# Time-Series and Frequency-Series Features for Musical Instrument Classification

Landon Buell

3 June 2020

## 1 Introduction

## 2 Time Series Features

Time series features (also called time-domain or time-space features) are pieces of information that can be extracted, derived, or otherwise originate from a signal that in expressed with a dependence on time. This can include weather forecasts, year-to-year population of a given demographic, audio information, or economic stock data. In the context of audio data, this is generally referred to as a *waveform*

### 2.1 Rise and Decay Time

Signal *rise time* measures the amount of time it takes for a time-domain signal, $S(t)$, to leave it's steady-state condition and *rise* to a given maximum $\max\left[|S(t)|\right]$. Signal *decay time* measures the amount of time it takes for a time-domain signal, $S(t)$, to *decay* from a given maximum value and return to it's steady-state value. Due to the nature of digitally stored audio in both time and amplitude, it may be impossible to find entries in the signal that satisfy exact equilibrium values (usually, 0).

To account for this, we employ a function within our feature-extraction program that measures the 10% to 90% amplitude rise time, and the subsequent 90% to 10% amplitude decay time. This means that we build a function within our feature-extraction program that isolates the produces two array-like objects for each signal. The first array will hold the position (or index) of all points that have a value that exceeds 90% of the waveform maximum (counting for negative amplitudes as well). The second array contains similarly index of values that exceed 10% of the maximum, but not 90%.

If we take the difference between the 0-th entry of each array, this is the number of samples that it took for the signal to reach (roughly) 10% of it's maximum amplitude to (roughly) 90% - this *rise time*, $\Delta t_R$. Similarly, if we take the difference between the final entry in each array, this gives the *decay time*, $\Delta t_D$. Both measurements are in units of the *number of samples*; to convert into more traditional units of time (e.g. seconds), one can multiply by the spacing between samples or divide by the sample rate.

**Algorithm 1** Compute amplitude rise and decay time of a discrete time-domain signal $S(t)$ for aa give lower bound percentage and upper bound percentage.

**Require:** Discrete time series signal: $S$
**Require:** Value to consider lower bound percentage (0,1): *low_bnd*
**Require:** Value to consider upper bound percentage (0,1): *high_bnd*
**Ensure:** low_bnd < high_bnd
  n_samples ← length($S$)
  $S \leftarrow |S|$
  max_amp ← maximum($S$)

### 2.2 Root-Mean-Square Energy of Frames

The Root-Mean-Square energy (RMS energy) of a frame is

### 2.3 Percentage of Frames with RMS Energy above value

### 2.4 Time Spectral Flux

## 3 Frequency Series Features

### 3.1 Number of Unique, Isolated Peak

### 3.2 Energy of Signal in Select Frequency Bands

## 4 Combined Series Features

### 4.1 Spectrogram Matrix

### 4.2 Phase-Space Matrix

## 5 Additional Features

## Summary

## References

[1] Khan, M. Kashif Saeed, and Wasfi G. Al-Khatib. "Machine-Learning Based Classification of Speech and Music." Multimedia Systems, vol. 12, no. 1, 2006, pp. 55–67.,

doi:10.1007/s00530-006-0034-0.

[2] Short, Garcia, "Signal Analysis Using the Complex Spectral Phase Evolution (CSPE) Method". Journal of the Audio Engineering Society.