or more nonlinear equations have to be solved iteratively for every nonlinear element in the circuit and every sample of the input data. In addition, the characteristic of every nonlinear circuit element has to be known or assumed to solve the nonlinear equations.

An alternate way of emulating analog audio units is system identification with input/output (I/O) measurements. By sending specifically designed input signals through the device under test (DUT) and measuring the output, the influence of the system on the input signal can be determined and recreated with a digital model to achieve the same characteristics [6].

Black-box identification is still an important topic and there are countless contributions to this domain from neural networks to wavelet-transform based models for identification. Sjöberg et al. published a summary of different nonlinear modeling techniques [7]. A method, that has been successfully used to model nonlinear audio systems, was introduced by Novak et al. [8]. They used a block-oriented Hammerstein model for the identification of a distortion guitar effect pedal.

Feature extraction from audio data is an important topic for music information retrieval (MIR) [9]. Nevertheless, typical MIR features like Zero Crossings or Spectral Flux can not be utilized to classify the subtle tonal characteristic of specific audio effect devices.

The majority of commercial digital audio effects, emulating a specific device, are parametric digital models which are usually tweaked by a professional sound engineer to approach the sound of the analog unit. This identification procedure can be automated using black-box modeling. The proposed feature set can facilitate the decision which digital model is best-suited to reproduce the characteristics of the analog effects unit under test. For this purpose specifically designed input signals are generated, replayed through the device, and, by recording the output, the influence of the system on these signals is measured and different characteristics are extracted.

In future work, the extracted features can be used to classify the DUT and choose an appropriate model from a model set. The classification can be done in several ways, e.g. with a neural network or a weighted decision tree. Once a model is chosen, an identification algorithm can be used to fit the model's characteristics to resemble the DUT.

This paper describes the signal model for several typical effects in section 2, the input signals designed for the feature computation are shown in section 3, and the computation of the features is explained in section 4. In section 5 the measurement-setup is shown and the features are evaluated by measuring some typical effect pedals and comparing the resemblance of the pedal characteristic and the computed feature. The usability of the features is demonstrated using a simple classifier in section 6. Section 7 concludes this paper.

2. SIGNAL MODEL

This section describes the influence of typical effects, as categorized in Fig. 1, on an input signal $x(t)$. This analysis was done to

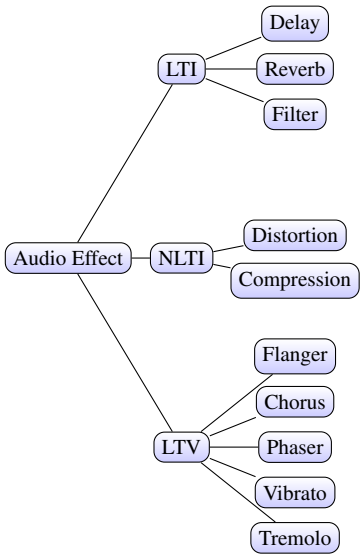


Figure 1: *Categorization of typical digital audio effects into linear time-invariant (LTI), non-linear time-invariant (NLTI), and linear time-variant (LTV).*

design proper input signals and select adequate features, calculated from the recorded (output) signals.

2.1. Linear Time-Invariant Effects

2.1.1. Filter

In the field of digital audio effects, a filter is a linear time-invariant (LTI) system, which is able to amplify or attenuate certain frequency regions of the input signal $x(t)$. The impulse response $h(t)$ defines the characteristic of the filter and is convolved with the input signal to produce the filtered output,

$$y(t) = x(t) * h(t). \tag{1}$$

2.1.2. Reverberation

Although, the mathematical representation of the reverberation effect

$$y(t) = x(t) * h_{\text{rev}}(t), \tag{2}$$

namely the convolution of the input signal with an impulse response h_{rev}, is very similar to the filter effect, those effects are easy to distinguish from each other. The impulse response of filters and reverberations differ significantly in length and noisiness and hence can be differentiated with measurements.

2.1.3. Delay

Repetitions of an input signal can be achieved using delay pedals. The output signal

$$y(t) = x(t) + g \cdot y(t - t_d) \tag{3}$$

can be modeled with a direct path and a delay line in the feedback path. The delay time t_d defines the temporal distance between two repetitions while the amplitude of the repetitions is controlled with the feedback gain g. It should be noted that this is a simple model which does not take into account that many delay effects employ some modulation or filtering in the feedback path to create a certain kind of tonal characteristic.

