A Classification Database for Short-Duration Acoustic Signals (Version 1.3)

Dr. Paul M. Baggenstoss
Naval Undersea Warfare Center
Newport RI, 02841
401-832-8240 (TEL)
p.m.baggenstoss@ieee.org
http://class-specific.com/csf/index.html

December 29, 2014

Abstract

We present a new database designed to develop classification algorithms for short-duration acoustic time-series. We describe the database and present experimental results for a variety of algorithms.

1 Introduction

In undersea or in-air acoustic classification problems, one often encounters a wide-variety of short-duration signals. Examples include marine mammal vocalizations, man-made sonar pings machinery noises. Widely-used public classification databases are typically unsuited to develop classification algorithms for such problems. Some databases are feature-based, eliminating the possibility for the designer to develop features. Others are more suited for image classification, or are focussed on obtaining good performance on huge amounts of training samples. In acoustic classifcation, there is often difficulty in acquiring large numbers of samples for training and there is a high-cost associated with making errors (need to use resources to examine a false-alarm). Thus, the goal should be near 100 percent classification performance with few training samples. The characteristics of a good database include:

1. Time-series only (no features).
2. A good variety of short-duration acoustic signal types containing diverse spectral and temporal characteristics.
3. Ability for humans to make the correct calls - thus a goal of near 100 percent performance. Thus, clean signals and consistency in signal duration and noise background are required.
4. Moderate number of training and testing samples (focus on classification error with fewer training samples).

We present a database with these characteristics.

2 Database description

2.1 Overview

Data was collected from 24 classes in a typical office environment using 16-bit PCM at 32Khz sampling rate. All samples were the same length, 16128 samples (roughly 0.5 sec). For each class there were 102 samples. We placed an emphasis on a both consistency and similarity. In other words, an attempt was made to be consistent in how a sound was made within a class. To make classes "similar", often the same object or objects were used to produce different distinct sounds. In the end, colored synthetic noise was added to mask any differences in the background noise during the recording of each class.

2.2 Data Holdout Sets

For the purpose of cross-validation, and to measure the degradation as fewer training data samples are used, we use two-fold, three-fold, and four-fold holdout sets. In two-fold, we divide the data into two sets, the first 51 samples of each class, and the last 51 samples of each class. We train using one set, and test using the other, then switch the sets. This way. we report results on all 102 samples of each class. For three-fold holdout sets, we train using one one third of the samples (34 samples) comprised of the first, second, or third contiguous 34 samples. We test using the remaing 2/3 of the data in

each case. As a result of this, we report resuts on 4080 samples since each sample is tested twice.

2.3 Classes

2.3.1 Class 1: Penny

A penny was dropped onto a sheet of paper laying on a table. The sound had sharp impulses and a faint resonance of the Penny. The spectrograms of three samples are shown in figure 1.

2.3.2 Class 2: Quarter

A quarter was dropped onto a sheet of paper laying on a table. Like "Penny", the sound had sharp impulses resonances at different frequencies from the quarter. The spectrograms of three samples are shown in figure 2.

2.3.3 Class 3: Book

Four different books were dropped on the table producing a "thud". The spectrograms of three samples are shown in figure 3.

2.3.4 Class 4: Coins

Three pennies, three quarters, and two dimes were briefly tossed in the air and caught to produce a "jingle" sound near the microphone. The spectrograms of three samples are shown in figure 4.

2.3.5 Class 5: Golf ball

A golf ball was bounced on the table. The spectrograms of three samples are shown in figure 5.

2.3.6 Class 6: Keys

A bundle of keys was dropped on the table. The spectrograms of three samples are shown in figure 6.

2.3.7 Class 7: Pretzels

A bag of pretzels was grabbed by the hand, producing a "rustle" sound. The spectrograms of three samples are shown in figure 7.

2.3.8 Class 8: Stapler

A sheet of paper was stapled once using a standard stapler. The spectrograms of three samples are shown in figure 8.

2.3.9 Class 9: Bottle

A bottle (one of three specimens) was placed on the table, producing a "knock" sound. The spectrograms of three samples are shown in figure 9.

2.3.10 Class 10: Door

A door was latched closed. The spectrograms of three samples are shown in figure 10.

2.3.11 Class 11: Jingle

A set of keys (same set as "Keys") was jingled by dangling the keys from the hand. The spectrograms of three samples are shown in figure 11.

