ONSET TIME ESTIMATION FOR THE EXPONENTIALLY DAMPED SINUSOIDS ANALYSIS OF PERCUSSIVE SOUNDS

Bertrand Scherrer, Philippe Depalle

SPCL/CIRMMT
McGill University
Montréal, QC, Canada
bertrand.scherrer@mail.mcgill.ca

ABSTRACT

Exponentially damped sinusoids (EDS) model-based analysis of sound signals often requires a precise estimation of initial amplitudes and phases of the components found in the sound, on top of a good estimation of their frequencies and damping. This can be of the utmost importance in many applications such as high-quality re-synthesis or identification of structural properties of sound generators (*e.g.* a physical coupling of vibrating devices). Therefore, in those specific applications, an accurate estimation of the onset time is required. In this paper we present a two-step onset time estimation procedure designed for that purpose. It consists of a “rough" estimation using an STFT-based method followed by a time-domain method to “refine" the previous results. Tests carried out on synthetic signals show that it is possible to estimate onset times with errors as small as 0.2ms. These tests also confirm that operating first in the frequency domain and then in the time domain allows to reach a better resolution *vs.* speed compromise than using only one frequency-based or one time-based onset detection method. Finally, experiments on real sounds (plucked strings and actual percussions) illustrate how well this method performs in more realistic situations.

1. INTRODUCTION

In this paper, the focus is set on percussive sounds that can be pitched (*e.g.* guitar, piano or glockenspiel sounds). Such sounds, and sometimes even non-pitched percussive sounds (see [1]), are well represented using a signal model of the form:

$$x[n] = \sum_{m=1}^{M} \left(\sum_{k=1}^{K_m} a_{m,k} e^{j\phi_{m,k}} z_{m,k}^{n} \right) u[n-n_m] + w[n] \quad (1)$$

where $x[n]$ is the real sound signal; M is the number of transient events in the sound; $u[n]$ is the unit step function; n_m is the sample marking the start of transient m; K_m is the order of the model for transient m; $z_{m,k} = e^{(\delta_{m,k}+j\omega_{m,k})}$ is its k^{th} pole with radian frequency $\omega_{m,k}$ and damping factor $\delta_{m,k}$; $a_{m,k}$ and $\phi_{m,k}$ are the initial amplitude and phase of $z_{m,k}$, respectively; and $w[n]$ represents the stochastic component of the signal.

The estimation of the parameters of this type of model has been extensively researched [1, 2, 3, 4]. In the scope of this study, however, it is relevant to note that it is necessary to have a good estimate of the time parameter n_m for those parametric methods to yield the best possible results. For example, when choosing on which segment to perform EDS modelling, it is important that the transient be close to the beginning of the segment to avoid pre-echo artifacts [1]. Also, one can desire a precise knowledge of the “initial” amplitudes and phases of the EDS of the model : for example, in [5], initial amplitudes and phases of the components forming a partial of a guitar’s string sound are central to the estimation of the angle at which a guitar string is released.

In this paper we present an onset detection scheme designed to obtain transients with very fine time resolution in a reasonable amount of time. To borrow the terminology introduced in [6], this method makes use of two different detection functions one after the other.[1] More specifically, the detection functions used are based on frequency- and time-domain energy features rather than on probabilistic models [6] or a combination of the two [8].

The goal of this paper is to show that the sequential application of two simple detection functions leads to significant improvements over the results achievable using these two functions in isolation. Although percussive sounds might be seen as “easy” sounds to detect onsets on, and despite the fact that methods based on variations of the energy of the signal to segment audio have been used for a very long time (*e.g.* [9]), the valuable contribution of this paper lies in that the method proposed is of prime interest in the particular context of the analysis of percussive sounds using exponentially damped sinusoids. Indeed, it allows to obtain a finer time resolution than well known methods such as spectral flux [10] with acceptable computational demands.