2.2. Non-Linear Time-Invariant Effects

2.2.1. Compression

Compression is a non-linear effect reducing the dynamic range of the input signal. Therefore, the input signal $x(t)$ is fed to time-variant variable-gain amplification stage, weighting $x(t)$ with a gain factor $g(t)$ to produce the output

$$y(t) = g(t) \cdot x(t). \tag{4}$$

The variable-gain amplifier can be modeled as an envelope signal $x_+(t)$ smoothed with a signal-dependent lowpass filter $\text{LP}_{\text{AT/RT}}$ like

$$g(t) = \text{LP}_{\text{AT/RT}}\{x_+(t)\}, \tag{5}$$

where $\text{LP}_{\text{AT/RT}}$ defines a lowpass filter for the attack (AT) and release (RT) case. Typical choices for the envelope signal $x_+(t)$ are the peak signal $|x(t)|$ or the *root-mean-square* signal $x_{\text{RMS}}(t)$, depending on the type of compressor.

2.2.2. Distortion

Distortion effects modify the input signal $x(t)$ with a nonlinear function $f(x)$ mapping the level of the input signal $x(t)$ to the level of the output signal $y(t)$, as shown in

$$y(t) = f(x(t)). \tag{6}$$

The shape of the nonlinear function defines the tonal quality of the effect. Musicians tend to sub-categorize distortion effects in overdrive, distortion, and fuzz in ascending order of nonlinearity. For an accurate signal model, representing an analog distortion effect unit, there should be additional input and an output filters. Here they are omitted for the sake of readability.

2.3. Linear Time-Variant Effects

Linear time-variant effects, often called *modulation* effects, modulate the input signal in terms of volume, frequency or phase using a low frequency oscillator (LFO) controlling the rate of the modulation.

2.3.1. Tremolo

The tremolo effect was introduced in the 1950s by companies like Fender or Vox. Initially it was an electronic circuit integrated in the guitar amplifier which modulated the volume of the output signal periodically. The modulating signal $x_{\text{LFO}}(t)$ is multiplied with the input signal $x(t)$

$$y(t) = x_{\text{LFO}}(t) \cdot x(t). \tag{7}$$

$x_{\text{LFO}}(t)$ is a periodic signal having, for example, sinusoidal, triangular, sawtooth or square-wave characteristic.

2.3.2. Vibrato

Vibrato is, like the tremolo, one of the earliest common guitar effects. The effect originated from the so called *whammy bar*, a mechanical lever typically located at the bridge of the guitar, which modulated the tension of the strings and therefore the frequency of the guitar tone. Another popular effect unit was called the *Leslie speaker*, where a rotating speaker created frequency modulation due to the Doppler effect. This behavior is recreated in effect units by producing an output signal

$$y(t) = x(t - x_{\mathrm{LFO}}(t)), \tag{8}$$

where $y(t)$ is a delayed version of the input signal $x(t)$. The periodic variation of the time delay by the LFO signal $\mathrm{x_{LFO}}(t)$ produces a recurrent pitch variation in the output signal.

2.3.3. Phaser

A phaser produces an output signal which is a mix of the unprocessed, *dry* input signal $d \cdot x(t)$ and an allpass-filtered, *wet* version of the same signal, $w \cdot \mathrm{AP}^M\{(x(t), x_{\mathrm{LFO}}(t))\}$, where d and w are gain factors weighting the unprocessed or processed parts of the input signal. The allpass filter cascade introduces a frequency dependent phase shift in respect to the dry signal. When both signals are added together to yield the output signal

$$y(t) = d \cdot x(t) + w \cdot \mathrm{AP}^M\{(x(t), x_{\mathrm{LFO}}(t))\}. \tag{9}$$

Phase cancellation and elevation for certain frequencies occur. This results in notches in the spectrum of the output signal. These notches are moving along the frequency axis over time, because $x_{\mathrm{LFO}}(t)$ modulates the center frequency of the allpass filters periodically. The amount of notches is controlled by the order M of the allpass filter cascade. For first order filters there are $\frac{M}{2}$ notches because a first-order allpass filter introduces a maximum phase shift of π [10].

