A PERCEPTUALLY BASED ONSET DETECTOR FOR REAL-TIME AND OFFLINE AUDIO PARSING

William Brent

American University, Washington DC
Audio Technology
w@williambrent.com

ABSTRACT

Use of the perceptually determined Bark frequency scale is investigated in an implementation of spectral difference onset detection. The study is carried out using a new external object for onset detection in Pure Data called bark∼, and contrasted with a similar object called bonk∼. While the filterbanks used in both of these objects feature more detailed resolution in low frequency bands, bark∼'s reliance on the Bark scale is intended to make transitional filter spacing from low to high frequency more appropriately gradual. A weighting curve is also applied in order to account for variabilities associated with equal loudness curves. In three performance evaluations involving piano note onsets, it is shown that bark∼ produced fewer erroneous reports and was able to capture onsets that bonk∼ (similarly configured) failed to report.

1. INTRODUCTION

At present, onset detection has been explored from a variety of angles. Many standard techniques exist and are often used in parallel to achieve the best possible results. An excellent overview is provided in [1], which describes functions that aim to identify sharp changes in amplitude, general spectral content, high frequency content, and phase. Here, the focus will be an aspect of the spectral difference approach, which compares the energy present in various frequency bands across two successive analysis frames. When total growth in bands of the latter frame surpasses a given threshold, an onset is reported. This technique is valued for its ability to capture onsets in cases where the overall amplitude remains roughly constant.

Onset detection based on spectral difference was implemented for real-time use in Pure Data (Pd) as an object called bonk∼ [3]. By default, its filterbank is composed of 11 filters, with the first three spaced by an equal amount of Hz over low frequencies, and the remaining spaced exponentially at middle and high frequencies. Creation arguments for the object allow for custom filter parameters that may change the number of linearly spaced filters, but the transition from linear to exponential is always a sudden one.

This paper investigates the effect of applying the Bark frequency scale in spectral difference onset detection, and provides a preliminary report on a new external object for Pd. Like bonk∼, the object introduced here (bark∼) looks for abrupt growth in bands of a signal's spectrum. However, its filters are spaced equally throughout the audio spectrum according to the Bark frequency scale. The resulting low frequency detail is comparable to that of linear spacing, but wider spacing in Hz towards higher frequencies is achieved in a more gradual fashion. Figure 1 shows the result of equal Bark spacing in relation to linear frequency up to 5 kHz. Filter width increases smoothly across the range. It is hoped that this strategy will be beneficial in onset detection, where choosing the appropriate emphasis for various frequency bands can be difficult.

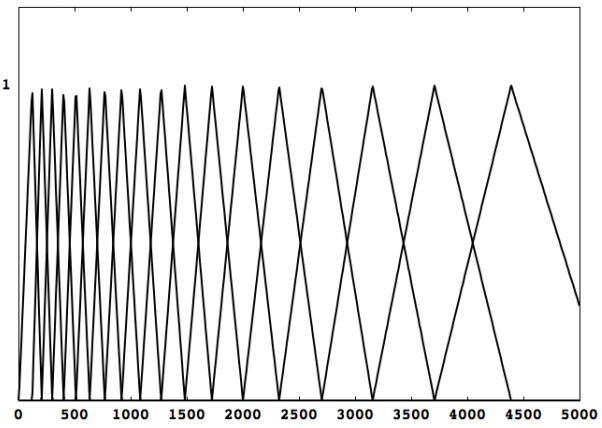

Figure 1. Overlapping filters spaced equally in Barks.

2. BASIS FOR THE BARK SCALE

The authors of [4] note strong relationships between curves plotting frequency in Hz against critical bandwidths, just-noticeable frequency difference, and the difference in frequency required to advance the point of maximum stimulation of the basilar membrane by 0.2 mm. The similarity of these curves indicates that the Bark scale based on documented critical bandwidths has physiological as well as perceptual validity. The relevance of critical bandwidth data has been verified repeatedly using differing approaches. Experiments using four unique strategies (based on threshold, masking, phase, and loudness summation) to establish the boundary and center frequencies of critical bands are reviewed in [5].

3. OBJECT DESIGN

Apart from employing the Bark scale, the onset detection algorithm used in bark~ is very similar to that described for bonk~ in [3]. As analysis frames are captured, relative growth is measured in each filter band with respect to the previous frame, and accumulated to produce a measure of total growth in each frame. Total growth can then be compared with appropriate thresholds to determine the presence of a new onset. An energy mask is applied in each band in order to prevent reporting of additional onsets that may occur due to spectral flux immediately following an attack. Once detected, onsets generate output from bark~'s three outlets, giving a list of growth values in each band, the total growth, and a "bang".