The onset detection method is presented in Sec. 2. Experiments on synthetic and real percussive musical sounds are carried out in Sec. 3 and Sec. 4, respectively. The conclusions drawn from these experiments, as well as potential extensions are discussed in Sec. 5.

2. ONSET TIME ESTIMATION PROCEDURE

The onset time estimation procedure studied in this paper is comprised of two steps: a first onset determination over the whole duration of the signal based on its STFT with a “rough” time resolution; the second step involves another onset detection with finer time resolution around each “rough” onset.

The “rough” onset detection starts by computing the STFT of the signal $x[n]$ as follows:

$$X[l,b] = \sum_{n=0}^{N-1} w[n].x[n+lH].e^{j2\pi nb/N} \text{with } b \in [0; N-1] \quad (2)$$

where $w[n]$ is a Hanning analysis window [11], l is the STFT frame index, b is the FFT bin index, N is the FFT size, and H

[1] A similar, though not identical, approach can be found in [7, p. 42].

is the hop size. The frequency-domain detection function $d_f[l]$ is given by:

$$d_f[l] = \sqrt{\sum_{b=0}^{N/2} \big(|X[l,b]| - |X[l-1,b]| \big)^2} \qquad (3)$$

where $|X|$ is the modulus of the complex number X. In essence, $d_f[l]$ measures how different two consecutive STFT frames are from each other using an L_2-norm.[2] It is clear that the maximum time resolution is limited by H and depends on N.

As "rough" onsets are often late (see Sec. 3), the "refining" stage of onset detection is performed on smaller data segments starting a few hop sizes before each "rough" onset. Another detection function is put to use at this stage: a time-domain method, based on the variations of the energy of the signal [1]. More specifically, for a given sample index n, the power of the signal is computed over $\big[x[n-J]; x[n-1]\big]$ (a "backward" window) and over $\big[x[n+1]; x[n+J]\big]$ (a "forward" window), where J is an integer number of samples. The detection function $d_t[n]$ is then computed as follows:

$$d_t[n] = \frac{1}{J} \log \left(\frac{\sum_{m=n+1}^{n+J} x^2[m]}{\sum_{l=n-J}^{n-1} x^2[l] + \upsilon} \right) \cdot \sum_{k=n+1}^{n+J} x^2[k]. \qquad (4)$$

The term in the log function is included in order to emphasize increases in energy. The variable υ is included in Eq. 4 as a regularization factor (*i.e.* to prevent divisions by zero).

The time offsets implied by the definitions of $X[l,b]$, $d_f[n]$ and $d_t[n]$[3] are compensated for in practice in order to be able to perform proper comparisons.

As suggested in [6], the detection function is first zero-meaned, normalized and finally smoothed using a normalized derivative filter:

$$\mathcal{H}(z) = \frac{1-\gamma}{1-\gamma z^{-1}}, \qquad (5)$$

Peaks are identified on the smoothed detection function using parabolic interpolation and considering an extremum to be a peak if it is α dB above the neighbouring minima [12, p. 42]. Once peaks are detected, an adaptive thresholding scheme [6] is used: only the peaks with amplitude higher than τ_{ad} are considered to be onsets. The expression of τ_{ad} is as follows:

$$\tau_{\text{ad}} = \tau + \ell d_{\text{median,p}}, \qquad (6)$$

where τ is an absolute threshold, $d_{\text{median,p}}$ is the normalized and smoothed detection function passed through a median filter of order p, and ℓ controls how much the absolute threshold is affected by $d_{\text{median,p}}$.

After both the "rough" and "refined" onset detection steps, a pruning mechanism is included to remove repeated onsets. That is, each onset is compared to neighbouring onsets within a given time interval (notated I in the rest of the paper). Then, in this interval, only the onset corresponding to the highest value of the detection function is kept.

