Audio Engineering Society

Convention Paper 9887

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

Analysis of Drum Machine Kick and Snare Sounds

Jordie Shier[1], Kirk McNally[1], and George Tzanetakis[1]

[1] *University of Victoria*

Correspondence should be addressed to Kirk McNally (kmcnally@uvic.ca)

ABSTRACT

The use of electronic drum samples is widespread in contemporary music productions, with music producers having an unprecedented number of samples available to them. The development of new tools to assist users organizing and managing libraries of this type requires comprehensive audio analysis that is distinct from that used for general classification or onset detection tasks. In this paper, 4230 kick and snare samples, representing 250 individual electronic drum machines are evaluated. Samples are segmented into different lengths and analyzed using comprehensive audio feature analysis. Audio classification is used to evaluate and compare the effect of this time segmentation and establish the overall effectiveness of the selected feature set. Results demonstrate that there is improvement in classification scores when using time segmentation as a pre-processing step.

1 Introduction

The first commercial electronic drum machine was released in 1959 by the Rudolph Wurlitzer Corporation. Marketed as an automatic rhythm accompaniment, the Side Man featured ten preset electronic drum sounds and twelve predefined rhythmic patterns. In recent decades the use of drum machine and drum samples has grown to such a degree that they are now ubiquitous in music productions, with contemporary music producers having an unprecedented number of drum samples available to them. Intelligent music production, including automated systems intended to aid creativity and improve user workflow, is a growing area of research. In [1] the drum track is identified as being of particular importance in electronic dance music and an experimental system to aid users in the creation of this musical element is studied in this work.

Another area that has been explored is the automatic classification and management of large drum sample libraries. Previous work includes the study by [2] where a set of features is used to analyze a large set of drum samples, including 33 different drum classes comprised of both acoustic and synthetic sources. In the work by [3] feature analysis and the use of metric learning and kernelized sorting to arrange drum samples in two dimensions is explored using the Audio-quilt interface. Related analysis studies include [4], where unpitched percussive sounds, including a mix of acoustic and synthetic kick and snare drum sounds are analyzed. In [5] acoustic kick and snare drum sounds are explored. In this paper, the focus is on achieving the best possible characterization of two specific drum classes: kick and snare drum samples. We explore using various time segmentation methods as a pre-processing step to audio analysis. Principle component analysis and audio classification is used to explore and evaluate the effect and efficacy of this approach.

2 Methods

2.1 Pre-processing

Pre-processing is applied to all samples prior to audio analysis. All audio samples are down-mixed to mono and resampled to a rate of 44.1kHz if required. A normalization step includes applying a ReplayGain of -6dB and an equal loudness filter. Various time segmentation methods are then applied as a final step during pre-processing.

2.1.1 Time Segmentation

Segmentation choices for this work are derived from [6], where a 23ms window is used to segment audio samples. A 50ms time window is used in [7] and it is reported in [5] that certain audio features can better characterize percussive sounds when using different sample lengths. In this work time segments of 25ms, 100ms and 250ms are used in the analysis. 500ms and full sample durations are also included for completeness. The starting position of a time segment is determined in relation to the signal envelope as the point in time when the signal reaches a threshold of its maximum value. Choices for this threshold are derived from the Essentia documentation[1] and include 20, 50, and 90 percent.

2.2 Audio Analysis

Audio feature extraction was performed using the Essentia library [8]. This library was selected based on the findings in [9], where it was found to be the most comprehensive library with regards to feature coverage. The features selected for use are from those defined both within the MPEG-7 format and prior work into the classification of percussion sounds. The features we used include Bark bands, MFCCs, HFC, spectral and temporal features, which together constitutes a 133-dimensional feature-space. A 2048 sample Hann window using a hop-size of 1/8, derived from [5], is used for the STFT required for computation of all spectral features. Calculations for each feature using the 2048 sample window are summarized over time using mean and standard deviation.

2.3 Principle Component Analysis

Principal component analysis (PCA), a technique that has been found useful in related intelligent music production tasks [10], is used to reduce the dimensionality of the original feature space. The principal components that result from PCA are a set of new axes that maximize the variance of the dataset, such that the first axis contains the most variance, the second axis contains the next most variance, etc. In [11] PCA is used to help characterize multitrack music mixes and explore the most relevant features in terms of the variance. We use PCA here to characterize kick and snare samples, using the methods described, and to examine the effect that time segmentation has on feature variance in lower dimensions.

