AUDIO ONSET DETECTION USING MATHEMATICAL MORPHOLOGY

Martin Eckart
University of Guelph
Department of Computer Science
meckart@uoguelph.ca

ABSTRACT

A novel note and beat onset detection algorithm is proposed using mathematical morphology. It is intended to be simple and fast and should work well on instruments with some attack characteristics.

1. INTRODUCTION

Mathematical morphology is a branch of image and signal process originally developed in the 1960s for textural analysis of iron ores [3]. At its most basic, mathematical morphology employs dilation $\oplus$ and erosion $\ominus$ operations along with basic algebra to create more complex non-linear filters. The advantage is that images as well as 1-dimensional signals can be processed as geometrical objects. This paper proposes a simple but novel audio onset detection algorithm based on morphological operations. Running in the time domain without regard to any frequency information, this morphological algorithm processes audio quickly (around 50-100 seconds of audio to each second of processing on modern CPUs) and is somewhat robust in the presence of noise. However, it has difficulty distinguishing instrument timbres that are light in the attack.

2. MATHEMATICAL MORPHOLOGY IN SIGNAL PROCESSING

Although it is mostly used in image processing, mathematical morphology has relevant applications to audio and signal processing. Morphological operations use a structuring element which can be understood in 1-dimensional signals as a sliding window of fixed size which calculates a local minimum (erosion) or maximum (dilation). Figure 1 shows the dilation and erosion of a 1-dimensional signal by a flat structuring element (a straight line 50 samples long). Note that these operations do not take signal polarity into account and -1 is considered a minimum.

Morphological processes can be combined to create filters that are useful for localized peak picking and edge finding. Iterating opening operations (erosion followed by

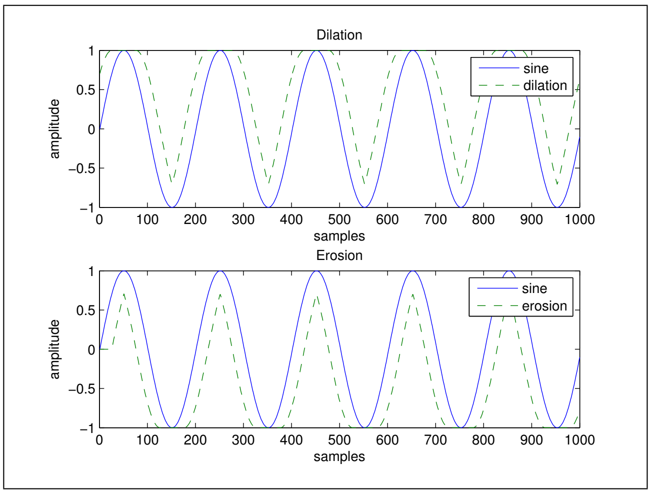

Figure 1. Dilation and erosion of a sine wave with a linear structuring element 50 samples long.

dilation) for example could be used to separate higher harmonic information in a low-pass filter [2]. In the proposed onset detection system, morphological operations are used to generalize the waveform intensity and find quickly rising edges which would denote onset attacks. These operations are further described in the method section.

3. METHOD

The proposed morphological onset detector is an analysis of amplitude intensity in the time domain and follows a basic process of *waveform conditioning*, *onset isolation*, *filtering* and *peak picking*. Each of these steps are described along with a discussion of the relevant morphological operations used.

3.1 Waveform Conditioning

Since morphological operations do not take signal polarity into account, the absolute value of the waveform is taken in order to create a graph of signal intensity. Then the maximum intensity over 50ms is calculated by dilating the waveform with a linear structuring element 50ms wide. Whereas root-mean-square (RMS) calculation would average the intensity within a 50ms window and flatten the waveform somewhat, dilation preserves larger edges by acting on maximum values, effectively outlining the waveform as seen in Figure 2.

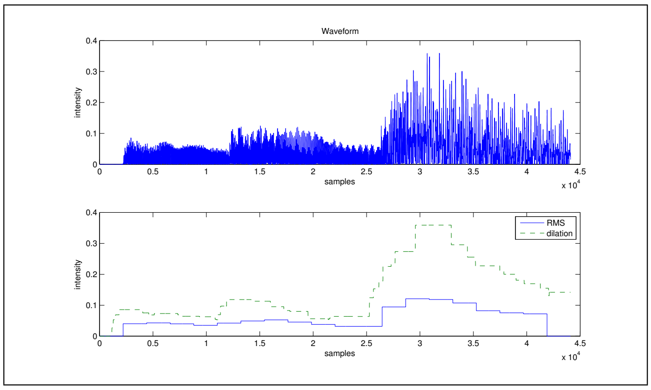

Figure 2. Comparison of RMS and dilation for 3 notes played on piano. The RMS window and structuring element are both 50ms wide.