Table 1: *Parameters of the onset estimation procedure used in Sec. 3.1. Their definition is found in Sec. 2. A sampling rate of 44.1kHz is used and the signals analyzed are such that* $|x[n]| < 1$.

Rough onsets	**Refined onsets**
N : 2048	J : 200
H : 1024	υ : 10^{-4}
γ : 0.3	γ : 0.1
τ : 0.1	τ : 0.5
p : 5	p : 5
ℓ : 0.5	ℓ : 0.5
α : 6dB	α : 6dB
I : 900	I : 900

3. EXPERIMENTS ON SYNTHETIC SOUNDS

In this section, the two-step onset detection procedure presented in Sec. 2 is evaluated in several experiments on synthetic pitched percussive sounds. These sounds reproduce the basic signal structure of sounds generated by a guitar or a piano. In other words, there are several modes, or poles $z_{m,k}$, inside a given string partial due to the coupling of strings through the bridge of the instrument [13, 5]. In this paper, synthetic signals are composed of EDS components grouped in pairs with very similar frequencies and quite different damping factors: 100 different sounds, lasting 1 s each (with $F_s = 44.1$ kHz), are synthesized with parameters randomly chosen within specific ranges as follows. Onset times are chosen within the first 0.5 s of sound segments. The number of partials, K, is such that $K \in [1; 6]$. In order to approach the ideal structure of guitar sounds, partials are chosen to be strictly harmonic, with a fundamental frequency between 82 Hz and 900 Hz. Moreover, amplitudes of harmonics are weighted with a formula of the type $1/k^2$ to mimic the expected spectral slope for the displacement of an ideally plucked string with rigid terminations [14]. Each harmonic consists of two EDS, with slightly different frequencies, damping factors, amplitudes and phases. Only the real part of each signal is kept, so that the phasors in Eq. (1) are replaced by *cos* functions. Finally, some noise is added to the signal in such a way that the ratio of the maximum value of the signal squared to the power of the noise is 30dB. These sounds, as well as the real sounds used in Sec. 4 can be downloaded from this paper's companion webpage.[4]

3.1. Comparing "rough" and "refined" estimations

The two-step onset detection is applied to these sounds with the parameters found in Table 1. The choice of N was motivated by a desire to ensure that spectral components would be separate enough that a rapid temporal variation would clearly translate in energy spreading in more bins. The "refined" onset detection is performed on a portion of the sound starting 5 hop sizes before the onset detected at the "rough" stage and ending 1 hop size after.

Fig. 1 depicts the distribution of errors between the true onset, n_0, and its estimate, $\hat{n}_0$ after both "rough" and "refined" onset estimations, over the 100 sounds of the experiment. From these plots, it is clear that the refining step improves the performance of the onset detection: the median of the error (red line in Fig. 1)

[2] Note that only half of the bins are considered since $x[n] \in \mathbb{R}$.

[3] $N/2$ in Eq. 2, $(N-H)/2$ in Eq. 3 and $J/2$ in Eq. 4

[4] http://www.music.mcgill.ca/~scherrer/dafx14/

goes from 500 samples after the “rough" onset detection to 0 after the “refined” onset detection stage. Also, the spread of the errors for the “refined" estimation is dramatically reduced compared to the “rough" onset stage. In particular, Fig. 2a shows that 75 % of the error lies between 0 and 3 sample of the target (0-0.06ms at 44.1kHz) for the “refined" onset detection compared to the [250;700] sample range (6-16ms) for the “rough" estimation. It also appears that most of the onsets detected at the “rough” stage were late compared to the actual onset time ($e_{n_0} < 0$). This justifies the choice to look for “refined” onsets 5 hop sizes before the estimated onset and 1 hop size after during the refinement stage.