2.3.4. Chorus

The chorus effect originated from the idea of emulating several musicians playing in unison (the same melody). Because they are not playing in perfect sync, there are small deviations in loudness and timing. To achieve this effect one or more copies of the input signal are delayed by 10 ms to 25 ms with small and random variations in the delay times and added to the original signal. It is a mix between the original (*dry*) signal $x(t)$ and a vibrato effect,

$$y(t) = x(t) + w \cdot x(t - x_{\mathrm{LFO}}(t)). \tag{10}$$

The delayed signal is realized with a variable delay line, controlled by the LFO-signal $x_{\mathrm{LFO}}(t)$ [11, 12].

2.3.5. Flanger

The flanger produces a similar sound to that of a chorus effect but differs in some aspects. The main difference is the delay time of the modulation path. It is between 0 ms and 15 ms. Secondly, for most flangers, the modulation is done with a feedback path while the chorus has a feed-forward architecture [11]. The output signal

$$y(t) = x(t) + w \cdot x_{\mathrm{SP}}(t), \tag{11}$$

consists of a direct path $x(t)$ and a weighted second path $w \cdot x_{\mathrm{SP}}(t)$ as shown in Fig. 2. The second path is the feedback path with

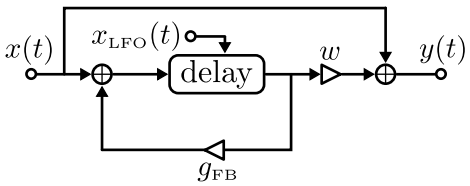


Figure 2: *Block diagram of a flanger.*

feedback gain g_{FB},

$$x_{\mathrm{SP}}(t) = x(t - x_{\mathrm{LFO}}(t)) + g_{\mathrm{FB}} \cdot x_{\mathrm{SP}}(t - x_{\mathrm{LFO}}(t)). \tag{12}$$

The oscillator signal $x_{\mathrm{LFO}}(t)$ modulates the delay of the second path with a sinusoidal signal [12].

3. INPUT SIGNALS

In this section the input signals which are used for the feature computation, are introduced. Because the input signals were digitally synthesized, the authors decided to change the notation from continuous time signals $x(t)$ to discrete time signals $x(n)$ where $x(t)$ is sampled with sampling rate f_s.

3.1. Sinusoidal Signals

3.1.1. Sine Wave with Fixed Amplitude

The first basic input signal is a sinusoidal wave with frequency f_0, duration of d seconds, and fixed amplitude A,

$$x_{\sin\,\mathrm{fix}}(n) = A \cdot \sin\left(2\pi \frac{f_0}{f_s} n\right), \tag{13}$$

with $n = [0, \ldots, d \cdot f_s]$ and sampling rate f_s.

3.1.2. Amplitude-Modulated Sine Wave

The sine signal can be modified to feature varying envelopes

$$x_{\sin\,\mathrm{var}}(n) = a(n) \cdot \sin\left(2\pi \frac{f_0}{f_s} n\right), \tag{14}$$

where $a(n)$ is a time-variant amplitude. In this work, the amplitude switched abruptly between a_{high} and a_{low} with a low frequency, e.g. $f_{\mathrm{mod}} = 2\,\mathrm{Hz}$.

3.2. Impulse

To identify a system's linear behaviour for a whole frequency range, a broadband stimulus has to be fed to the system. The discrete unit impulse

$$x_{\mathrm{pulse}}(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{else} \end{cases}, \tag{15}$$

achieves the excitation of a system for the whole frequency range uniformly and is used to measure the impulse response of the DUT.

3.3. Sweep

Another way of exciting a system within a certain frequency range is to feed an exponentially swept sine wave, as described in [13]. The sweep

$$x_{\text{sweep}}(n) = \sin\left(\omega_1 \cdot \frac{N}{\ln(\frac{\omega_2}{\omega_1})} \cdot \left[\mathrm{e}^{\frac{n}{N}\cdot\ln(\frac{\omega_2}{\omega_1})} - 1\right]\right), \qquad (16)$$

has a duration of N samples and covers frequencies from $\omega_1 = \frac{2\pi f_{\text{start}}}{f_s}$ to $\omega_2 = \frac{2\pi f_{\text{stop}}}{f_s}$.

4. FEATURES

As previously mentioned, this study is groundwork for a classification procedure that shall present the features on which a following classification can be based. In this section, the features are derived and adapted to create a comparable quantitative representation.