2.3.12 Class 12: Paper

A sheet of paper was ripped. The spectrograms of three samples are shown in figure 12.

2.3.13 Class 13: Hangup

A telephone receiver was hung up. The spectrograms of three samples are shown in figure 13.

2.3.14 Class 14: Scissors

A sheet of paper was cut using scissors. The spectrograms of three samples are shown in figure 14.

2.3.15 Class 15: Sticks

A handful (about 6) wooden coffee stir stics were dropped into a ceramic cup. The spectrograms of three samples are shown in figure 15.

2.3.16 Class 16: Cup

A handful of "Skittles" (small hard sugar-coated candies) was dropped into a cermic cup (same cup as for "Sticks"). The spectrograms of three samples are shown in figure 16.

2.3.17 Class 17: Skittles

A handful of "Skittles" was dropped into the table. The spectrograms of three samples are shown in figure 17.

2.3.18 Class 18: Two Skittles

Just two "Skittles" were dropped into a cermic cup (same cup as for "Sticks" and "Cup"). The spectrograms of three samples are shown in figure 18.

2.3.19 Class 19: Spoon

A spoon and a fork were dropped onto the table. The spectrograms of three samples are shown in figure 19.

2.3.20 Class 20: Pens

Three pens (one mechanical pencil, one ball-point pen, and one "Sharpie") were dropped onto a table. The spectrograms of three samples are shown in figure 20.

2.3.21 Class 21: Dry

Four "Expo" brand dry-erase markers were dropped on the table. The spectrograms of three samples are shown in figure 21.

2.3.22 Class 22: Caps

A handful of plastic ball-point pen caps were dropped on the table. The spectrograms of three samples are shown in figure 22.

2.3.23 Class 23: CD

A standard CD was dropped on the table. The spectrograms of three samples are shown in figure 23.

2.3.24 Class 24: Clips

A handful of wooden clothesline clips were dropped on the table. The spectrograms of three samples are shown in figure 24.

3 Classifiers

3.1 Support Vector Machine (SVM)

For a performance baseline, we used a support-vector machine (SVM).

3.1.1 Features

We tried two different feature sets with the SVM. The first feature set "spectrogram" was extracted using the following procedure:

1. To time-align the data, circularly shift the data so that the "event" starts always at the same time (sample 1000). This was done by detecting the onset of the sound by computing instantaneous power (IP) and determining when the IP exceeded the larger of two thresholds. The first threshold was the 1/4 of the peak IP. The second was the midpoint between the peak IP and minimum IP.

Classifier	Features	% Error (2:1)	% Error (3:1)
SVM-Light	Specgram	5.59	7.21
SVM-Light	FFT	2.84	4.6
libSVM	FFT	2.84	3.43

Table 1: Error in percent for SVM classifiers using 2:1 (two-fold) and 3:1 (three fold) data houldout.

2. Compute spectrogram of the data (256-point FFT, 50% overlap, hanning weighting).
3. Compute the log of the magnitude of the spectrogram.
4. Concatenate the spectrogram into one large column vector.
5. Gather all the spectrogram vectors from the training data and compute the SVD, retain the top 128 singular vectors.
6. Project the data onto these to obtain a 128-dimensional feature.

The second feature set "fft" was extracted using the same prodedure as above, but replacing the spectrogram with the FFT of the entire 16128-sample time-series. Note that in this case, the circular alignment will not affect the feature, so it is optional.

3.1.2 Classifier

We used two publicly available SVM classifiers: (1) SVM-light toolkit by Thorsten Joachims of Cornell, and (2) libSVM from Taiwan University. We used a linear kernel for both classifiers.

3.1.3 Performance

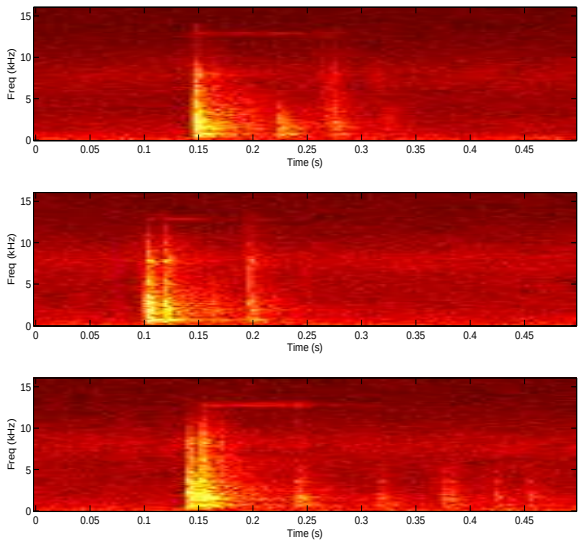

Figure 1: Three samples of class 1 “Penny”.