3 Analysis

3.1 Feature Analysis

Analysis of the results from audio feature extraction demonstrates that the choice of the time segmentation has an effect on the variance for each feature. The variance of each audio feature responds uniquely to the time segmentation method used and the sample type (kick or snare) being analyzed. The most relevant feature to kick drum characterization as described by PCA was found to be the distribution of high frequency content (HFC). Feature distribution plots illustrating the effect of time segmentation for both kick and snare samples are shown in figures 1 and 2. Distribution of spectral energy, the most relevant feature to snare characterization as described by PCA, is shown for kick and snare samples in figures 3 and 4. These figures show how time segmentation effects the distribution of the selected features across a large set of samples as well as how kick and snare drum samples respond uniquely to specific features and segmentation methods.

3.2 Principal Component Analysis

Results of PCA give insight into how the time segmentation effects variance and which features are most useful for characterizing kick and snare drum samples. Variance is maximized in the first two dimensions when using a 100ms window starting at 20% of the attack for kick sounds, and a 250ms window starting at 90% of the attack for snare sounds. The first two dimensions explain 31.65% and 32.79% of the variance for kick

[1] `www.essentia.upf.edu/documentation`

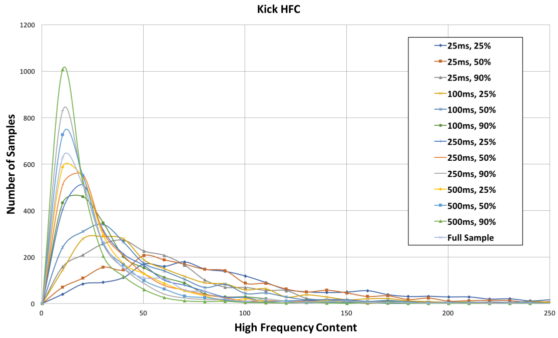

Fig. 1: Distribution of Kick HFC

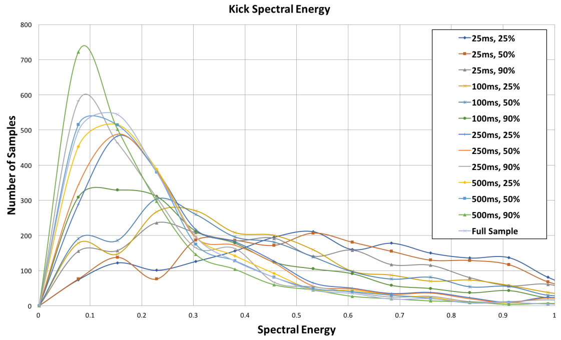

Fig. 3: Distribution of Kick Spectral Energy

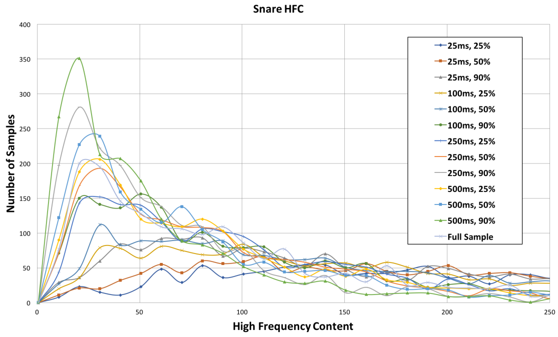

Fig. 2: Distribution of Snare HFC

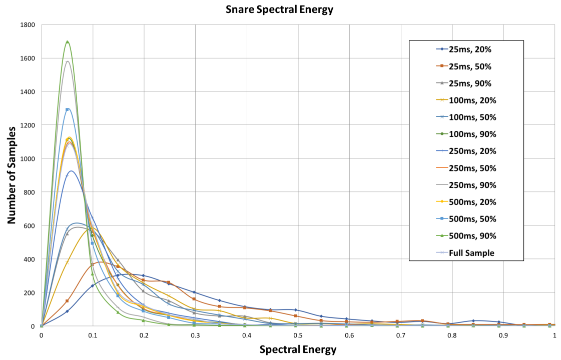

Fig. 4: Distribution of Snare Spectral Energy

and snare drums respectively. The main contributing features for the first dimension of kick drums after PCA are the mean and standard deviation of the HFC, and the high spectral energyband. The second MFCC band and the mean and standard deviation of the middle low spectral eneryband are main contributing features to the second dimension of kick drums. For snare drums, spectral energy (SE), and the 18th and 19th bark bands contribute highly to the first dimension and the standard deviation of bark spread, the standard deviation of the zero crossing rate, and MFCC band 5 are main contributors to the second dimension. Tables 1 and 2 summarize results for kick and snare drum analyses respectively.

4 Experimental Results

4.1 Classification

Different audio classification tasks are used to evaluate and compare the effect of the time segmentation choices. The classification tasks performed are sample type, drum machine and manufacturer classification. Three different classification algorithms implemented in Scikit-learn [12] are used: Support Vector Machine, Perceptron, and Random Forest. 10-fold cross-validation was used for each classification task and accuracy scores are calculated as an average between the three algorithms. ZeroR is used to determine the baseline accuracy for each task.