4.1. Distortion Feature

A typical description of distortion for audio devices is the so-called total harmonic distortion (THD)

$$\text{THD} = 20 \cdot \log_{10}\left(\frac{\sqrt{V_1^2 + V_3^2 + \ldots + V_6^2}}{V_0}\right), \qquad (17)$$

where V_i is the RMS amplitude of the i_{th} harmonic. To calculate the THD, signal $x_{\text{sin fix}}$ (see Eq. 13) is sent through the DUT and recorded. Afterwards the recorded signal is segmented to contain k times the period length $k \cdot \frac{f_s}{f_0}$ samples of the sine wave. This has the advantage that there is no leakage effect for the frequency f_0 and all of its harmonics, i.e. all the energy of f_0 falls into one Discrete Fourier Transform (DFT) bin.

The THD is calculated by extracting the RMS amplitudes of the fundamental frequency and the RMS amplitude of the harmonics. The highest harmonic considered is V_6. The related metric s_{THD} is defined as

$$s_{\text{THD}} = \frac{\text{THD} + 70}{80} \qquad (18)$$

and maps the values from the initial range of $[-70, \ldots, 10]$ to $[0, \ldots, 1]$. Outlying values are clipped to the maximum and minimum value, respectively.

4.2. Compression Feature

To detect a compressor, the signal levels of the varying amplitude sine wave $x_{\text{sin var}}$ from (Eq. 14) are analyzed. If the signal-level-difference of the output would be smaller than the signal-level-difference of the input, dynamic range compression is occurring. But reduction of the signal-level-difference is also occurring in a highly nonlinear distortion effect. For this reason the compression feature is based on the attack and release regions of a compressor and not on the signal-level-difference. Hence, it is assumed that attack and release times differ significantly.

The computation of the compression feature is depicted in Fig. 3 and described in the following. After recording the replayed test signal $x_{\text{sin var}}$ (see Eq. 14), with two different amplitude levels (see Fig. 4 (a)), the compression feature is calculated. At first, the envelope of the signal is computed by linearly interpolating between all its local maxima. The result of this processing step is shown in Fig. 4 (b).

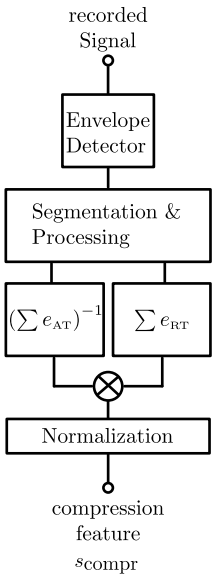


Figure 3: *Block-diagram of the compression feature calculation.*

The resulting envelope is segmented into attack and release part by cutting each part according to the time-domain positions of the input signal. The beginning of the attack-part corresponds to the region of the input signal $x_{\text{sin var}}$ where it changes from high signal level a_{low} to a_{high}. And the start point of the release-part is the region where the input signal changes from low signal level a_{low} to a_{high}. The segmented envelope parts of attack and release regions are depicted in Fig. 4 (c) and (d).

After attack and release part have been segmented, the release envelope is mirrored horizontally. Then, the minimum of both curves is subtracted from the signal. The result of this processing step yields only the overshoot of attack and release filter and is shown in Fig. 4 (e) and (f). The colored areas in said figure represent the integral over the curves.

