ONSET DETECTION REVISITED

Simon Dixon

Austrian Research Institute for Artificial Intelligence
Freyung 6/6, Vienna 1010, Austria
simon.dixon@ofai.at

ABSTRACT

Various methods have been proposed for detecting the onset times of musical notes in audio signals. We examine recent work on onset detection using spectral features such as the magnitude, phase and complex domain representations, and propose improvements to these methods: a weighted phase deviation function and a half-wave rectified complex difference. These new algorithms are compared with several state-of-the-art algorithms from the literature, and these are tested using a standard data set of short excerpts from a range of instruments (1060 onsets), plus a much larger data set of piano music (106054 onsets). Some of the results contradict previously published results and suggest that a similarly high level of performance can be obtained with a magnitude-based (spectral flux), a phase-based (weighted phase deviation) or a complex domain (complex difference) onset detection function.

1. INTRODUCTION

Many music signal analysis applications require the accurate detection of onsets of musical tones, and it is not surprising that several different methods have been proposed for performing onset detection. At first sight, onset detection is a well-defined task: the aim is to find the starting time of each musical note (where a musical note is not restricted to those having a clear pitch or harmonic partials). However, in polyphonic music, where nominally simultaneous notes (chords) might be spread over tens of milliseconds, the definition of onsets starts to become blurred. Likewise, instruments with long attack times (e.g. flute) produce notes for which it is difficult to define an unambiguous and precise onset time.

The most natural way to approach this problem is to attempt to define onset times in line with human perception, for example using the work of Vos and Rasch [1], who distinguished between the physical and perceptual onset times of musical tones, and showed that the perceptual onset time occurs when the tone reaches a level of approximately 6 – 15 dB below its maximum value. However, their research did not deal with the type of situations faced in analysing audio recordings of complex musical works, where factors such as masking, temporal order thresholds and just noticeable differences lay to rest any hope of a crisp definition of onset for real-world data.

In this paper, we take a more pragmatic approach to onset detection, allowing the available data sets to guide the definition of onsets, sometimes corresponding to perceived onset times, and sometimes to physical onset times. The bulk of the first data set is hand-labelled, that is, listeners have marked the positions in the audio file at which they perceive onsets. If multiple listeners are used for each file, a reasonably robust data set can be developed, but this involves much work, so only small data sets can be produced in this way. The remaining data is collected from computer-monitored pianos, which are able to measure the physical onset times of notes with a high degree of accuracy. Since it is not feasible for listeners to annotate the perceived onsets of all 106054 notes in this collection of piano sonatas, this is a method by which large data sets can be quickly collected. The only drawback is the disparity between physical and perceptual onset times (assuming that the goal is to find perceived onsets). In the case of percussive instruments, the difference is of the order of a few milliseconds, which is sufficiently precise for our purposes, and at least as accurate as human-labelled data.

In a recent tutorial article, Bello et al. [2] reviewed a number of onset detection algorithms, making a theoretical and empirical comparison of several state-of-the-art approaches. In this paper, we complement and extend their work by introducing new onset detection functions based on their work, and by testing the new methods alongside independent implementations of a subset of the published methods on the same data set and on a second data set which is two orders of magnitude larger. Other comparisons of onset detection methods can be found in [3, 4]. We restrict our comparison to methods based on short term spectral coefficients, which are the most widely used methods, and the most successful according to the 2005 MIREX audio onset detection evaluation [4].

In the next section, we introduce onset detection functions, review three state-of-the-art functions as presented in [2, 5], and describe three new functions which are extensions of the published algorithms. Section 3 addresses the evaluation of these onset detection functions, starting with methodological concerns, then describing the test data and finally presenting and discussing the results of the tests. The final section contains the conclusions and some ideas for further work.