In order to test the effect of maximizing the variance of all features on classification results, a mixed time segmentation scheme was devised and included in testing. Using the results from Section 3.1, segment start position and length was selected independently for each feature such that the variance for that feature was maximized for the type of sample being classified. Results of all classification tasks performed are summarized in Table 3.

Table 1: Kick drum PCA results

Length	Start	Dim.1	Dim.2	Combined
25*ms*	20%	17.95%	12.46%	30.41%
25*ms*	50%	17.64%	11.93%	29.57%
25*ms*	90%	15.69%	11.53%	27.21%
100*ms*	20%	17.30%	14.35%	31.65%[3]
100*ms*	50%	16.72%	13.65%	30.37%
100*ms*	90%	15.62%	12.45%	28.08%
250*ms*	20%	17.01%	14.40%[2]	31.41%
250*ms*	50%	16.48%	13.83%	30.30%
250*ms*	90%	15.31%	12.92%	28.23%
500*ms*	20%	17.16%	13.52%	30.68%
500*ms*	50%	16.52%	13.10%	29.63%
500*ms*	90%	15.16%	12.70%	27.86%
Full	0%	18.03%[1]	13.44%	31.47%

Main contributing features:
[1] HFC, HFC Std Dev, Mid-High Spectral Energyband
[2] Spectral Flatness dB, Spectral Centroid, Spectral Kurtosis
[3] **Dim 1:** HFC, HFC Std Dev, High Spectral Energyband **Dim 2:** MFCC Band 2, Mid-Low Spectral Energyband Std Dev, Mid Low Spectral Energyband

Table 2: Snare drum PCA results

Length	Start	Dim.1	Dim.2	Combined
25*ms*	20%	17.11%	13.85%[2]	30.97%
25*ms*	50%	18.08%	13.73%	31.81%
25*ms*	90%	19.38%	12.81%	32.18%
100*ms*	20%	20.16%	11.03%	31.19%
100*ms*	50%	20.57%	10.75%	31.32%
100*ms*	90%	21.22%	10.15%	31.37%
250*ms*	20%	21.22%	10.73%	31.95%
250*ms*	50%	22.70%	10.54%	32.24%
250*ms*	90%	22.75%[1]	10.04%	32.79%[3]
500*ms*	20%	21.43%	10.19%	31.62%
500*ms*	50%	21.86%	10.01%	31.87%
500*ms*	90%	22.70%	9.69%	32.39%
Full	0%	21.01%	10.38%	31.39%

Main contributing features:
[1] Spectral Energy, Bark Band 18 and 19
[2] Spectral Decrease, Spectral Decrease Std Dev, Spectral RMS
[3] **Dim 1**:Spectral Energy, Bark Band 18 and 19 **Dim 2**: Bark Spread Std Dev, Zero Crossing Rate Std Dev, MFCC Band 5

4.1.1 Sample Type Classification

Sample type classification seeks to distinguish between kick and snare samples. All of the samples from the dataset were used and the baseline accuracy score was calculated to be 52.11%. The highest accuracy for kick and snare classification was 97.76 % and was achieved using a 250ms time segment positioned at 90% of the signal envelope.

4.1.2 Drum Machine Classification

For drum machine classification, machines were selected for kicks and snares separately such that each drum machine would have at least 50 samples for each type. Six distinct classes were used for kick drums which resulted in 464 kick samples in total and a baseline accuracy of 22.20%. The machines used for kick drums were the Alesis DM5, Alesis SR-16, Roland SH-09, Roland TR-808, Roland TR-909 and Yamaha RM50. Nine distinct classes were used for snare drums which resulted in 726 snare samples and a baseline accuracy of 16.54%. Machines used for snare classification were the Alesis DM5, Alesis SR-16, Boss DR-660, Roland TR-808, Roland TR-909, Yamaha CS-6, Yamaha RM 50, and Yamaha RY-30 Classification tasks performed best when using mixed time segments for both kicks and snares, reporting 84.20% and 69.88% accuracy respectively.

4.1.3 Manufacturer Classification

Manufacturers were selected such that each manufacturer was represented by at least 100 samples of each type. The same six manufacturers were used for both kicks and snares and included Alesis, Boss, E-MU, Korg, Roland, and Yamaha. For kick drums a total of 1329 samples were included reporting a baseline accuracy score of 39%. For snare drums a total of 1556 samples were used reporting a baseline accuracy of 33.10%. The results show that manufacturer classification was a more difficult task than the previous two tasks; kick drum classification reported 46.45% accuracy using a 500ms time segment at 20% of the signal envelope, and snares reported 48.21% accuracy using the full sample length.