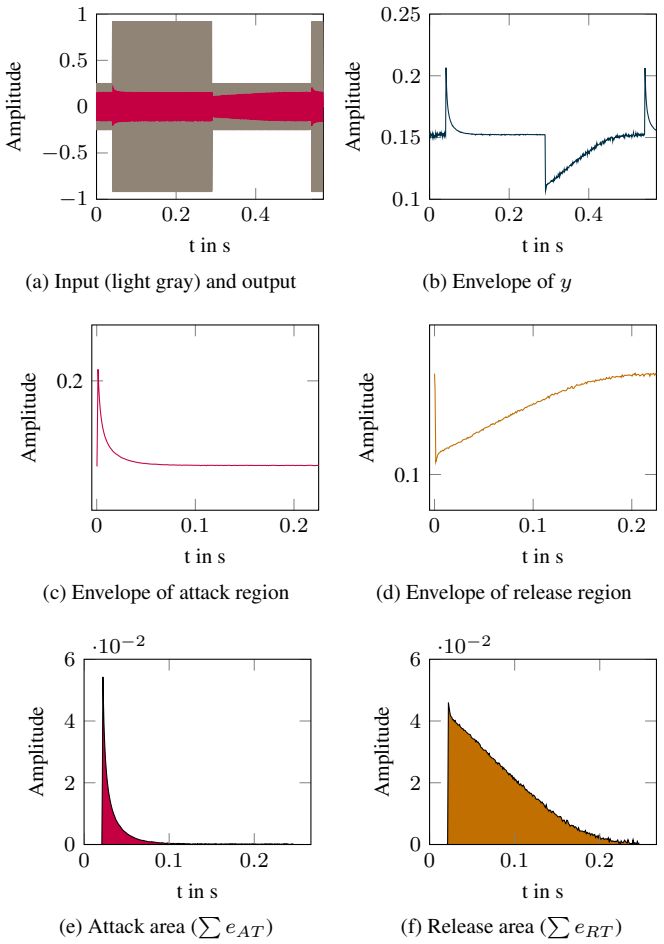
The ratio of the integrals is computed and normalized to form the compression feature

$$s_{\text{compr}} = |F_C - 1|, \qquad (19)$$

where $F_C = \frac{\sum e_{RT}(n)}{\sum e_{AT}(n)}$ is the ratio of the release and attack integrals. The initial assumption that a compressor has different settings for attack and release time does not always have to be true, but most commercial compressor units tend to have a much longer release time than attack time, whereas release and attack time of a distortion effect unit should be roughly the same. s_{compr} becomes zero if attack and release time are the same and approaches a higher value if attack and release time differ. s_{compr} is clipped to one if it should exceed this value.

4.3. Time Variance Feature

To detect the time-variance of a system, the exponential sine sweep, input signal x_{sweep}, is sent three times through the DUT and recorded. To avoid sending the signal with the same periodicity as a potential LFO of the time-variant effect could feature, the pauses between each sweep are non-uniform. The recorded signal is segmented into the three recorded sweeps $y_1(n)$, $y_2(n)$, and



(a) Input (light gray) and output (b) Envelope of y

(c) Envelope of attack region (d) Envelope of release region

(e) Attack area ($\sum e_{AT}$) (f) Release area ($\sum e_{RT}$)

Figure 4: *Compression Feature - Measured waveforms of an MXR DynaComp.*

$y_3(n)$ which are then subtracted from each other to yield three error signals $e_1(n) = y_1(n) - y_2(n)$, $e_2(n) = y_1(n) - y_3(n)$, and $e_3 = y_2(n) - y_3(n)$. The *root mean square* (RMS) value of $e_1(n)$, $e_2(n)$, and $e_3(n)$ is calculated and the maximum value e_{rms} is selected to form the time-variance metric

$$s_{\text{tvar}} = 1 - e^{-10 \cdot e_{rms}}. \tag{20}$$

Since a system is either time-invariant or time-variant the time-variance metric is distorted with a nonlinear function (see Eq. 20), to emphasize values of s_{tvar} which are non negligible.

Please note that the time-variance metric is sensitive to the level of the error signals $e_1(n)$, $e_2(n)$ and $e_3(n)$. Therefore the recorded signal $y_{\text{sweep}}(n)$ is normalized before segmentation to have a maximum amplitude of one before calculating the time-domain error.

4.4. Impulse Response Length Feature

As previously mentioned, the class of linear time-invariant effects is quite versatile. To distinguish simple filters from effects which massively influence the temporal structure of a source signal, like

