Homework 3: Onset Detection

Introduction

Onset detection deals with finding a starting point of each note or a specific sound in a recording. While this might be a fairly straightforward problem in monophonic music, when it comes to polyphonic music due to numerous overlaps in the musical notes it is often hard to tell individually when exactly each note started. The paper I will be overviewing provides several algorithms that use different approaches for onset detection. Namely, spectral flux based approach, phase deviation approach and complex domain methods. Moreover, the paper also mentions several potential improvement suggestions.

Methodology Overview

When creating an onset function we want the peaks of the function to coincide (match) with the times of note onsets. While onset functions ultimately have a goal of detecting change in audio signal, due to ever-changing nature of audio signal and various types of changes such as onsets, offsets, vibratos, amplitude modulations and noise, they have a goal of detecting very specific changes. Moreover, onset detection functions are able to achieve data reduction by using a low sampling rate while simultaneously preserving necessary information to detect onsets. In order to achieve aforementioned lower frame rate, the signal is processed using short time Fourier transform with a Hamming window $w(m)$ at a frame rate of 100 Hz:

$$X(n,k) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(hn+m)w(m)e^{-\frac{2j\pi mk}{N}}$$

where window size $N = 2048$ equal to 46 ms at a sampling rate of 44.1 kHz and hop size $h = 441$ equal to 10 ms.

Spectral Flux

Spectral flux allows us to measure change in magnitude of each frequency bin. By computing spectral flux restricted to positive changes and summing it across all frequency bins:

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|)$$

where

$$H(x) = \frac{x + |x|}{2}$$

which represents half-wave rectifier function. Moreover, paper states that empirical comparison of using L_1 and L_2 norms favored usage of L_1-norm in combination with linear magnitude instead of logarithmic. Results of the onset detection of this algorithm using a drum sample as an input can be seen in Figure 1.

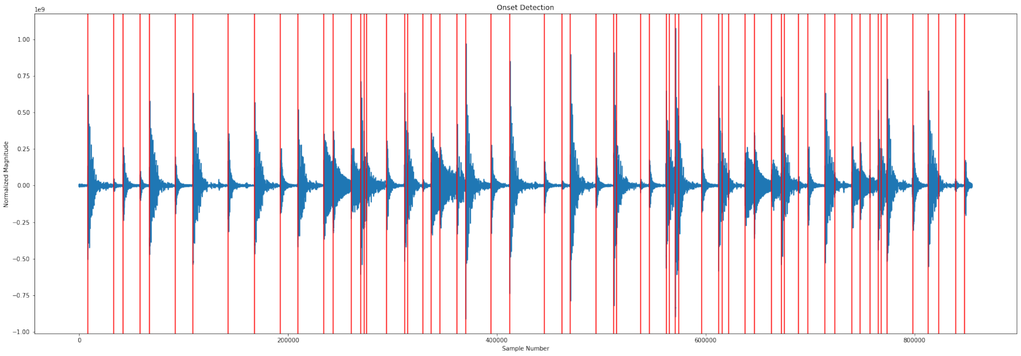

Figure 1: Showcase of onset detection using spectral flux.

Phase Deviation

Another method we can use to detect onset is tracking of phase rate of change in an STFT frequency bin. Given that our signal is defined as:

$$X(n,k) = |X(n,k)|e^{j\psi(n,k)}$$

where $-\pi < \psi(n,k) \leq \pi$. Instantaneous frequency is given by the first derivative

$$\psi'(n,k) = \psi(n,k) - \psi(n-1,k)$$

mapped onto the same range as the original phase. The change in instantaneous frequency change is then given by the second difference of the phase:

$$\psi''(n,k) = \psi'(n,k) - \psi'(n-1,k)$$

once again mapped onto the same range. Taking mean absolute changes across all bins we obtain our onset detection function:

$$PD(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |\psi''(n,k)|$$

Complex Domain

Instead of looking at phase and magnitude individually, we can consider them jointly to detect sudden changes. We define target value as:

$$X_T(n,k) = |X(n-1,k)|e^{\psi(n-1,k)+\psi'(n-1,k)}$$

Hereafter, our onset detection function becomes sum of absolute deviations from the target:

$$CD(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n,k) - X_T(n,k)|$$

Weighted Phase Deviation

This approach is provided in the paper as an improvement over the regular phase deviation method. It proposes weighting different frequency bins by their corresponding magnitude:

$$WPD(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n,k)\psi''(n,k)|$$

This is done in order to alleviate the problem of the noise introduced by components that have no significant energy in the original phase deviation method. Another improvement can be done by normalizing the weighted phase deviation function:

$$NWPD(n) = \frac{\frac{1}{N}\sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n,k)\psi''(n,k)|}{\frac{1}{N}\sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n,k)|}$$

Rectified Complex Domain

Last mentioned method deals with the inability of complex domain method to distinguish between increases and decreases of the amplitude. By using a similar approach to spectral flux method of half-wave rectification we ensure that we preserve only the increases of energy:

$$RCD(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} RCD(n,k)$$

$$RCD(n,k) = \begin{cases} |X(n,k) - X_T(n,k)|, \text{ if} |X(n,k)| \geq |X(n-1,k)| \\ 0, \text{ otherwise} \end{cases}$$

Onset Selection

Lastly, the paper mentions that setup of onset selection threshold is very important. Setting a threshold too high reduces the number of false positives, but also increases the chance of a false negative, and setting it too low achieves the opposite effect. Therefore, we need to find a balance between the two where our function performs the best. Threshold value also depends on use case and how undesirable false positives are in comparison to false negatives.
Peak picking is done by normalizing each onset function f(n) - meaning it has a mean of 0 and standard deviation of 1. Hereafter, the peak is selected as an onset if it fulfils following three equations:

$$f(n) \geq f(k) \text{ for all } k \text{ such that } n - w \leq k \leq n + w$$

$$f(n) \geq \frac{\sum_{k=n-mw}^{n+w} f(k)}{mw + w + 1} + \delta$$

$$f(n) \geq g_\alpha(n-1)$$

where time of the peak can be computed using $t = \frac{nh}{r}$. Window size and multipliers used were both $w = m = 3$. Threshold function $g_\alpha(n)$ is given by:

$$g_\alpha(n) = max(f(n), \alpha g_\alpha(n-1) + (1-\alpha)f(n))$$

α and δ are arbitrary constants that are to be experienced with to see which setup gives the best results. For personal testing I selected $\delta = 0.4$ and omitted usage of last equation as it did not improve the outcome of the tested example.

Recent Developments

With significant modern developments in artificial intelligence algorithms, there have been significant advancements made in terms of onset detection. Novel approaches harness the power of deep learning to surpass previously available algorithms. Current state of the art by Schulter & Bock [2] was introduced in 2014. and is based on usage of convolutional neural networks.

Conclusion

In this summary I showed several methods of detecting onsets in audio recordings accompanied by an implementation of the spectral flux based approach. While this methods work really well and are computationally efficient, recent developments show that usage of new approaches can significantly improve results.

References

[1] Dixon S. Onset detection revisited. InProceedings of the 9th International Conference on Digital Audio Effects 2006 Sep 18 (Vol. 120, pp. 133-137).

[2] Schluter, Jan & Bock, Sebastian. (2014). Improved musical onset detection with Convolutional Neural Networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 6979-6983. 10.1109/ICASSP.2014.6854953.