2. ONSET DETECTION FUNCTIONS

An onset detection function is a function whose peaks are intended to coincide with the times of note onsets. Onset detection functions usually have a low sampling rate (e.g. 100Hz) compared to audio signals; thus they achieve a high level of data reduction whilst preserving the necessary information about onsets. Most onset detection functions are based on the idea of detecting changes in one or more properties of the audio signal. But audio signals, whether composed of natural or synthetic sounds, are in a continual state of change, so the task of onset detection also involves distinguishing between the various types of change, such as onsets, offsets, vibrato, amplitude modulation and noise.

If an audio signal is observed in the time-frequency plane, the onset of a new sound has noticeable energy in the frequency bands in which the sound is not masked by other simultaneous components. Thus an increase in energy (or amplitude) within some frequency band(s) is a simple indicator of an onset. Alternatively, if we consider the phase of the signal in various frequency bands, it is unlikely that the frequency components of the new sound are in phase with previous sounds, so irregularities in the phase of various frequency components can also indicate the presence of an onset. Further, the phase and energy (or magnitude) can be combined in various ways to produce more complex onset detection functions. This is the basis of the onset detection functions presented in this section.

The next 3 subsections briefly review existing approaches to onset detection using spectral flux, phase deviation and complex domain methods (for a more in depth review, see [2]). Then we present potential improvements to these methods, defining weighted phase deviation, normalised weighted phase deviation and half-wave rectified complex domain onset detection functions. All of these methods make use of a time-frequency representation of the signal based on a short time Fourier transform using a Hamming window $w(m)$, and calculated at a frame rate of 100 Hz. If $X(n,k)$ represents the kth frequency bin of the nth frame, then:

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

where the window size $N = 2048$ (46 ms at a sampling rate of $r = 44100$ Hz) and hop size $h = 441$ (10 ms, or 78.5% overlap).

2.1. Spectral Flux

Spectral flux measures the change in magnitude in each frequency bin, and if this is restricted to the positive changes and summed across all frequency bins, it gives the onset function SF [6]:

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

where $H(x) = \frac{x+|x|}{2}$ is the half-wave rectifier function. Empirical tests favoured the use of the L_1-norm here over the L_2-norm used in [7, 2], and the linear magnitude over the logarithmic (relative or normalised) function proposed by Klapuri [8].

2.2. Phase Deviation

The rate of change of phase in an STFT frequency bin is an estimate of the instantaneous frequency of that component. This can be calculated via the first difference of the phase of $X(n,k)$. Let $\psi(n,k)$ be the phase of $X(n,k)$, that is:

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

where $-\pi < \psi(n,k) \leq \pi$. Then the instantaneous frequency is given by the first difference $\psi'(n,k)$:

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

mapped onto the range $(-\pi, \pi]$. The change in instantaneous frequency, which is an indicator of a possible onset, is given by the second difference of the phase:

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

which is also mapped onto the range $(-\pi, \pi]$. Large discontinuities in the unwrapped phase or its derivatives can wrap around to 0, but the onset detection function based on phase deviation, PD, takes the mean of the absolute changes in instantaneous frequency across all bins [9, 2], which reduces the chance of a missed detection:

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

2.3. Complex Domain

Amplitude and phase can be considered jointly to search for departures from steady-state behaviour by calculating the expected amplitude and phase of the current bin $X(n,k)$, based on the previous two bins $X(n-1,k)$ and $X(n-2,k)$. The target value $X_T(n,k)$ is estimated by assuming constant amplitude and rate of phase change:

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

and therefore a complex domain onset detection function CD can be defined as the sum of absolute deviations from the target values:

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

This formulation is simpler but equivalent to the complex domain detection function in [2, 5].