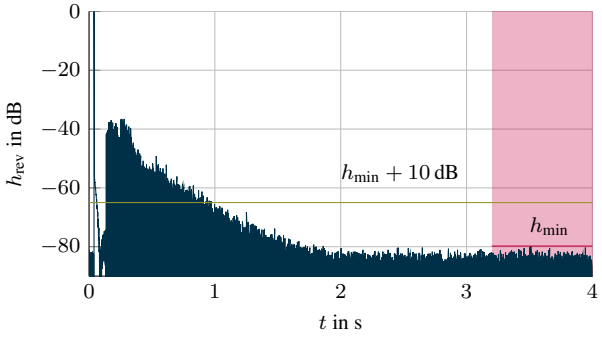


Figure 5: *Exemplary impulse response of reverb effect.*

reverb and delay effects, the length of the impulse response is a valuable information. The impulse response length feature is determined by the duration of the impulse response to decrease to a certain value. For this purpose, the DUT is excited with x_{pulse} to record the impulse response h_{rev}, which is normalized to an amplitude of 1 in the following. The maximum value h_{min} of the last section of the normalized impulse response, shown as red area in Fig. 5, is then weighted to define the threshold $h_{\text{thr}} = h_{\text{min}} + 10\,\text{dB}$. Afterwards the impulse response is scanned in time-inverse direction to find the point of time s_{len} where the first sample exceeds h_{thr}. In Fig. 5, this corresponds to $s_{\text{len}} = 0.94$ s. The feature yields the time it takes the impulse response to decrease below the threshold in seconds. So the metric only takes values greater than zero, $s_{\text{len}} \geq 0$, and is clipped to one if it exceeds one second.

5. MEASUREMENTS

5.1. Setup

The measurement setup is depicted in Fig. 6. A computer is connected to a high quality external *USB audio interface*. The output of the audio interface is connected to the input of the DUT and the output of the DUT is connected to the input of the audio interface. True amplitude scaling is guaranteed by sending a sine

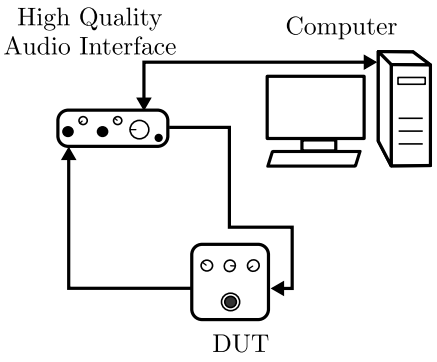


Figure 6: *Measurement Setup.*

with 0 dBFS (full scale), measuring the maximum amplitude of the recorded signal and computing the gain factor. Every recorded signal is weighted with this gain factor.

Time synchronization is achieved by computing the cross correlation $r_{xy}(m)$ between recorded signal $y(n)$ and input signal to the DUT $x(n)$. The shift of the main peak of $r_{xy}(m)$ from the center position indicates the amount of samples the recorded signal is delayed in respect to the input signal.

The different input signals, described in section 3, are merged into one long signal before they are sent through the DUT. After the signal has been recorded, it is cut into its original parts and each part is given to the feature extraction.

For extreme settings of the effect units, e.g. very long decay times for reverb or delay, problems arose while recording the output of the DUT. To assure that the effects do not alter the signal beyond recognition, all potentiometers of the effect units were set to a neutral (12 o'clock) position during the measurements. This also assured that the effect would not be too light to alter the signal enough.

5.2. Results

To get an impression of the applicability of the proposed features, they are computed for several real effect pedals. All analyzed pedals and the resulting features are listed in Tab. 1. First, the Harley Benton 7-band EQ was utilized. The distortion feature of $s_{\text{THD}} = 0.15$ and the time variance feature $s_{\text{tvar}} = 0.001$ clearly show the linear, time-invariant behavior of this system. The Tube-Screamer, ZenDrive, and the FuzzFace DIY clones from Musikding also show negligible time variance values. In contrast, the distortion feature values show a rising trend from 0.54 to 0.78. The strong non-linearity of those pedals is reflected in those values.

The MXR Dyna Comp compressor also shows a high distortion value of $s_{\text{THD}} = 0.64$. Nevertheless, it can be clearly distinguished from the distortion pedals due to the s_{compr} value of 0.87, which doesn't exceed 0.06 for the distortion pedals. The s_{tvar} measurements for the already mentioned pedals are as expected consistently very low. When the measurements are repeated with modulation pedals (MXR Flanger, MXR Phase90, and the BOSS CEB-3 Bass Chorus) a significant difference is detected. The smallest value is $s_{\text{tvar}} = 0.44$ for the Flanger. Unexpectedly, the modulation pedals also show much distortion. The verification of this finding by spectral analysis and actual listening affirmed the nonlinear characteristics of these devices. Nevertheless, the categorization into the group of time-variant effects is still possible with proposed time variance feature.