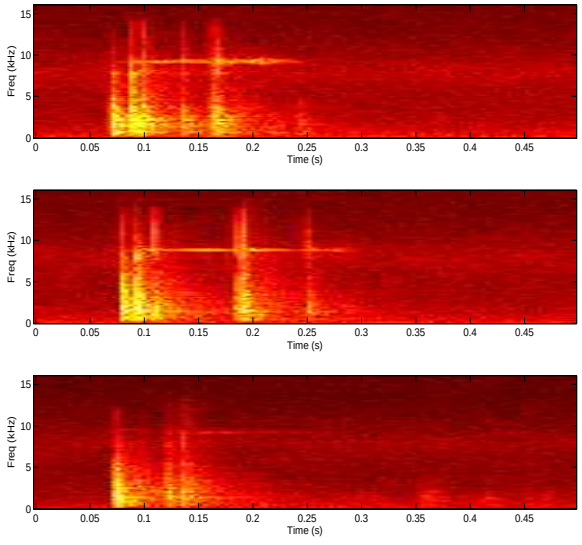

Figure 2: Three samples of class 2 “Quarter”.

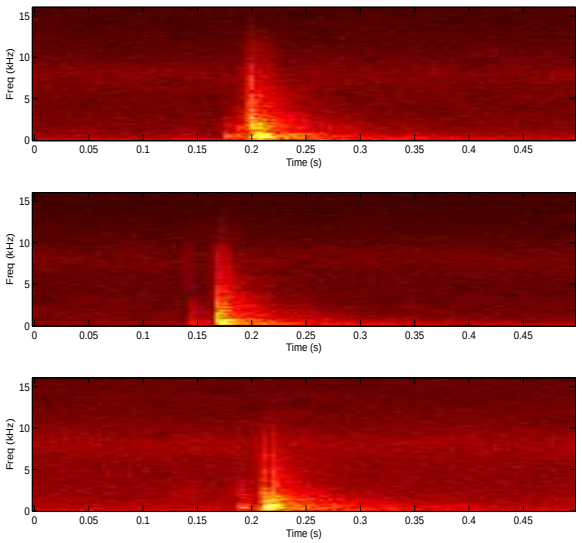

Figure 3: Three samples of class 3 “Book”.

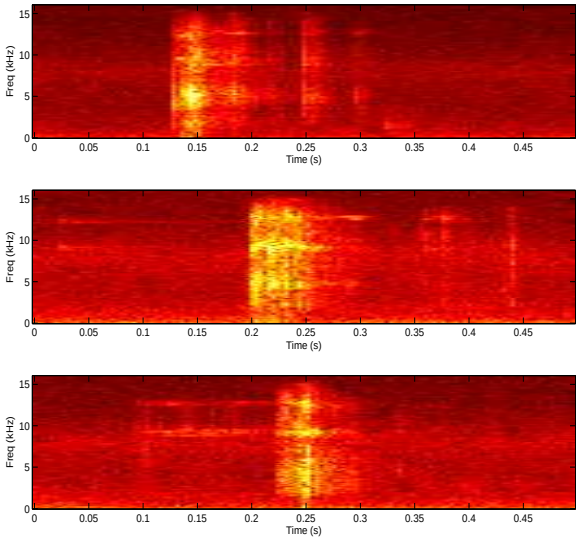

Figure 4: Three samples of class 4 “Coins”.

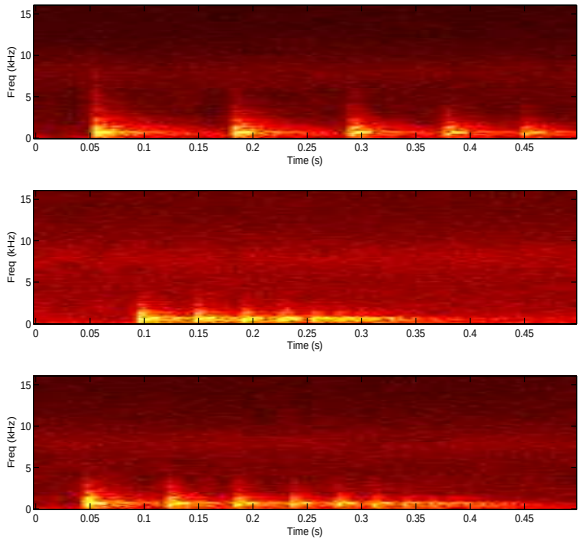

Figure 5: Three samples of class 5 “Golf ball”.