2.4. Weighted Phase Deviation

In the remainder of this section we propose improvements to the onset detection functions described in the literature. The first idea addresses the problem that the PD function "is susceptible ... to noise introduced by components with no significant energy" [2]. That is, the function considers all frequency bins k equally, although the energy of the signal is concentrated within the bins containing the partials of the currently sounding tones. We propose weighting the frequency bins by their magnitude, giving a new onset detection function which we call the *weighted phase deviation* (WPD):

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

This is similar to the CD function, in that the magnitude and phase are considered jointly, but with a different manner of combination. A further option is to define a *normalised weighted phase deviation* ($NWPD$) function, where the sum of the weights is factored out to give a weighted average phase deviation:

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

2.5. Rectified Complex Domain

One problem with the CD method is that it does not distinguish between increases and decreases in amplitude of the signal. Since it is important to distinguish onsets from offsets, we propose using a similar idea to that used in the SF function, where half-wave rectification is used to preserve only the increases in energy in spectral bins. This idea can easily be incorporated into the CD method, giving a (half-wave) rectified complex domain (RCD) onset detection function as follows:

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

where

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

2.6. Onset Selection

The onsets are selected from the detection function by a peak-picking algorithm which finds local maxima in the detection function, subject to various constraints. The thresholds and constraints used in peak-picking have a large impact on the results, specifically on the ratio of false positives to false negatives. For example, a higher threshold generally reduces the number of false positives and increases the number of false negatives. The best values for thresholds are dependent on the application and the relative undesirability of false positives and false negatives. It is difficult to generate threshold values automatically, so we follow Bello et al. [2] in reporting results for optimal parameter settings, which also allows a fair comparison with their published results.

Peak picking is performed as follows: each onset detection function $f(n)$ is normalised to have a mean of 0 and standard deviation of 1. Then a peak at time $t = \frac{nh}{r}$ is selected as an onset if it fulfils the following three conditions:

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

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

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

where $w = 3$ is the size of the window used to find a local maximum, $m = 3$ is a multiplier so that the mean is calculated over a larger range before the peak, δ is the threshold above the local mean which an onset must reach, and $g_\alpha(n)$ is a threshold function with parameter α given by:

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

Experiments were performed with various values of the two parameters δ and α, and it was found that best results were obtained using both parameters, but the improvement in results due to the use of the function $g_\alpha(n)$ was marginal, assuming a suitable value for δ is chosen.

3. EVALUATION OF ONSET DETECTION FUNCTIONS

In this section, we discuss our testing methodology, describe the data sets, and present the results from testing the above onset detection functions on each data set.

3.1. Methodology

The main difficulty with the evaluation of onset detection algorithms is that of obtaining a significantly large and balanced set of recordings for which the onset times are known (ground truth data). Precise measurements of onset times are only available for a small fraction of music, such as piano music recorded on computer-monitored pianos, and music generated with a MIDI synthesiser. Other data must be labelled by hand, which is a laborious and error-prone task. By balanced, we mean that the data set should be representative of the full range of data that the system is intended to be used for, including the proportions of pieces per instrument, per style, per level of complexity, etc. If the test set is not representative of real-world data, then the results reported will be overly optimistic (or in some cases pessimistic) of the actual performance of the algorithm. Finally, the data set should be large enough that separate training and test sets can be established in order to avoid overfitting.

A second methodological problem is determining how to report and compare results. Each onset detection function has parameters which can be tuned to alter the proportion of false positives (reported detections where no onset exists) and false negatives (missed detections). These proportions are often expressed in terms of precision and recall (defined below), but it is not sufficient to report the precision and recall alone, that is, via one pair of values. The relationship of precision to recall is often shown graphically in a receiver operating characteristic (ROC) curve, and if a single scalar statistic is desired (to make comparisons simple), the the precision and recall can be combined into a single value such as the the area under the ROC curve or the F-measure, which represents the optimal point on the ROC curve.

A third problem is how to deal with situations where a number of onsets are very close together, for example when a chord is played on a guitar or piano. Depending on the time between the notes, one or more onsets might be perceived, but this is dependent on the instrument and presence of other simultaneous sounds. The MIREX 2005 [10] onset detection evaluation addressed this problem by counting the number of merged onsets (two onsets detected as a single onset) and double onsets (a single onset recognised as two) in addition to the standard counts of correct detections, false positives and false negatives[1].