The s_{compr} value yields no noteworthy results for all time-variant effect units. Due to time variant phase-shifts and signal superpositions, the envelope of the recorded signal behaves in an, in this context, unpredictable way. This leads to s_{compr} values varying between 0.07 to 1, for repeated measurements. Even though the delay is technically not a time-variant effect, the results for the compression metric do not yield an accurate or even reproducible result. The computation of the compression metric is based on the envelope of the recorded signal, which is corrupted by the copies of the input signal which are added to the output signal. The remaining effect pedals of the analysis are the TC Electronics *Hall of Fame* reverb and the *Flashback X4* delay. Apparently, these devices tend to perform quite linear ($s_{\text{THD}} = 0.14, 0.06$) but differ strongly from the EQ pedal in terms of their impulse response length s_{len} which are 0.94 and over 1 seconds respectively. Considering the insignificant s_{len} values for other pedals the authors assume that the identification of reverb and delay pedals using the proposed feature is doubtlessly realizable.

Table 2: *Template vectors for different effect categories.*

Filter	$\hat{v}_F = [0,\ 0,\ 0,\ 0]$
Distortion	$\hat{v}_D = [1,\ 0,\ 0,\ 0]$
Compression	$\hat{v}_C = [0,\ 0,\ 1,\ 0]$
Linear Time-variant	$\hat{v}_{TV} = [0,\ 1,\ 0,\ 0]$
Reverb o. Delay	$\hat{v}_{RV} = [0,\ 0,\ 0,\ 1]$

6. EXEMPLARY CLASSIFICATION

To demonstrate the usability of the proposed features a very simple classifier was implemented. It is based on pattern matching as suggested by [14] for the purpose of chord identification based on chroma vectors. Therefore, the features from section 4 for a specific device are combined in a feature vector

$$v = [s_{\text{THD}},\ s_{\text{tvar}},\ s_{\text{compr}},\ s_{\text{len}}]. \tag{21}$$

In the next step, the squared euclidean distance Δv_i

$$\Delta v_i = \sum (v - \hat{v}_i)^2 \tag{22}$$

of the current feature vector v to all template vectors $\hat{v}_i$, listed in Tab. 2 is computed. The output of the classifier is the effect corresponding to the template vector resulting in the smallest Δv_i. The results of these calculations can be seen in Fig. 7. The euclidean distance from the current feature vector to each Δv_i (Tab. 2) is shown for each effect unit under test. The lowest distance yields the result of the classification. Already this very simple approach delivers a flawless classification for the values of Tab. 1.

7. CONCLUSION

This goal of this study was to derive features that allow the classification of audio effect units via I/O measurements. Based on simple signal models of typical effects and specifically designed input sequences, the features are derived. The features describe the characteristic of a DUT in terms of distortion, compression, time variance, and impulse response length.

The experimental application of the features on 10 different guitar effect units (*stompboxes*) proved the usability of the features to differentiate between effect classes. A very simple classifier based on pattern matching was already capable of flawlessly distinguishing between the tested devices. Advanced classifiers based on neural networks or support vector machines are expected to perform even better and be able to handle more effect categories using potentially more features. This process flow allows the autonomous modeling of various audio processing equipment.

Further improvements could be implemented to differentiate even more effects. Currently it is for example not possible to distinguish a reverb effect unit from a delay effect unit or a chorus from a flanger. Further features can be designed for enhanced classification offering detailed information about the device under test.

8. REFERENCES

[1] U. Zölzer, *DAFX: Digital Audio Effects*, vol. 2, Wiley Online Library, 2011.

[2] M. Holters and U. Zölzer, "Physical modelling of a wah-wah effect pedal as a case study for application of the nodal

Table 1: *Overview of results for various effect pedals.*

Effect Category	**Effect Unit**	$s_{\mathbf{THD}}$	$s_{\mathbf{tvar}}$	$s_{\mathbf{compr}}$	$s_{\mathbf{len}}$
Filter	Harley Benton EQ	0.15 (-57.72 dB)	0.005	0.04	0.11
Distortion	Musikding TubeScreamer	0.54 (-27.2 dB)	0.005	0.03	0.002
Distortion	Musikding Zendrive	0.64 (-18.89 dB)	0	0.06	0.06
Distortion	Musikding FuzzFace	0.78 (-7.57 dB)	0.0198	0.03	0.002
Compression	MXR Dyna Comp	0.64 (-18.9 dB)	0.1813	0.87	0
Flanger	MXR Flanger	0.45 (-33.94 dB)	0.8892	(0.1 - 0.8) time variant	0.03
Phaser	MXR Phase90	0.93 (4.61 dB)	0.970	(0.07 - 0.8) time variant	0.004
Chorus	BOSS CEB-3 Bass Chorus	0.4 (-37.65 dB)	0.993	(0.3 - 1) time variant	0.006
Reverb	TC Hall of Fame	0.14 (-58.6 dB)	0.30	0.138	0.94
Delay	TC Flashback X4 (2290)	0.06 (-65.07 dB)	0.020	(0 - 1) superposition	1

