An algorithm for creating groove templates for application to programmed MIDI drum data

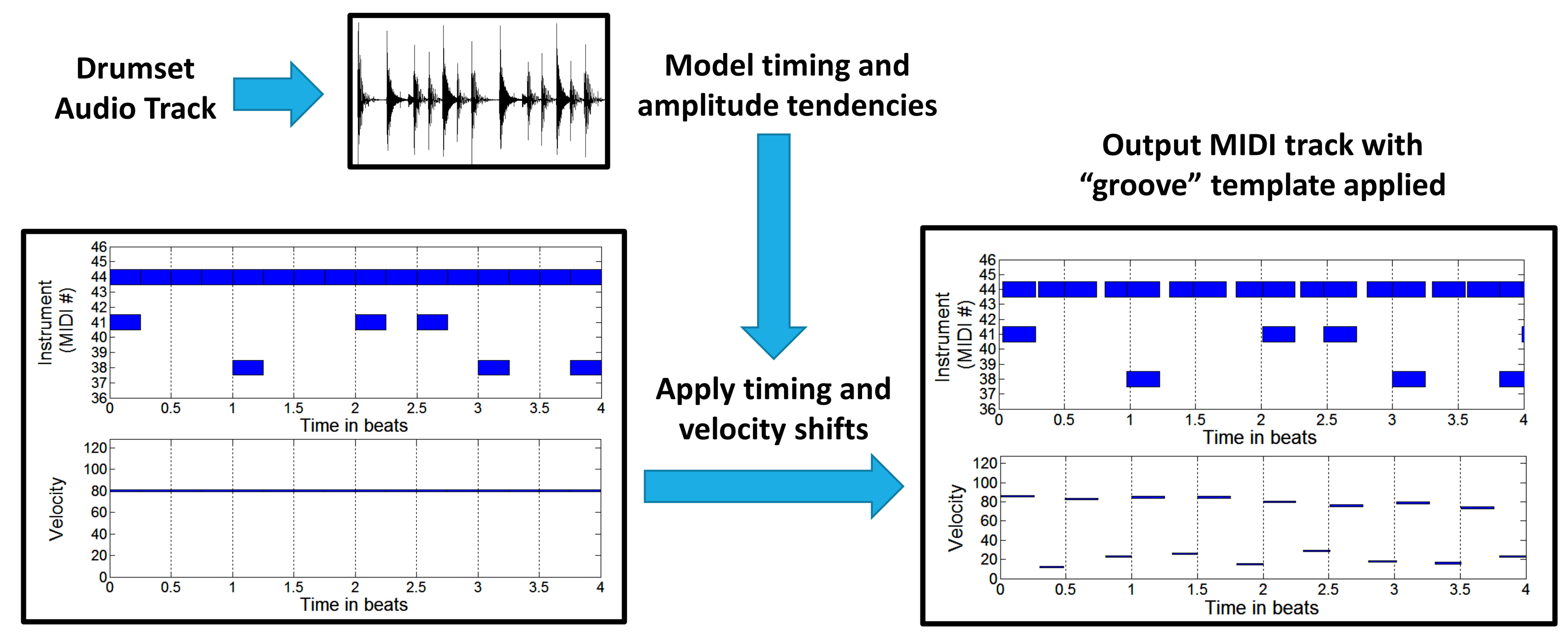
By Jeremy Hassett
AME 272 – Audio Signal Processing