In this work, we consider an onset to be correctly matched if a detected onset is reported within 50 ms of the ground truth onset time. We do not penalise merged onsets, since the data we have contains many simultaneous or almost-simultaneous notes, and we are not attempting to recognise the notes. The results are summarised by three statistics: the precision P, recall R and F-measure F (for the optimal parameter settings), which are given by:

$$P = \frac{c}{c + f^+}$$

$$R = \frac{c}{c + f^-}$$

$$F = \frac{2PR}{P + R} = \frac{2c}{2c + f^+ + f^-}$$

where c is the number of correct detections, f^+ is the number of false positives and f^- is the number of false negatives. Parameters were chosen to maximise F; for certain applications where false positives and false negatives are not equally undesirable, different parameter values would be more suitable. Further discussion of onset detection evaluation can be found in [11].

3.2. Data

The first set of tests were performed on the data used by Bello et al. [2], consisting of 4 sets of short excerpts from a range of instruments, classed into the following groups: NP — non-pitched percussion, such as drums (119 onsets); PP — pitched percussion, such as piano and guitar (577 onsets); PN — pitched non-percussion, in this case solo violin (93 onsets); and CM — complex mixtures from popular and jazz music (271 onsets). Although there are only 1060 onsets in these sets of excerpts, they offer two important advantages: to test the algorithms on a range of different

[1] http://www.music-ir.org/mirex2005/index.php/Audio_Onset_Detection

instruments, and second, to enable a direct comparison with other published work.

The second set of data contains about 4 hours of solo piano music played by a professional pianist on a Bösendorfer computer-monitored grand piano[2]. This data consists of 106054 notes — two orders of magnitude more than that used in other evaluations — and includes complex passages such as trills, fast scale passages with pedal and arpeggiated chords. The level of complexity is such that a human annotator would have immense difficulty marking all the onsets precisely.

3.3. Results and Discussion

Table 1 shows the results for 8 different onset detection functions tested on the 4 data sets used in [2]. In each case, the results are shown for the point on the ROC curve which gives the maximum value of the F-measure. That is, the ground-truth data was used to select optimal values of δ and α. A similar approach was taken in [2], so the comparison with the published results, which are included in the table in the rows marked by asterisks (SF* and PD*), is fair.

The first point to note is that there are some large discrepancies between the published results and our own implementations of the same functions. For example, SF* performs particularly well with the data set PP in comparison with the PN and NP data sets, but our implementation (SF) shows much smaller performance differences across these 3 sets of excerpts. SF also achieves better performance across the entire range of data, presumably due to a better peak-picking function. Even greater differences are evident in the results of the phase deviation functions, where our PD function achieved much worse performance than the published PD* results. The closeness of the PD* results to the WPD and NWPD results raised the suspicion that perhaps some weighting scheme been used in the PD* algorithm, and this was later confirmed by one of the authors, who had mistakenly thought it was an unimportant detail. It is noteworthy that relatively small differences in implementation have a large impact on results, and that some of the differences are specific to particular data sets. However, we add two caveats: first, that parameter settings greatly influence the results, and this could be the source of some of the differences, and second, that the differences in performance are not necessarily significant considering the size of the test sets.

The second point that we note from Table 1 is that the WPD and NWPD are both very significant improvements on the PD function, but the normalisation is only an improvement on the WPD in two cases (PP and CM), while for the other two cases a slight degradation in performance results. Finally, the RCD method offers a small improvement on the CD on this data, but considering the small size of the data set, this difference might not be significant.

Overall, these results show that spectral flux, weighted phase deviation and complex domain methods can all achieve a similarly high level of performance on these data sets. Since the data sets are small and not sufficiently general, we are not willing to draw further conclusions about the differences between these methods, except to state that spectral flux has the advantage of being the simplest and fastest algorithm.

We now turn to the second set of results (shown in Table 2), which comes from a much larger but more homogeneous set of data, 13 complete piano sonatas by Mozart. Results were collected using a single parameter set for each function, and then summed across all sonatas, but the F-measure was also calculated separately for each piece, to give an indication of the variation in results. Since piano is a pitched percussion instrument, one would expect the results to be similar to those shown in the PP columns of Table 1, and this is basically true. The top 3 algorithms are the same in both cases (SF, CD and RCD), with a range of less than one standard deviation between them. The fact that their order has changed does not appear to be significant. The remaining algorithms preserved their ranking, with NWPD very close to the third placed algorithm, WPD somewhat further behind, and PD performing very poorly.