3.1. Settings and features

Several of bark~'s parameters can be configured while searching for optimal onset detection conditions. The object's creation arguments are analysis window size and hop in samples, and spacing between filters in Barks. Larger Bark spacings will create fewer filters. The default spacing is 0.5 Barks, which creates 47 filters.

Other options related to bark~'s filterbank are the use of power vs. magnitude spectrum, and the choice of scaling energy in each filter relative to its width. Additionally, a subset of filters can be specified for use in the total growth sum. Under these conditions, growth in other bands will not be accumulated into the sum. This may be useful for cases where it is known that unique spectral characteristics of an instrument's attack occur in a certain frequency band. For instance, this feature can be used to consider high frequency content exclusively.

As with bonk~, lower and upper growth thresholds can be chosen in order to identify a trend toward stability after sudden spectral growth. An onset is reported when total growth is larger than the upper threshold then falls below the lower threshold. The lower threshold can be omitted so that an onset is reported the moment that total growth beyond the upper threshold begins to show any sign of decay. A "measure" function is provided to aid in the choice of appropriate thresholds. It reports the average and peak growth obtained during the measurement period. Measuring a period of relative silence or background noise will thus provide information useful for the tuning process.

The parameters of bark~'s energy mask can be altered to control its rate of decay over time. As with bonk~, this is specified by two values: the first determines the number of analysis periods for which masking will be complete, and the second determines how quickly the mask will decay (e.g., by a factor of 0.75 in every frame thereafter). Another strategy for avoiding erroneous onset reports immediately following an attack is to simply ignore growth patterns for a given amount of time. This can be set using bark~'s "debounce" message, which disables onset reporting and schedules a clock callback to re-enable it after a specified number of milliseconds.

Finally, a weighting curve based on equal loudness contours can be applied to filterbank output before total growth is measured. The most noteworthy effect of this option is a compensatory reduction of energy in the lowest frequency bands, as a much greater amount of energy is required for low frequencies to be perceived as equal in loudness to mid-frequency tones. This may aid in cases where perceptually unobtrusive low-frequency sounds cause undesirable onset reports.

Unlike bonk~, bark~ does not offer a means for classifying instrument attacks according to the spectral properties of onsets. However, instrument classification according to Bark spectrum is already provided by another library of Pd externals [2].

4. A PRELIMINARY PERFORMANCE EVALUATION

To gauge the impact that bark~'s perceptual components have on accurate onset detection, a simple performance evaluation was carried out relative to bonk~. The piano was chosen as an appropriately challenging yet reasonable instrument on which to detect attacks. In this context, the instrument's most interesting characteristics are extremely large pitch and dynamic ranges, and the potential for a tremendous amount of background resonance with complex patterns of spectrum and loudness fluctuation.

To serve as test input, three pieces were selected from the piano literature based on having relatively short durations, relaxed tempi, varying dynamics, and exploiting a large portion of the piano's full playing range. These pieces are Satie's first *Gymnopédie* (performed by Aldo Ciccolini, 1966), the second movement of Beethoven's "Waldstein" sonata, Op. 53 (performed by Emil Gilels, 1972), and *Le Gibet* from Ravel's *Gaspard de la Nuit* (performed by Ivo Pogorelich, 1981). The pieces are considered in order of complexity, with the Satie—having simple rhythms and plenty of inter-onset time between notes—being the least challenging. Durations and number of onsets for each piece are given in Table 1.

	Duration	No. of onsets
Satie	3'03"	188
Beethoven	4'38"	253
Ravel	6'52"	407

Table 1. Durations and number of onsets for test input pieces.

All of these stereo recordings were reduced to the left channel only, but kept at the standard sampling rate and bit depth of 44.1 kHz and 16 bits respectively. Because a variety of variable parameters are offered by both bonk~ and bark~, it is difficult to make a comparison aimed at isolating the effects of Bark spacing and loudness weighting. Based on the author's experience using these objects in practice, the best possible effort was made to achieve an optimal tuning state for both objects in each case. In all

of the tests below, window size and hop were set to 2048 and 128 samples respectively for both objects. To make bonk~'s filterbank as comparable as possible to that of bark~, "numfilters" was set to 47 and "halftones" was set to 1.9. The break point between linearly and exponentially spaced filters resulting from these settings was around 540 Hz.