Abstract

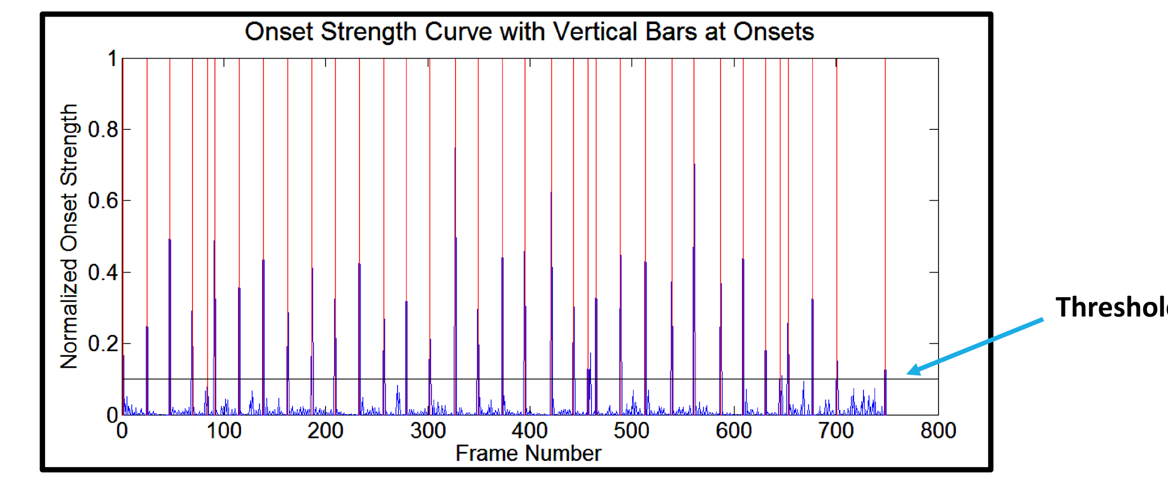
In order to provide an alternative to the existing high-cost and lengthy drum recording process, a method of providing human-like computer generated drum tracks is implemented. Drum recordings for most types of popular music must be conducted in suitable acoustic environments and require high-quality microphones to capture the sound effectively. The drummer must also play the parts without flaws, which often results in a lengthy recording process. As a result, drum beats for popular music are often programmed using Digital Audio Workstations (DAWs). This solution often produces very different results than those provided by a human drummer, as programmed drum patterns exhibit ideal timing and invariant amplitude among notes. Human drummers provide much more dynamic timing and note velocity variations, which result in more "natural" sounding drum tracks [1]. These tendencies are often unique to individual drummers as well as different styles of music. By investigating the reoccurring patterns of timing and velocity in a human performance, an effective groove template is realized and applied to programmed MIDI drum tracks to achieve a more realistic synthesized drum sound.

1. Tidemann, Axel, and Yiannis Demiris. "Imitating the Groove: Making Drum Machines More Human." *Proceeding KI '08 Proceedings of the 31st Annual German Conference on Advances in Artificial Intelligence* (2008): 144-51. Web.

Summary of algorithm functionality and results



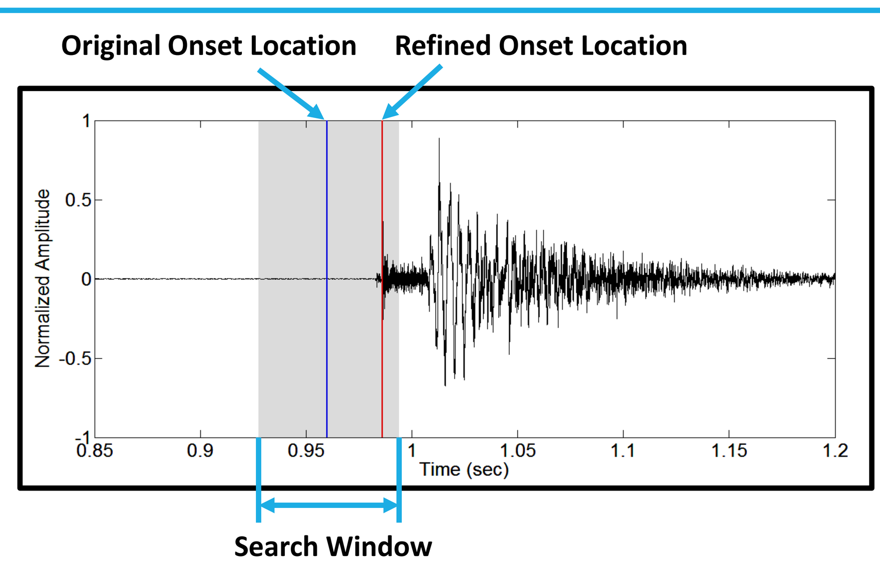
An effective onset detection method is realized by investigating the spectral-flux characteristics of the signal.



- Threshold for peak-picking is set intentionally low to provide more false positive detections than false negatives.
- A minimum onset separation is used to ensure that several onsets are not detected for a single onset event in the audio file.

2. J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. "A tutorial on onset detection in music signals." IEEE Trans. on Speech and Audio Processing, vol. 13, no. 5, (2005).

Onset locations are refined by performing a local onset search using an energy-based onset detection method.