[2] Sonatas K. 279–284, 330–333, 457, 475 and 533, from *Wolfgang Amadeus Mozart: The Complete Piano Sonatas*, played by Roland Batik, Gramola 98701–705, 1990.

	P	R	F	E (ms)
SF	0.958	0.969	0.964±0.017	8.8
PD	0.555	0.868	0.677±0.044	19.5
WPD	0.903	0.921	0.912±0.028	9.6
NWPD	0.945	0.944	0.944±0.021	10.3
CD	0.970	0.962	0.966±0.015	12.8
RCD	0.952	0.958	0.955±0.018	9.3

Table 2: *Results for a database of complex piano music consisting of 106054 onsets: the columns are precision (P), recall (R), F-measure with standard deviation of F-measures across sonatas (F) and average absolute error in ms (E).*

The results for the piano sonatas are overall lower than for the PP data set. This would be expected, since the PP data set contains relatively simple music, and most of the piano music does not even use the damper pedal, so the likelihood of sustained tones masking new onsets is greatly reduced. Considering the difference in complexity of the music, it is surprising that the drop in algorithm performance is not greater; on the contrary, the results are very encouraging.

Another factor that we can take into account with this data is the errors in correctly detected onsets, which are summarised in the right column of Table 2 as an average absolute value. If precision of onset detection is important, the SF function has a slight advantage over other methods. Such a comparison is not possible with the hand-labelled data, since the timing errors in hand-labelling are much greater than the errors we observe in Table 2.

One previous study involved the use of genetic algorithms to learn an optimal set of parameters for combining various simple onset detection functions [12]. Testing was performed on 10 of the same piano sonatas, and the results were equivalent to an average F-measure of 0.94 across different training sets, with an average error of 11 ms. This is slightly worse than the best results achieved here, and the detection window was greater (70 ms instead of 50 ms), which would give higher detection rates. Further, the system had to be trained to learn a larger set of parameters than the one or two parameters which were varied in this work.

4. CONCLUSIONS

We revisited a recent study on onset detection and proposed 3 new onset detection functions and a new peak-picking algorithm as improvements on the published methods. Tests on a common data set supported the claim that the new methods are better, but more extensive tests using a large set of piano music showed the spectral flux and complex domain functions to be marginally better than the weighted phase deviation functions. The test results contradicted some findings in the literature, which probably indicates an instability in the results with respect to small differences in imple-

	PN data			PP data			NP data			CM data		
	P	R	**F**	P	R	**F**	P	R	**F**	P	R	**F**
SF*	0.914	0.871	**0.892**	0.984	0.949	**0.966**	0.945	0.816	**0.876**	0.896	0.804	**0.848**
SF	0.938	0.968	**0.952**	0.981	0.988	**0.984**	0.959	0.975	**0.967**	0.882	0.882	**0.882**
PD*	0.957	0.957	**0.957**	0.997	0.955	**0.976**	0.945	0.807	**0.871**	0.753	0.801	**0.776**
PD	0.654	0.935	**0.770**	0.482	0.865	**0.619**	0.750	0.933	**0.831**	0.663	0.749	**0.704**
WPD	0.937	0.957	**0.947**	0.899	0.925	**0.912**	0.974	0.958	**0.966**	0.843	0.830	**0.836**
NWPD	0.909	0.968	**0.938**	0.961	0.981	**0.971**	0.950	0.966	**0.958**	0.916	0.845	**0.879**
CD	0.946	0.946	**0.946**	0.971	0.984	**0.978**	0.948	0.924	**0.936**	0.941	0.819	**0.876**
RCD	0.948	0.978	**0.963**	0.983	0.979	**0.981**	0.944	0.983	**0.963**	0.945	0.819	**0.877**

Table 1: *Results of onset detection tests, showing precision (P), recall (R) and F-measure (F) for the data sets pitched non-percussive (PN), pitched percussive (PP), non-pitched percussive (NP) and complex mixture (CM), for 8 different onset detection functions (see section 2). The functions marked with asterisks are results in [2].*

mentation details or parameter settings.