5 Summary and Discussion

A dataset of 4230 drum machine samples was prepared and analyzed using comprehensive audio analysis and time segmentation as a preprocessing step. Results show that time segmentation effects the variance and distribution of each audio feature in a unique way, and

Table 3: Classification Results

Length	Start	Sample Type	Drum Machine		Manufacturer	
			Kick	**Snare**	**Kick**	**Snare**
25*ms*	20%	94.32%	74.86%	58.65%	44.95%	39.67%
25*ms*	50%	95.02%	72.99%	57.68%	42.89%	39.58%
25*ms*	90%	96.05%	71.26%	55.28%	41.08%	40.57%
100*ms*	20%	97.18%	81.97%	65.39%	45.77%	43.76%
100*ms*	50%	97.63%	80.46%	62.28%	46.45%	43.74%
100*ms*	90%	97.52%	76.01%	63.91%	45.42%	45.41%
250*ms*	20%	97.55%	81.75%	63.91%	44.07%	46.35%
250*ms*	50%	97.67%	76.01%	62.94%	44.09%	45.51%
250*ms*	90%	97.67%	69.18%	59.72%	44.32%	47.51%
500*ms*	20%	97.73%	76.80%	65.03%	44.77%	47.33%
500*ms*	50%	97.83%	77.01%	65.24%	44.14%	47.61%
500*ms*	90%	97.70%	71.34%	62.94%	42.34%	47.02%
Full[1]	0%	97.43%	79.09%	67.94%	44.54%	48.21%
Mixed[2]	-	97.52%	84.20%	69.88%	46.23%	46.03%

[1] Entire duration of sample

[2] Time segmentation and start position selected independently for each feature so that the variance for that feature is maximized

that utilizing different time segmentation methods, including a mixed segmentation approach, leads to improved classification results and ability to retain variance in lower dimensions when using PCA. This technique shows potential for improving the effectiveness of applications related to drum sound characterization and classification. The findings of this work will be useful for tasks associated to large percussion libraries within the field of intelligent music production.

Future work includes the perceptual testing of the techniques described in order to determine whether utilizing time segmentation methods as a pre-processing step leads to more perceptually relevant characterization of audio samples. The authors welcome any feedback and contributions on the GitHub page [2] in accordance with the recommendations for open access and reproducibility in signal processing re- search presented in [13]. The dataset is also available upon request.

References

[1] Vogl, R., Leimeister, M., Nuanáin, C. Ó., Jorda, S., Hlatky, M., and Knees, P., "An intelligent interface for drum pattern variation and comparative evaluation of algorithms," *Journal of the Audio Engineering Society*, 64(7/8), pp. 503–513, 2016.

[2] Herrera, P., Yeterian, A., and Gouyon, F., "Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques," in *Music and Artificial Intelligence*, pp. 69–80, Springer, 2002.

[3] Fried, O., Jin, Z., Oda, R., and Finkelstein, A., "AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting." in *NIME*, pp. 281–286, 2014.

[4] Herrera, P., Dehamel, A., and Gouyon, F., "Automatic labeling of unpitched percussion sounds," in *Audio Engineering Society Convention 114*, Audio Engineering Society, 2003.

[5] Pampalk, E., Herrera, P., and Goto, M., "Computational models of similarity for drum samples," *IEEE transactions on audio, speech, and language processing*, 16(2), pp. 408–423, 2008.

[6] Danielsen, A., Waadeland, C. H., Sundt, H. G., and Witek, M. A., "Effects of instructed timing and tempo on snare drum sound in drum kit performance," *The Journal of the Acoustical Society of America*, 138(4), pp. 2301–2316, 2015.

[7] Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and Sandler, M. B., "A tutorial on

[2] `https://github.com/jorshi/sample_analysis`

onset detection in music signals," *IEEE Transactions on speech and audio processing*, 13(5), pp. 1035–1047, 2005.

[8] Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X., "Essentia: an open-source library for sound and music analysis," in *Proceedings of the 21st ACM international conference on Multimedia*, pp. 855–858, ACM, 2013.

[9] Moffat, D., Ronan, D., Reiss, J. D., et al., "An evaluation of audio feature extraction toolboxes," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway*, 2015.

[10] Tzanetakis, G. and Cook, P., "3D graphics tools for sound collections," in *Proc. COSTG6 Conference on Digital Audio Effects, DAFX*, 2000.

[11] Wilson, A. and Fazenda, B., "Variation in multitrack mixes: analysis of low-level audio signal features," *Journal of the Audio Engineering Society*, 64(7/8), pp. 466–473, 2016.

[12] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, 12(Oct), pp. 2825–2830, 2011.

[13] Vandewalle, P., Kovacevic, J., and Vetterli, M., "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, 26(3), 2009.