- Detected onsets using spectral method are often slightly displaced from actual onsets due to the reduced sampling rate of the short-time Fourier transform (hop size = 1024 samples).
- Energy-based method allows much higher location resolution (hop size = 4 samples).

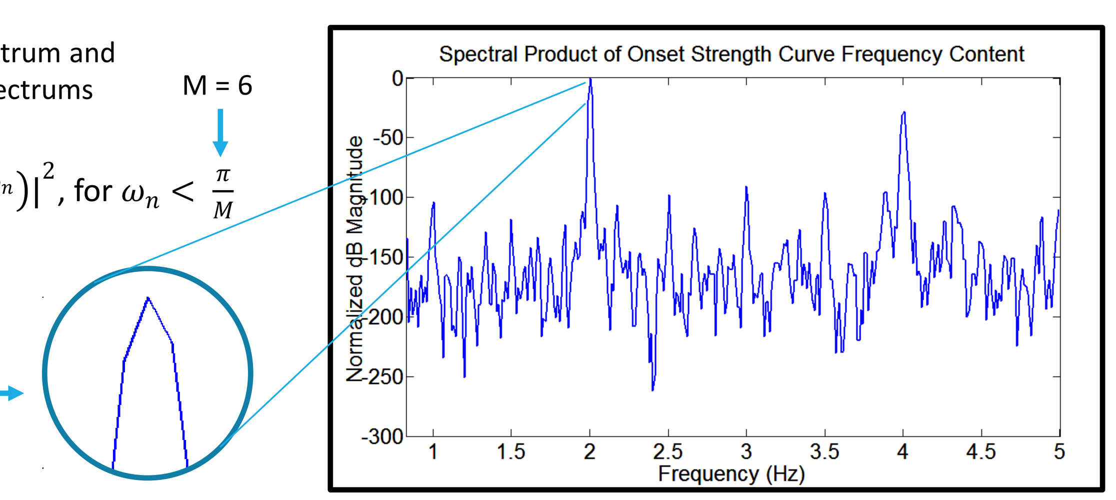
2. J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. "A tutorial on onset detection in music signals." IEEE Trans. on Speech and Audio Processing, vol. 13, no. 5, (2005).

Tempo of audio is accurately determined using modified spectral product pitch detection method.

Spectrum for tempo search — Product of spectrum and compressed spectrums — M = 6

$$S(e^{j\omega_n}) = \prod_{l=1}^{M} \left|X\left(e^{jl\omega_n}\right)\right|^2, \text{ for } \omega_n < \frac{\pi}{M}$$

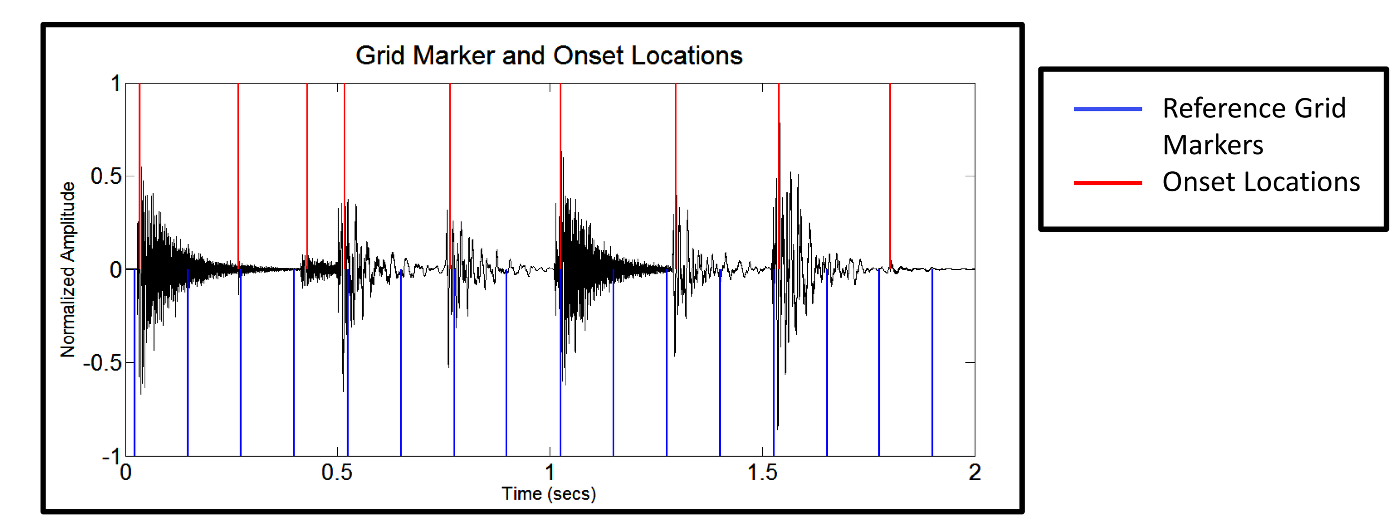
Find center of peak using quadratic interpolation