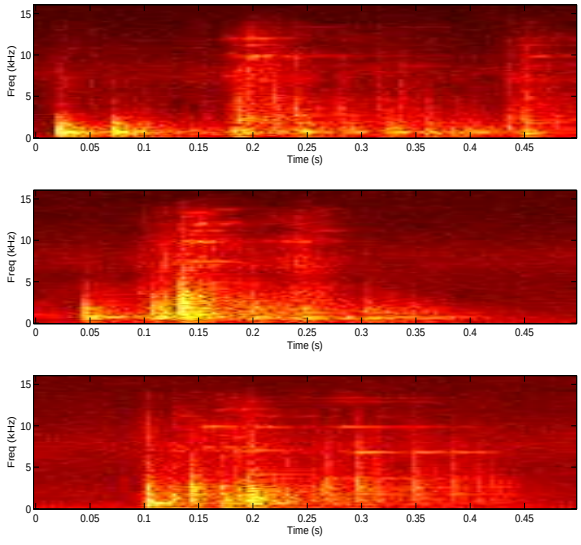

Figure 6: Three samples of class 6 “Keys”.

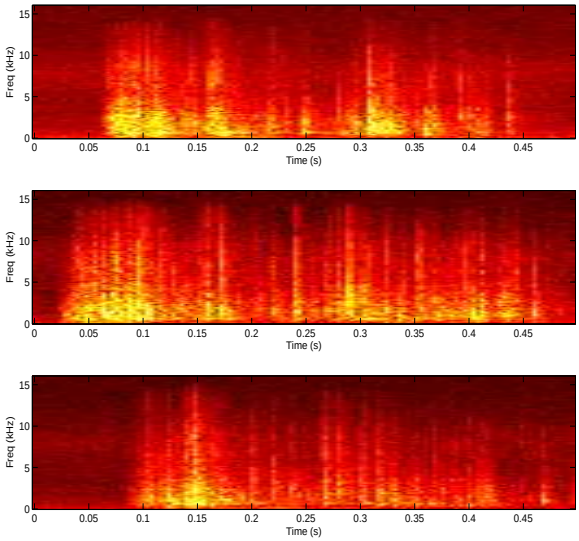

Figure 7: Three samples of class 7 “Pretzels”.

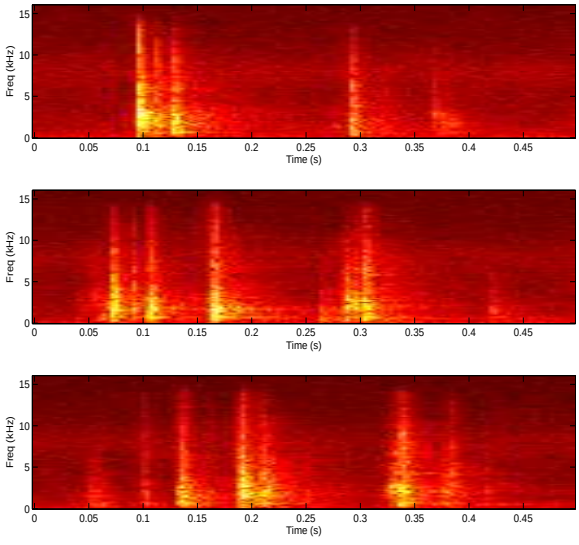

Figure 8: Three samples of class 8 “Stapler”.

