SIEVE: A PLUGIN FOR THE AUTOMATIC CLASSIFICATION AND INTELLIGENT BROWSING OF KICK AND SNARE SAMPLES

Jordie Shier, Kirk McNally and George Tzanetakis

University of Victoria

Introduction and Motivation

- Audio plug-in designed to assist music producers with the sorting, selection and auditioning tasks associated with the use of large electronic kick and snare drum sample libraries within a music production context.
- A database of 4230 kick and snare samples representing 250 individual electronic drum machines was used in this study.

Methodology

- Analysis of kick and snare drum sounds using time segmentation as a pre-processing step to audio feature extraction informed by prior work into the characterization of percussive sounds. [1- 3].
- Audio feature extraction performed using the Essentia library [4].
- Audio features used include: Bark bands, MFCCs, HFC, spectral and temporal features (133-dimension feature space).
- Principle Component Analysis (PCA) is used to reduce the 133-dimensional feature space down to two dimensions for visual plotting on plug-in user interface.

Experimental Results

- Audio classification and PCA using Scikit-learn [5] is used to compare the effect of the time segmentation on audio feature extraction.
- Classification tasks included: sample type, drum machine, and manufacturer classification, using three different classification algorithms: Support Vector Machine, Perceptron, and Random Forest.
- Table 1 shows classification results.
- PCA provides insight into how time segmentation effects variance and which features are most useful for characterizing kick and snare drum samples.
- Table 2 shows PCA variance ratios for both Kick and Snare.

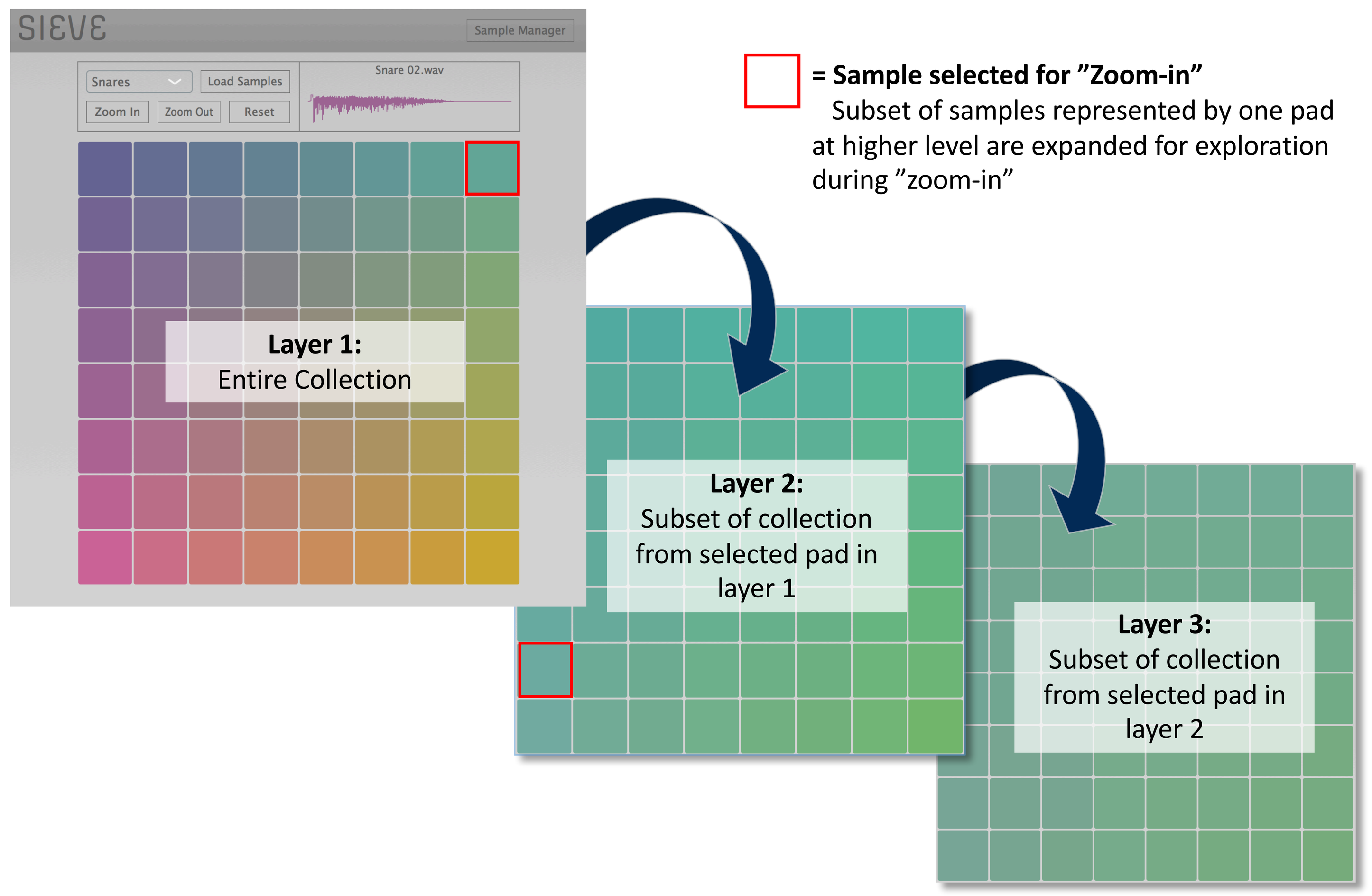

Figure 1: Sieve Plugin Interface and browsing overview

Table 1: Classification Scores for various tasks preprocessed with time segmentation