3.2 Onset Isolation

In order to identify onset attacks, the system looks for edges which are defined by large and quick gradients from low to high intensity. The morphological internal gradient is found by subtracting an erosion from the original signal [1]:

$$G_i(f) = f - f \ominus b \tag{1}$$

Where G_i is the internal gradient, f is the original signal and $f \ominus b$ is the erosion of f. In order to process only the edges which go from low to high intensity and not high to low, a structuring element of half a line is used as in the following example of a linear structuring element 9 samples wide:

$$\begin{bmatrix}1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0\end{bmatrix} \tag{2}$$

Figure 3 shows the result of calculating the morphological internal gradient on the dilated piano note waveform. The dilation is closed then opened to filter out small dips or spikes before applying the gradient filter.

3.3 Filtering and Peak Picking

Insignificant onsets are filtered out of the internal gradient by treating each spike as a separate object and summing the area below. Any gradient object sum below a given threshold is filtered out. For each gradient object, the first local maximum value is recorded as the note onset in a very simple peak picking algorithm.

4. DISCUSSION

It is expected that this algorithm will run very quickly due to the simplistic nature of the morphological operations and ignorance of any frequency domain characteristics. As such, it is reasonable to conjecture that the algorithm will not perform well when analyzing strings or other pitched instruments with nonexistent or slow attacks. It is expected that rhythmic and stronger attacking instruments such as woodwinds will produce much better results. In order to maximize detection results for varying instrument classes, more experimentation must be done by varying the size of

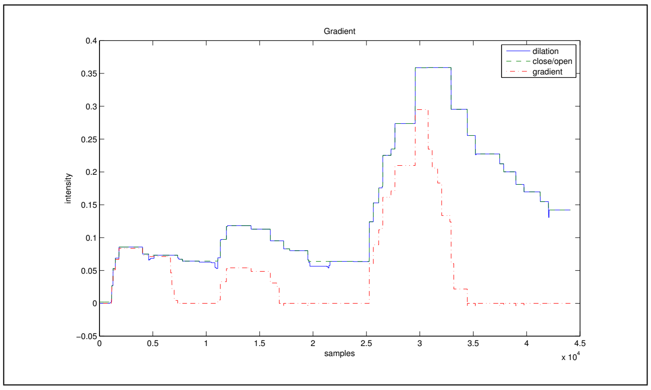

Figure 3. Internal gradient of piano waveform dilation from Figure 2.

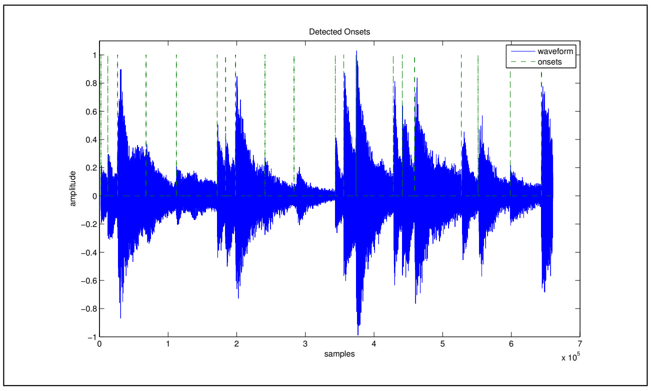

Figure 4. Detected onsets in 15 second piano excerpt.

structuring elements. For the moment, the structuring elements chosen are rather arbitrary.

The intended value of using this novel morphological process is to provide a fast pre-processing onset detection algorithm for use in other applications such as beat tracking and synchronization.

5. CONCLUSION AND FUTURE WORK

This novel morphological onset detection algorithm is fast and provides reasonable results when analyzing instruments with some attack characteristics. The algorithm does not perform well with pitched instruments that exhibit slow attacks such as stringed instruments. In order to increase the robustness of the algorithm, some analysis should also be done in the frequency domain. This can include setting up a series of bandpass filters (implementable in mathematical morphology using openings and closings) or doing morphological image analysis on short time Fourier transforms (STFT). This will slow processing time but would certainly increase robustness for detecting pitched onsets. Emphasis should also be placed on making this algorithm run in real-time as the processing footprint is rather low and could be effective in situations when many different instruments are being monitored for onsets.

6. REFERENCES

[1] John Goutsias, Luc M. Vincent, and Dan S. Bloomberg. *Mathematical morphology and its applications to image and signal processing*. Springer, 2000.

[2] Z Lu, QH Wu, and J Fitch. A Morphological Filter For Estimation of Power System Harmonics. *International Conference on Power System Technology*, (2):1–5, 2006.

[3] Jean Paul Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.