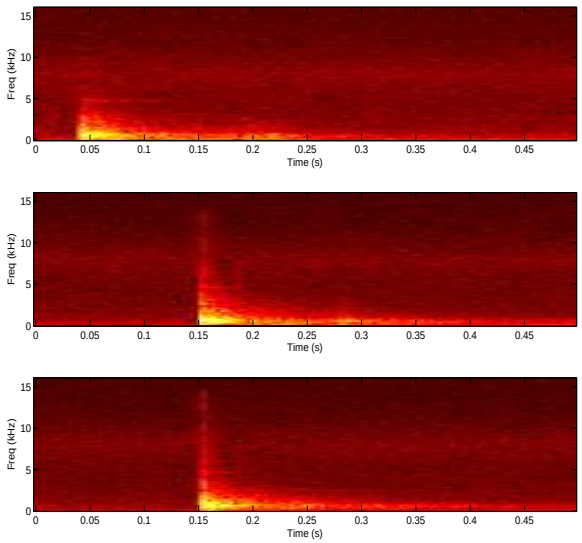

Figure 9: Three samples of class 9 “Bottle”.

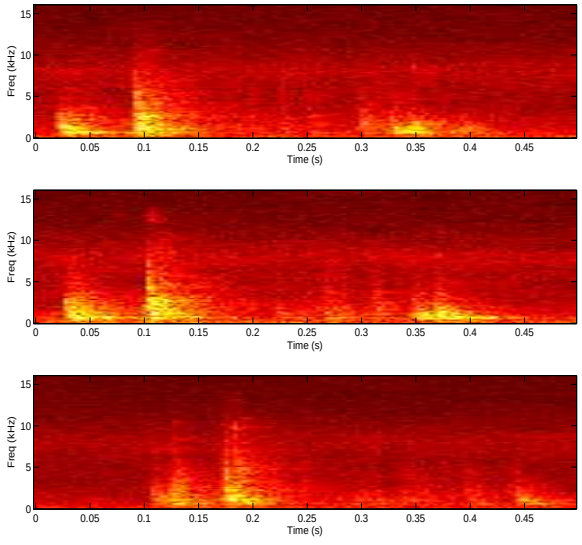

Figure 10: Three samples of class 10 “Door”.

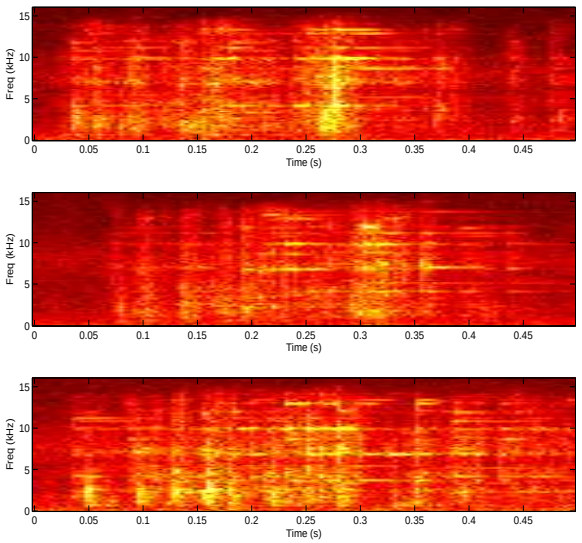

Figure 11: Three samples of class 11 “Jingle”.

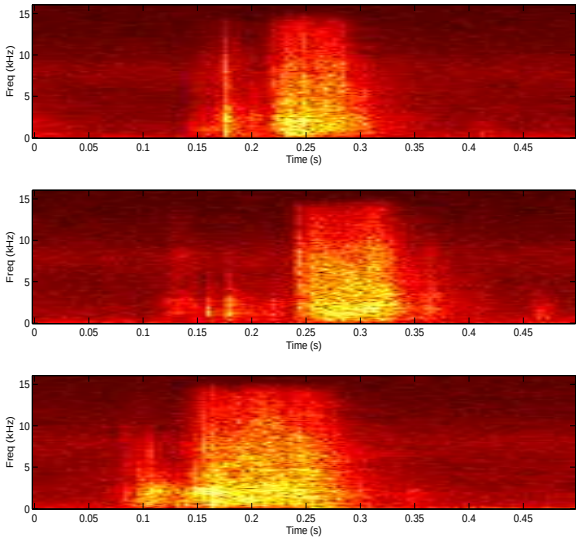

Figure 12: Three samples of class 12 “Paper”.