Length	Start	Sample Type	Drum Machine		Manufacturer	
			Kick	Snare	Kick	Snare
25ms	20%	94.32%	74.86%	58.65%	44.95%	39.67%
25ms	50%	95.02%	72.99%	57.68%	42.89%	39.58%
25ms	90%	96.05%	71.26%	55.28%	41.08%	40.57%
100ms	20%	97.18%	81.97%	65.39%	45.77%	43.76%
100ms	50%	97.63%	80.46%	62.28%	46.45%	43.74%
100ms	90%	97.52%	76.01%	63.91%	45.42%	45.41%
250ms	20%	97.55%	81.75%	63.91%	44.07%	46.35%
250ms	50%	97.67%	76.01%	62.94%	44.09%	45.51%
250ms	90%	97.67%	69.18%	59.72%	44.32%	47.51%
500ms	20%	97.73%	76.80%	65.03%	44.77%	47.33%
500ms	50%	97.83%	77.01%	65.24%	44.14%	47.61%
500ms	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

Table 2: Variance ratios from PCA after applying various time segmentations

Length	Start	Kick					Snare				
		Dim 1	Dim 2	Dim 3	Dim 4	1+2	Dim 1	Dim 2	Dim 3	Dim 4	1+2
25ms	20%	17.95%	12.46%	9.44%	7.91%	30.41%	17.11%	13.85%[5]	9.51%	5.48%	30.97%
25ms	50%	17.64%	11.93%	9.52%	7.65%	29.57%	18.08%	13.73%	9.39%	5.37%	31.81%
25ms	90%	15.69%	11.53%	9.60%	7.76%	27.21%	19.38%	12.81%	8.79%	5.36%	32.18%
100ms	20%	17.30%	14.35%	9.43%	7.80%	31.65%[3]	20.16%	11.03%	9.85%	5.93%	31.19%
100ms	50%	16.72%	13.65%	9.41%	8.22%	30.37%	20.75%	10.75%	9.77%	5.95%	31.32%
100ms	90%	15.62%	12.45%	10.57%	8.70%	28.08%	21.22%	10.5%	9.37%	6.26%	31.37%
250ms	20%	17.01%	14.40%[2]	8.95%	8.18%	31.41%	21.22%	10.73%	9.22%	6.52%	31.95%
250ms	50%	16.48%	13.83%	9.16%	8.16%	30.30%	21.70%	10.54%	9.08%	6.52%	32.24%
250ms	90%	15.31%	12.92%	9.91%	8.79%	28.23%	22.75%[4]	10.04%	8.89%	6.63%	32.79%[6]
500ms	20%	17.16%	13.52%	8.87%	7.83%	30.68%	21.43%	10.19%	9.15%	6.94%	31.62%
500ms	50%	16.52%	13.10%	8.92%	8.06%	29.63%	21.86%	10.01%	9.03%	6.97%	31.87%
500ms	90%	15.16%	12.70%	9.47%	8.71%	27.86%	22.70%	9.69%	8.71%	7.01%	32.39%
Full	0%	18.03%[1]	13.44%	9.07%	7.37%	31.47%	21.01%	10.38%	9.15%	7.08%	31.39%

Main contributing features:
[1] **Dim 1:** HFC, HFC Std Dev, Mid-High Spectral Energyband
[2] **Dim 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
[4] **Dim 1:** Spectral Energy, Bark Band 18 and 19
[5] **Dim 2:** Spectral Decrease, Spectral Decrease Std Dev, Spectral RMS
[6] **Dim 1**:Spectral Energy, Bark Band 18 and 19 **Dim 2**: Bark Spread Std Dev, Zero Crossing Rate Std Dev, MFCC Band 5

Audio Plugin Architecture

- Plugin developed using the JUCE C++ software framework.
- Essentia is used for feature extraction and computation of PCA.
- SQLite database was implemented to manage loaded samples, results of feature extraction and PCA.
- Figure 1 shows plugin UI and sample arrangement scheme.

Conclusion and Future Work

- Shorter time segments of audio for analysis improved the majority of kick and snare drum classification tasks.
- Manufacturer classification proved more challenging and is an area for future exploration.
- User tests using the plug-in are necessary to determine perceptual relevance of time segmentation scheme.
- Exploration of alternative methods of dimensionality reduction, for example, self organizing maps used in [6].

References

- [1] A. Danielsen, C. H. Waadeland, H. G. Sundt, and M. A. Witek, "Effects of instructed timing and tempo on snare drum sound in drum kit performance," Journal of the Acoustical Society of America , vol. 138, no. 4, pp. 2301–2316, 2015.
- [2] J. P. Bello et al. , "A tutorial on onset detection in music signals," IEEE Transactions on speech and audio processing , vol. 13, no. 5, pp. 1035–1047, 2005.
- [3] E. Pampalk, P. Herrera, and M. Goto, "Computational models of similarity for drum samples," IEEE Transactions on Audio, Speech, and Language Processing , vol. 16, no. 2, pp. 408–423, 2008.
- [4] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an open-source library for sound and music analysis," in 21st ACM International Conference on Multimedia , pp. 855–858, ACM, 2013.
- [5] F. Pedregosa et al. , "Scikit-learn: Machine learning in python," Journal of Machine Learning Research , vol. 12, no. Oct, pp. 2825–2830, 2011.
- [6] M. Cartwright, B. Pardo, and J. Reiss, "Mixploration: Rethinking the audio mixer interface," in The 19th international conference on Intelligent User Interfaces , pp. 365–370, ACM, 2014.