- Assumption is that spectrum of onset strength curve will exhibit strong peaks at multiples of its fundamental frequency, which typically corresponds to the tempo of the piece.
- Spectrum is multiplied with compressed version of the same spectrum to further emphasize these peaks.

3. Alonso, M., B. David, and G. Richard. "A Study of Tempo Tracking Algorithms from Polyphonic Music Signals." Ecole Nationale Superieure Des Telecommunications (ENST), 1 Apr. 2003. Web.

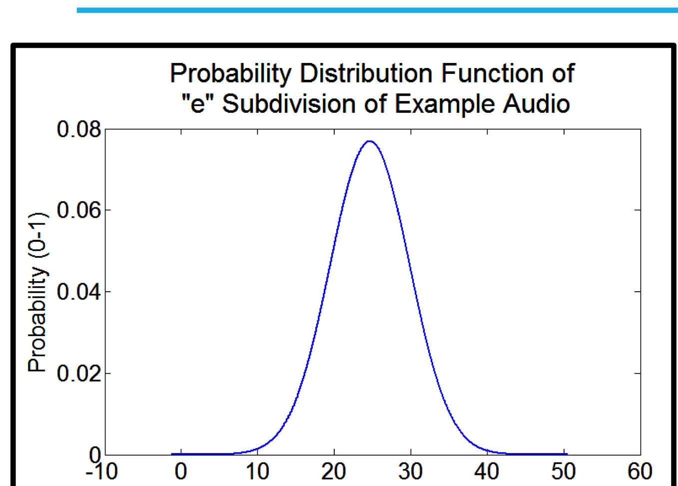
A reference grid of subdivided beats is superimposed onto signal by maximizing the cross-correlation between audio and comb signal.

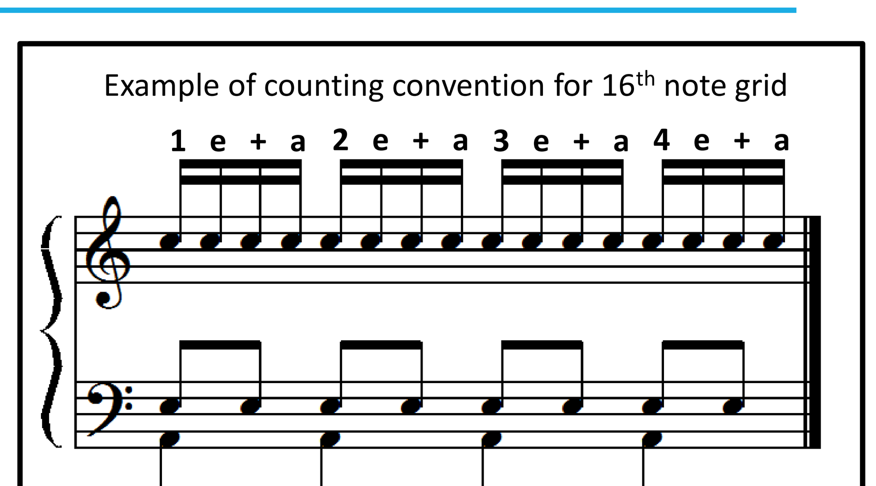


- Grid spacing is calculated using tempo information and phase is found by maximizing cross-correlation between audio signal and evenly spaced impulse train.
- Shifts of onsets from their closest grid marker are tracked in order to model timing tendencies.

4. Duan, Zhiyao. "Lecture 6: Rhythm Analysis." ECE 272/472 (AME 272, TEE 272) – Audio Signal Processing, 2015.

Timing differences between actual onsets and reference grid markers are mapped with normal distributions for each beat subdivision.