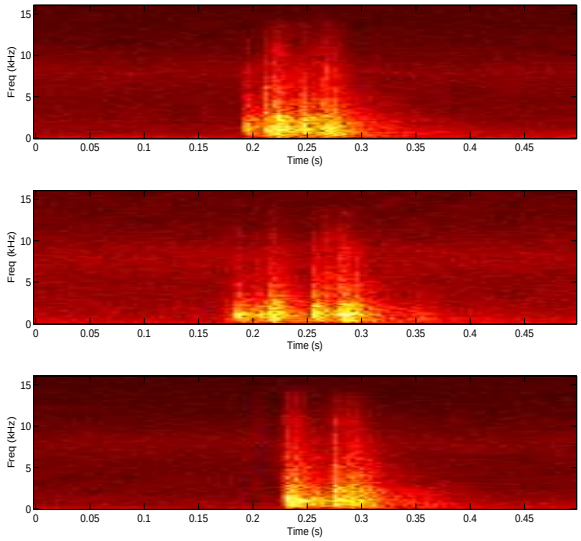

Figure 13: Three samples of class 13 “Hangup”.

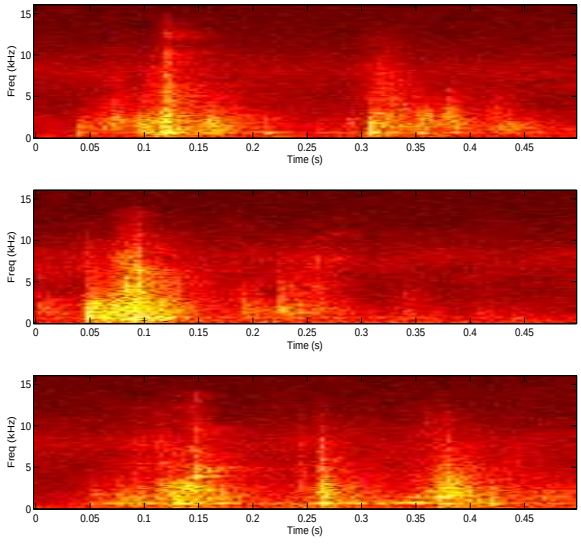

Figure 14: Three samples of class 14 “Scissors”.

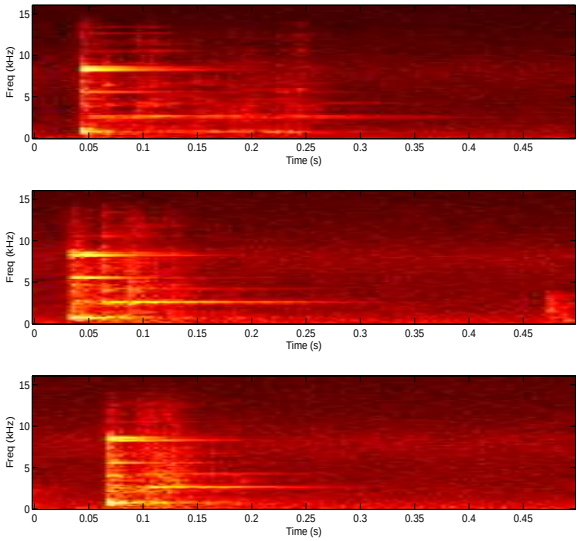

Figure 15: Three samples of class 15 “Sticks”.

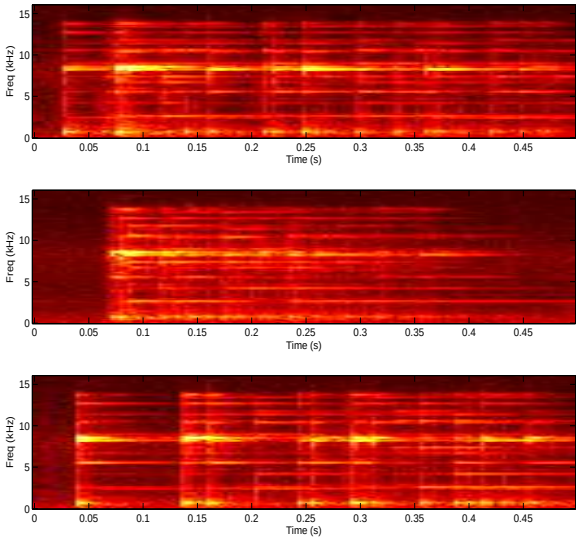

Figure 16: Three samples of class 16 “Cup”.