4.1. Satie: *1st Gymnopédie*

As the shortest inter-onset time between notes in this recording is about 550 ms, a debounce setting of 200 ms was used to reduce the number of erroneous onset reports immediately following an attack. The debounce feature for bonk~ is only available during its "learn" mode, so an equivalent debounce system was arranged using additional objects in the patch. The lower and upper thresholds for bonk~ were carefully chosen so that spectral flux during the sustain portion of notes produced a minimum of false positives, and were set to 2 and 6 respectively. Minimum velocity ("minvel") was set to 7. Thresholds for bark~ were set so that onsets would be reported immediately upon any sign of a decrease in growth once the upper growth threshold of 5.5 was surpassed. The spectrum weighting curve was applied as well. Masking parameters for both objects were set to mask completely for 4 analysis periods, then decay by 80% in each subsequent frame.

A total of 188 onsets were manually identified in the recording. In these tests, expressive timing differences between the right and left hands were treated as if they were simultaneous—that is, as written in the score. In cases where the performer articulated a bass note before a (nominally simultaneous) melody note, detection of the bass note alone was considered an accurate report, even if debounce settings suppressed reporting of the melody note that followed.

Neither object failed to catch articulations of primary melody notes in the mid- and upper-registers, but a total of 6 bass notes that were captured by bark~ went undetected by bonk~. For example, the D2 in measure 10 is very quiet beneath the existing resonance of previous notes. To account for the low amplitude of these missed bass notes, bonk~'s "minvel" parameter was set as low as 1, but this failed to improve detection.

The test patches were designed to record onset reports to separate label files. Figure 2 shows labels for measures 9 and 10 produced by bark~ and bonk~ after being imported into Audacity. The third label produced by bark~ (upper labels) relates to the quiet D2 described above. The fact that this attack is not visible in the waveform's amplitude points out the need for onset detection based on spectral growth in general, and the absence of a label in bonk~'s output (lower labels) can be seen as well. The following label in bark~'s output catches the next attack, while bonk~ erroneously reported additional onsets immediately following the actual attack. Such false positives are caused by fluctuating resonance from previous chords.

Repeated adjustments were made to the masking, minimum velocity, and threshold settings in an attempt to reduce errors. Settings that reduced extraneous onset reports caused an increase in attacks that were missed altogether, and vice-versa. In the best case, resonance remaining after actual attacks triggered a total of 51 false positives from bonk~. This occurred only 8 times using bark~. Results are summarized in Table 2, with true positives given as a percent of total onsets in the piece. As onset detection by bonk~ is already reasonably accurate for this recording, the differences are not drastic. However, bark~ produced fewer errors, and was more sensitive to extremely quiet bass note attacks.

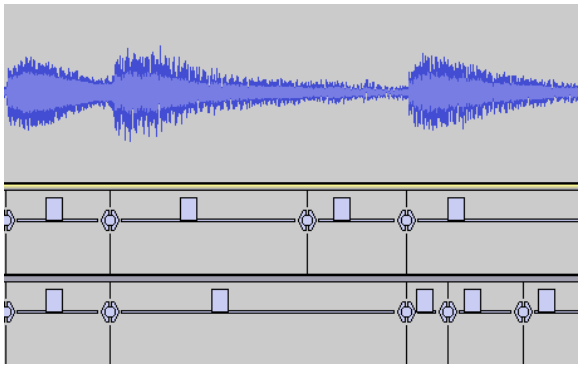

Figure 2. Label files generated by bark~ (upper labels) and bonk~ (lower labels) imported into Audacity.

	True positives (%)	False positives
bonk~	96.81	51
bark~	100	8

Table 2. True positives and no. of false positives in Satie.

4.2. Beethoven: Op. 53, 2nd movement

Tuning processes for the Beethoven example yielded similar settings. For bonk~, best results were achieved with upper and lower thresholds set to 3 and 7. Minimum velocity was set at 7, and masking was set to decay at a rate of 80% after 4 analysis periods. For bark~, the spectrum weighting curve was active, and the upper threshold was set to 5.25 (with the lower threshold again omitted). Masking parameters were identical to those used for bonk~. The debounce setting was lowered to 100 ms for both objects, as the shortest inter-onset time in the piece was just over 100 ms.