A large-scale evaluation was performed on a database of piano music, and it was found that the differences in F-measure between the best algorithms are not significant, implying that the choice of algorithm could be based on other factors such as simplicity of programming, speed of execution and accuracy of correct onsets (all of which speak for SF, the spectral flux onset detection function). The results for the large-scale test were worse than previously published results for pitched percussion instruments, which were somewhat optimistic as they were based on simple data. The present results are also optimistic, since the parameter values were generated using feedback from the ground truth data, and all recordings came from the same instrument and recording conditions.

In future work, we will address the issue of automatic parameter estimation, with the aim of producing a fully automatic onset detection algorithm. (The MIREX 2005 audio onset detection competition was won by a neural network which was trained to detect onsets from spectral data [4].) We also intend to compare these results with another recording of the same data on a different piano with different recording conditions. Finally, an analysis of errors will be performed to determine the extent to which the methods fail at different points, so that by combining the methods (e.g. by voting [13]) a more robust onset detection algorithm could be developed.

5. ACKNOWLEDGEMENTS

Many thanks to Juan Bello of Queen Mary University of London for providing the data for the first set of experiments. This work was supported by the Vienna Science and Technology Fund, project CI010 *Interfaces to Music*, and the EU project S2S². OFAI acknowledges the support of the ministries BMBWK and BMVIT.

6. REFERENCES

[1] J. Vos and R. Rasch, "The perceptual onset of musical tones," *Perception and Psychophysics*, vol. 29, no. 4, pp. 323–335, 1981.

[2] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in musical signals," *IEEE Trans. Speech and Audio Proc.*, vol. 13, no. 5, pp. 1035–1047, 2005.

[3] N. Collins, "A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions," in *118th Conv. Audio Eng. Soc.*, Barcelona, Spain, 2005.

[4] J. Downie, "2005 MIREX contest results – audio onset detection," [Online] http:www.music-ir.org/evaluation/mirex-results/audio-onset, 2005.

[5] J. Bello, C. Duxbury, M. Davies, and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," *IEEE Sig. Proc. Letters*, vol. 11, no. 6, pp. 553–556, 2004.

[6] P. Masri, "Computer modelling of sound for transformation and synthesis of musical signal," Ph.D. dissertation, University of Bristol, UK, 1996.

[7] C. Duxbury, M. Sandler, and M. Davies, "A hybrid approach to musical note onset detection," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02),* Hamburg, Germany, 2002, pp. 33–38.

[8] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, Phoenix, Arizona, 1999.

[9] C. Duxbury, J. Bello, M. Davies, and M. Sandler, "A combined phase and amplitude based approach to onset detection for audio segmentation," in *Proc. 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03)*, 2003, pp. 275–280.

[10] J. Downie, K. West, A. Ehmann, and E. Vincent, "The 2005 music information retrieval evaluation exchange (MIREX 2005): Preliminary overview," in *Proc. Int. Conf. Music Information Retrieval (ISMIR'05),* London, UK, 2005, pp. 320–323.

[11] P. Leveau, L. Daudet, and G. Richard, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," in *Proc. Int. Conf. Music Information Retrieval (ISMIR'04),* Barcelona, Spain, 2004, pp. 72–75.

[12] S. Dixon, "Learning to detect onsets of acoustic piano tones," in *Proc. MOSART Workshop on Current Dir. in Computer Music Res.,* IUA-UPF, Barcelona, 2001, pp. 147–151.

[13] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, and C. Uhle, "An experimental comparison of audio tempo induction algorithms," *IEEE Trans. Speech and Audio Proc.*, vol. 14, no. 5, 2006, to appear.