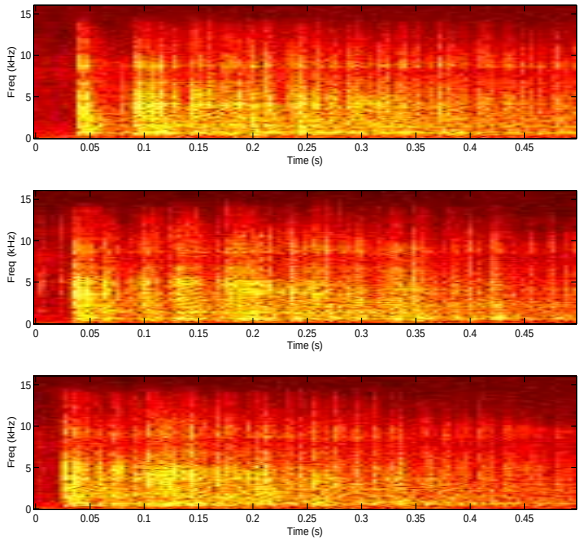

Figure 17: Three samples of class 17 “Skittles”.

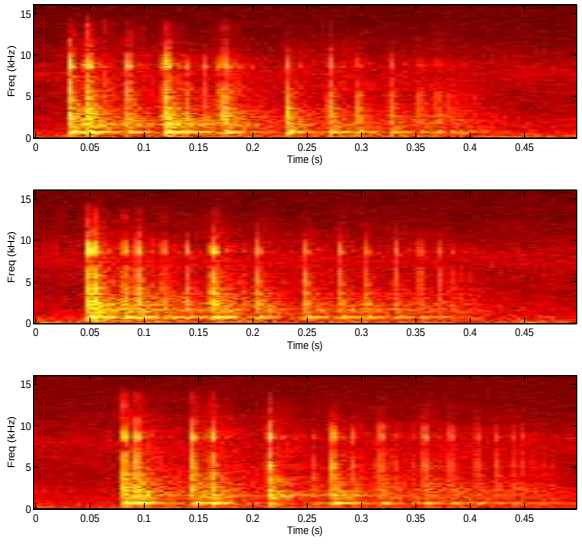

Figure 18: Three samples of class 18 “Two Skittles”.

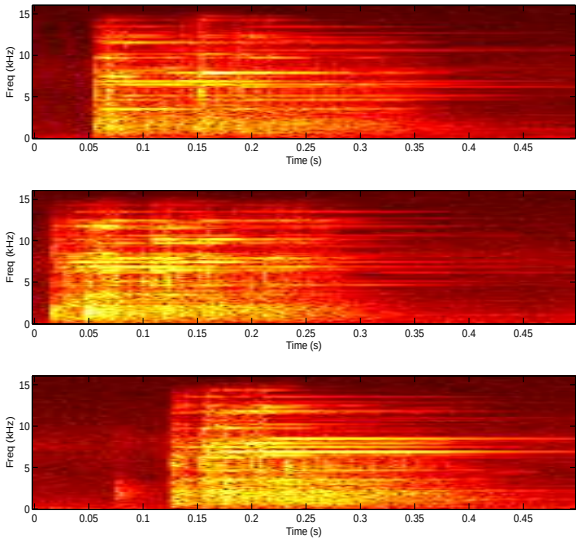

Figure 19: Three samples of class 19 “Spoon”.

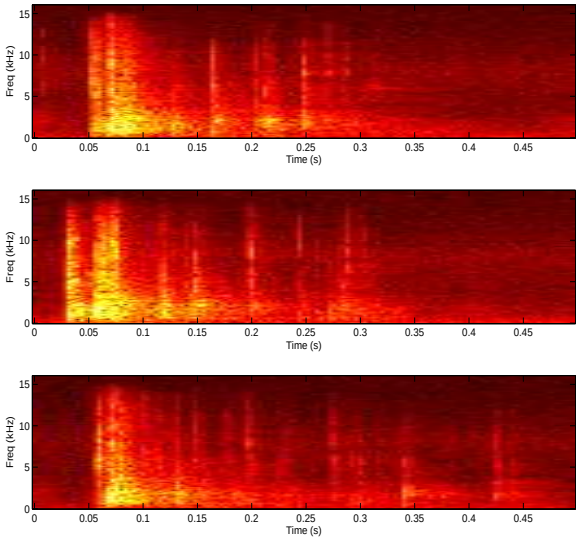

Figure 20: Three samples of class 20 “Pens”.