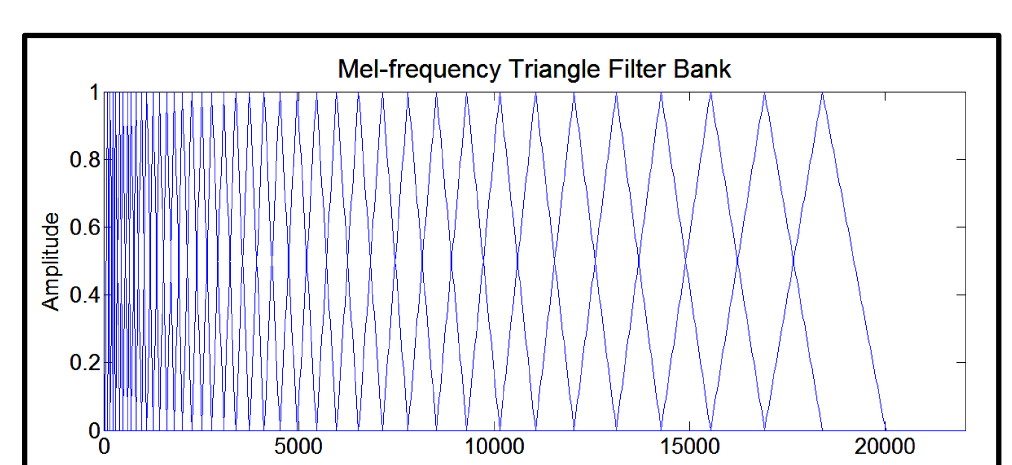




- Timing variation of human drummers is effectively modelled with the normal distribution function [1].
- Tendencies for onsets to lie a certain distance from grid markers is correlated to the onsets beat subdivision to provide consistent timing variations and model the "feel" of a drum beat.

1. Tidemann, Axel, and Yiannis Demiris. "Imitating the Groove: Making Drum Machines More Human." *Proceeding KI '08 Proceedings of the 31st Annual German Conference on Advances in Artificial Intelligence* (2008): 144-51. Web.

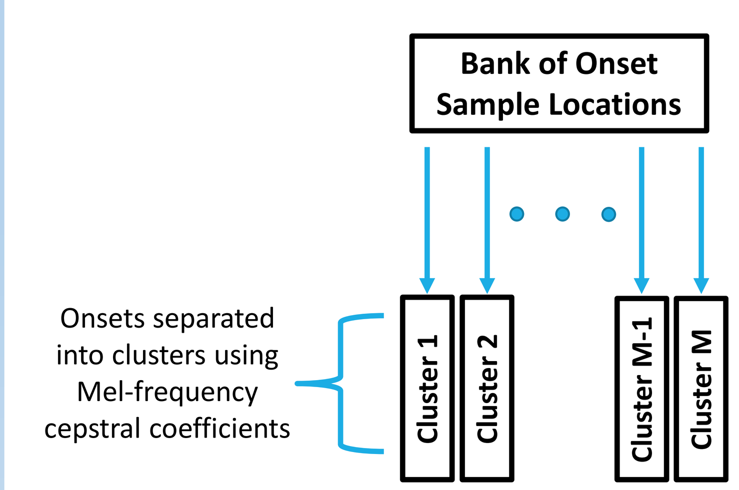
Drum instrument sources are separated using hierarchal clustering of Mel-frequency cepstral coefficients.



- Triangle filters are applied to power spectrum of signal and power is summed within for each filter.
- Power for each filter is used signal of length N, where N represents the total number of filters.
- Fourier transform of this signal gives Mel-frequency cepstral coefficients (First 12-13 are used for source separation).

5. "Mel Frequency Cepstral Coefficient (MFCC) Tutorial." *Practical Cryptography*. N.p., n.d. Web. 28 Apr. 2015.

Onset relative strengths are calculated with respect to mean amplitude within each cluster before being mapped to MIDI velocity.



RMS amplitude for each onset:

$RMS = \sqrt{\langle x^2 \rangle}$, where $\langle x^2 \rangle$ denotes the mean of the square of the signal throughout local window surrounding onset.

Relative amplitude for each onset:

$RA = 20\log_{10}\left(\frac{RMS}{\langle C_{n,RMS} \rangle}\right)$, where $\langle C_{n,RMS} \rangle$ is the mean RMS value for the given cluster.

- Relative amplitudes (in dB) are mapped linearly to shifts in MIDI velocity.
- MIDI velocity shifts are also modelled using the normal distribution function for each beat subdivision.