Feature Extraction Toolbox for Transients

Gaetano Andreisek[1], Bernhard U. Seeber[1]

[1] *Audio Information Processing, 80333 Munich, E-Mail: gaetano.andreisek@tum.de, seeber@tum.de*

Introduction

In acoustics, transient signals can be generated by impacting, striking or tapping objects. When recorded, the sound features an abrupt increase of amplitude and a subsequent damping process (see figure 1). This damping process is unique to the bearing and material of the sounding object: e.g. when struck with the same force and set of sticks, a drum sounds different from a cymbal. Transient signals therefore convey information about the nature of the sounding object (material, shape, mode of excitation) [1]. This information can be abstracted with acoustic features that describe the size, shape or proportion of different representations of the sound signal (e.g. time domain, Fourier domain). Many software toolboxes enable automated and efficient extraction of acoustic features from sound recordings [2, 3, 4, 5], and are often are optimized for a specific type of signal (e.g. music, speech, etc.). The Feature Extraction Toolbox for Transients (FETT) presented here has be optimized to extract acoustic features from transient signals.

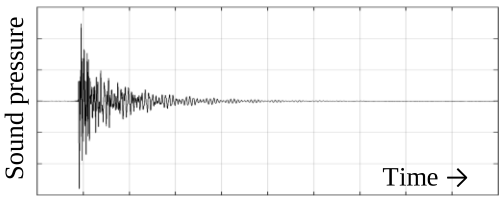

Figure 1: Typical transient signal for FETT. After a sharp increase in sound pressure, a damped decay follows.

Applications

FETT can be used to analyse any transient signal with damped decaying energy. Applications include recordings from percussion instruments, acoustic-based non-destructive material testing, room impulse responses, sound quality applications, and other impact sounds.

Software architecture of FETT

FETT is a Matlab-toolbox that operates on transient signals with damped energy decays (main input). Signals from the excitatory source, such as the time-force history of an impulse hammer, will be accepted and a set of features can be extracted as well. As known from other toolboxes, the core computation of FETT is divided into two parts: (i) transformation of the main input into 'input representations' and (ii) extraction of acoustic features (e.g. implemented in the Timbre Toolbox, [4]). A central configuration table stores all relevant parameters for the computations such as length of various windows, window hop sizes, frequency ranges, etc. In order to avoid unnecessary computations, individual input representations and corresponding features can be (de-)selected. A central feature of FETT is the estimation of the signal onset and offset, which is performed before transforming the main input into the corresponding input representations. This step is crucial since some features rely on a robust estimation of the signal boundaries. Furthermore, FETT is designed to easily integrate self-defined input representations and features.

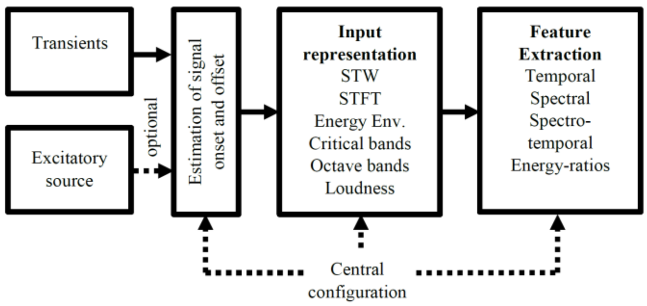

Figure 2: Stepwise process from input to features as done by FETT. Input from excitatory source (e.g. impulse hammer) is optional. A central configuration table store all necessary parameters for estimation of signal boundaries (onset and end), input representation and feature extraction.

Input Representation

Before features can be extracted, the sound recordings have to be transformed to a suitable format ('input representations'). These transformations include short-time windowing, temporal energy envelope, short-time Fourier transformation, Critical Band filtering, octave band filtering and loudness according to DIN 45631/A1 [6].

Feature Extraction

Following transformation, temporal, spectral and spectro-temporal features can be extracted. Temporal features are extracted from the temporal energy envelope and short–time windowing of the input, and include attack time, attack slope, time above threshold, zero-crossings over time or slope of decay (linear or higher order regressions). The list of spectral features contain among others spectral energy ratios, spectral flux, or spectral roll-off. Spectro-temporal features mainly comprise energy decays in frequency bands. Some features can be extracted from more than one input representation. These include energy ratios, statistical moments (centroid, spread, skew, and kurtosis) and decay estimations. In total, more than 350 features can be extracted.

Acknowledgement

This work is supported by the GreenTech Wind initiative of the EuroTech universities and the International Graduate School of Science and Engineering (IGSSE) at the Technical University of Munich.

Literature

[1] Lutfi, R. A.: Sound source identification. In: Springer Handbook of Auditory Research: Auditory Perception of Sound Sources, edited by Yost, W. A. und Popper, A. N., Springer US, New York (2008), 13-42

[2] Søndergaard,P. und Majdak, P.: The Auditory Modeling Toolbox. In: The Technology of Binaural Listening, edited by Blauert, J., Springer, Berlin, Heidelberg, (2013), 33-56

[3] Boersma, P. und Weenink, D.: Praat: doing phonetics by computer. Version 6.0.26, URL: `http://www.praat.org/` (2017)

[4] Peeters, G., Giordno, B. L., Susini, P., Misdariis, N. und McAdams, S.: The Timbre Toolbox: Extracting audio descriptors from musical signals. J. Acoust. Soc. Am. 130(5) (2010), 2902-2916

[5] Lartillot, O. und Toiviainen, P: A Matlab Toolbox for Musical Feature Extraction from Audio. International Conference in Digital Audio Effects, Bordeaux (2007)

[6] DIN 45631/A1:2010-03, Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sounds. Beuth-Verlag, Berlin, 2010