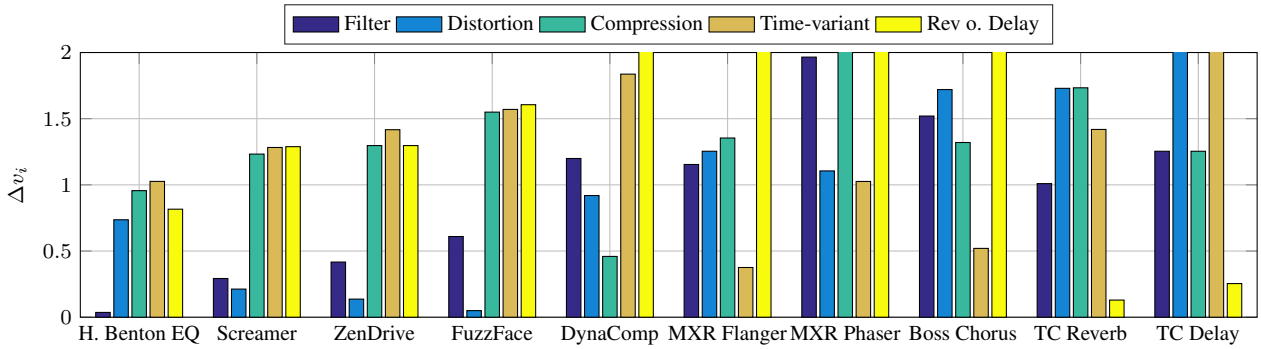


Figure 7: *Squared euclidean distance of feature vector and template vector for all measured devices.*

dk method to circuits with variable parts," in *Proceedings of Digital Audio Effects (DAFx-11)*, Paris, France, Sept. 19-23, 2011, pp. 31–35.

[3] D. Yeh and J.O. Smith, "Simulating guitar distortion circuits using wave digital and nonlinear state-space formulations," in *Proceedings of Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sept. 1-4, 2008, pp. 19–26.

[4] J. Macak, *Real-time Digital Simulation of Guitar Amplifiers as Audio Effects*, Ph.D. thesis, Brno University of Technology, 2011.

[5] M. Karjalainen and J. Pakarinen, "Wave digital simulation of a vacuum-tube amplifier.," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, May 2006, pp. 153–156.

[6] L. Ljung, *System identification*, Springer, 1998.

[7] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[8] A. Novak, L. Simon, P. Lotton, and J. Gilbert, "Chebyshev model and synchronized swept sine method in nonlinear audio effect modeling," in *Proc. Digital Audio Effects (DAFx-10)*, Graz, Austria, Sept. 6-10, 2010.

[9] C. McKay, I. Fujinaga, and P. Depalle, "jaudio: A feature extraction library," in *Proceedings of the International Conference on Music Information Retrieval*, London, UK, September 2005, pp. 600–3.

[10] F. Eichas, M. Holters, M. Fink, and U. Zölzer, "Physical modeling of the mxr phase 90 guitar effect pedal," in *Proceedings of Digital Audio Effects (DAFx-14)*, Erlangen, Germany, Sept. 1-5, 2014, pp. 153–158.

[11] P. Dutilleux, "Filters, delays, modulations and demodulations: A tutorial," in *Proceedings of DFX98, Digital Audio Effects Workshop*, Barcelona, Spain, 1998, pp. 4–11.

[12] S. Orfanidis, *Introduction to Signal Procesing*, vol. 1, Prentice-Hall, Inc., 1996.

[13] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Paris, France, February 2000.

[14] T. Fujishima, "Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music," *Proc. of International Computer Music Conference (ICMC), 1999*, pp. 464–467, 1999.