In comparison with the Satie, this piece is longer and has more onsets (253). It is also more varied in terms of rhythm and figuration. The most challenging areas were those with grace notes, repeated notes, and quiet bass notes articulated beneath sustained mid- and high-frequency resonance. The first onset missed by either object—in fact, both objects—was the quiet C3 grace note at the outset of measure 12. In this case, the two previous notes were sustained and at the same pitch, so there was not a sufficient level of spectral change.

The final measures were by far the most problematic,

for which both objects failed to detect several quiet bass notes. Through the highly resonant passage in measures 17-27, bonk~ missed 16 of the 70 bass notes (most of which are unsupported by simultaneous upper register notes), while bark~ missed 12. Overall, bark~ reported fewer false positives and slightly more true positives. Results are given in Table 3.

	True positives (%)	False positives
bonk~	91.3	36
bark~	94.47	8

Table 3. True positives and no. of false positives in Beethoven.

4.3. Ravel: *Le Gibet*

The final test considered the second movement of Ravel's *Gaspard de la Nuit*. With 407 onsets, a very broad dynamic range, and three staves for the majority of the piece (which requires extensive use of the sustain pedal), it presented a difficult challenge in terms of tuning. Several incidental noises in the recording, such as creaking of the piano bench, were detected by both objects. These were noted but not considered in the report below.

After repeated adjustments, bonk~ was configured with lower and upper thresholds of 3 and 6.25, a minimum velocity of 7, and masking decay as before (complete masking for 4 periods, and 80% decay thereafter). Identical masking was used for bark~. The upper growth threshold for bark~ was set to 5 (with lower threshold again omitted), and the loudness weighting curve was active. A debounce setting of 100 ms was used for both objects.

The underlying motive for the entire piece involves quiet pairs of repeated B-flats, with the second less accentuated than the first. Because this motive persists at a low dynamic beneath an increasingly thick harmonic texture, it was a challenge for both objects. Over the first 14 measures, there are 69 of these B-flat instances. With the settings described above, bark~ failed to detect 4, while bonk~ missed 11. For both objects, the B-flat motive was most difficult to detect when preceded by relatively loud bass octaves, such as in measure 17. The small amount of spectral growth caused by the quiet motive was apparently overshadowed by changes related to fluctuating bass resonance. Similar failures were found in moments with resonance from relatively quiet upper register chords, such as in measure 23. Here, spectral growth from the B-flat motive (now an A-sharp) is lost beneath complex patterns of spectral flux.

Overall results are shown in Table 4, where it can be seen that bark~ once again captured a greater number of true positives and fewer false positives than bonk~.

	True positives (%)	False positives
bonk~	84.77	49
bark~	92.38	12

Table 4. True positives and no. of false positives in Ravel.

5. CONCLUSION

The reports above indicate that a filterbank based on the Bark scale and spectrum weighting based on equal loudness contours is beneficial for piano onset detection. With respect to tuning parameters, configuring bonk~ to have the highest number of true positives caused the number of false positives to be much higher than that of bark~. Generally, much less time was needed to find parameter settings for bark~ that led to reasonable levels of accuracy.

Differences in the number of true positives between the two onset reports were for the most part related to a failure of bonk~ to detect quiet notes in the middle and bass registers during moments with significant previous resonance. Thus, the potential benefits offered by bark~ may be most relevant for instruments with a large frequency range and a great capacity for sustained resonance. In an additional study not described here, guitar onsets were considered, and the output from both objects was roughly the same.

As non-real-time audio parsing is useful in a variety of sample-based synthesis applications, an alternate version of bark~ was designed to analyze samples loaded into RAM via Pd's graphical arrays. Binaries and source code for both versions (as well as the patches used to generate labels in these tests) are available for download from the author's website.

6. REFERENCES

[1] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in music signals," in *IEEE Transactions on Speech and Audio Processing*, 2005.

[2] W. Brent, "A timbre analysis and classification toolkit for pure data," in *Proceedings of the International Computer Music Conference*, 2010, pp. 224–229.

[3] M. Puckette, T. Apel, and D. Zicarelli, "Real-time audio analysis tools for pd and msp," in *Proceedings of the International Computer Music Conference*, 1998, pp. 109–112.

[4] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Berlin: Springer Verlag, 1990.

[5] E. Zwicker, G. Flottorp, and S. Stevens, "Critical bandwidth in loudness summation," *Journal of the Acoustical Society of America*, vol. 29, pp. 548–557, 1957.