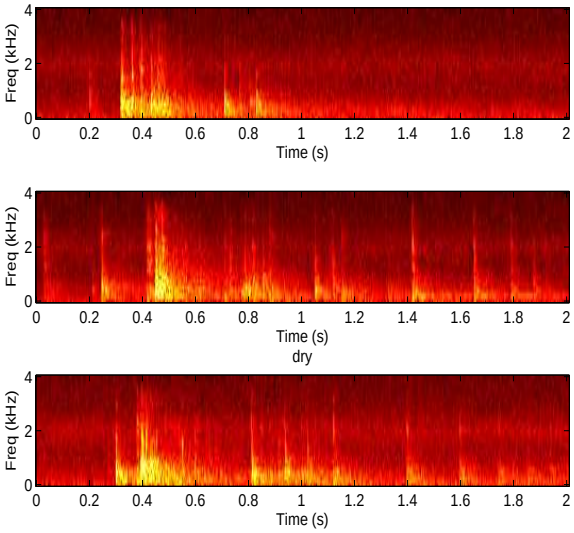

Figure 21: Three samples of class 21 “Dry”.

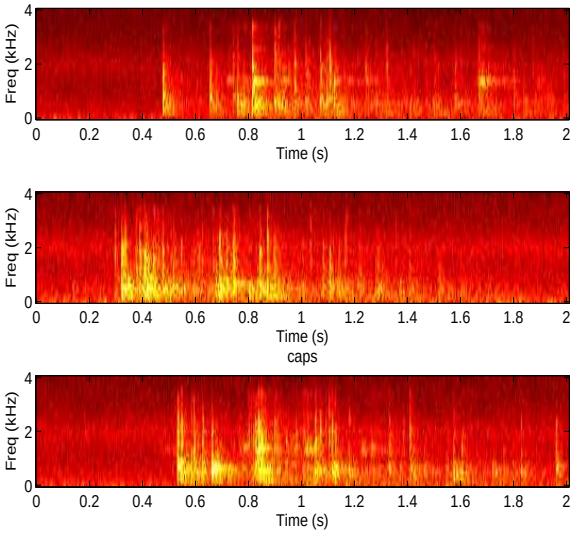

Figure 22: Three samples of class 22 “Caps”.

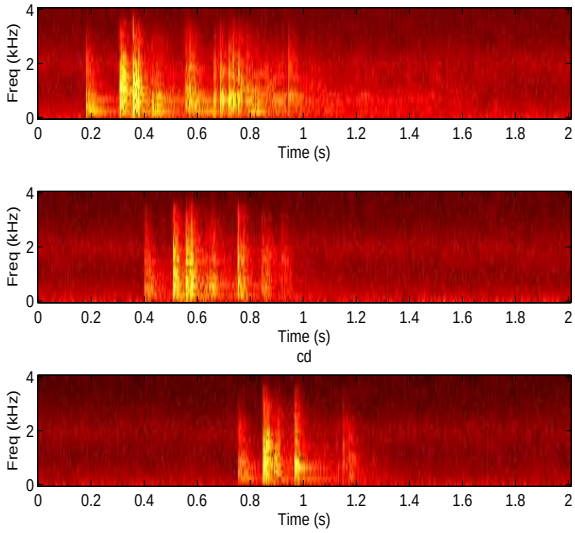

Figure 23: Three samples of class 23 “CD”.

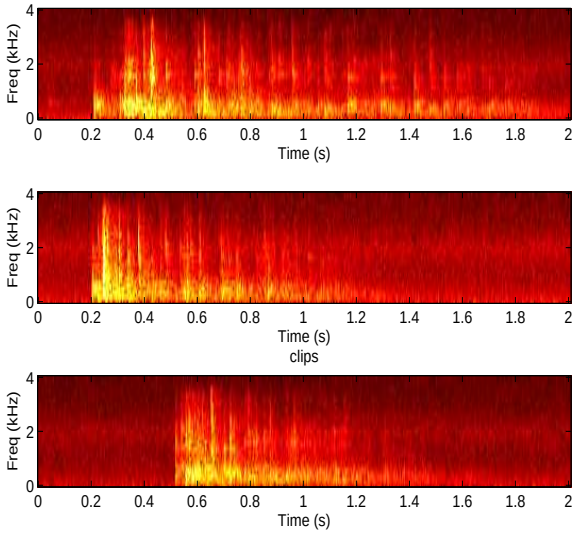

Figure 24: Three samples of class 24 “Clips”.