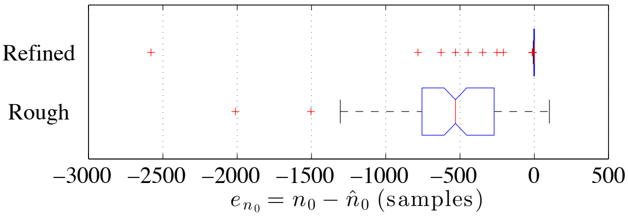


Figure 1: *Distribution of errors made on the estimation of the onset time (n_0) for synthetic signals with randomly chosen parameters. The plot labelled “Rough” corresponds to the distribution of errors made at the “rough” estimation stage, while the plot labelled “Refined” represents the error distribution after refinement.*

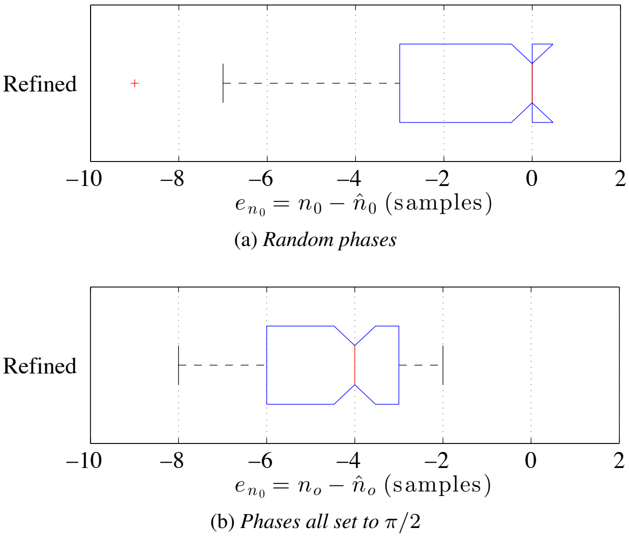
3.2. Testing the robustness to “soft” onsets

After closer inspection of Fig. 1, it appears the outliers (red crosses) in the “refined" stage correspond to signals where the waveform of the signal has a smoother start form 0 compared to the other sounds; signals with “soft" onsets. Thus, another experiment is carried out to better quantify the performance of the method on such sounds.

To that end, another set of synthetic sounds with the same general structure as those studied in Sec.3.1 is generated. The difference lies in the fact that the phases of the EDS’s are now all set to $\pi/2$. The results of this experiment are presented in Fig. 3. The advantage of using the two-step method is still clear, judging from the drastic improvement of the median of the error between “rough” and “refined” steps. When comparing the “refined” onset detection in this experiment and in the previous experiment, as in Fig. 2b, one can note a very slight degradation of performances when all phases are set to $\pi/2$. For example, there are slightly more outliers at the “refined" stage in the case where all the phases are set to $\pi/2$ than when the phases are all random. Also, as shown in Fig. 2b, when all phases are set to $\pi/2$, there is a small increase in the error: a median of -4 instead of 0 for the random phases case, and now 75% of the error is within [-3;-5] samples ([0.07ms-0.11ms] at 44.1kHz). Despite this slight degradation in this adverse scenario, the performance of the method is still very satisfying in terms of time resolution.

3.3. Computational time *vs.* onset time error

This last experiment on synthetic sounds aims at characterizing how the two-pass onset estimation procedure compares to onset estimations using only $d_f[l]$ (*cf.* Eq.3), or only $d_t[n]$ (*cf.* Eq.4). The sounds analyzed are the same as those used in Sec. 3.1. The parameters for the STFT-based method are the same as those in Table 1 for the “rough" estimation, except that the hop size has been varied between 1024 and 256 samples (so between 23ms and 6ms at 44.1kHz). Hop sizes smaller than 256 samples yield computation speeds higher than that of the two-pass estimation procedure so they are not included in the plot. The parameters for the time-based method are identical to those in Table 1.



(a) *Random phases*

(b) *Phases all set to $\pi/2$*

Figure 2: *Distributions of the errors on the detected onset time after refinement, a) for the case of partials with random phases, b) with phases set to $\pi/2$.*

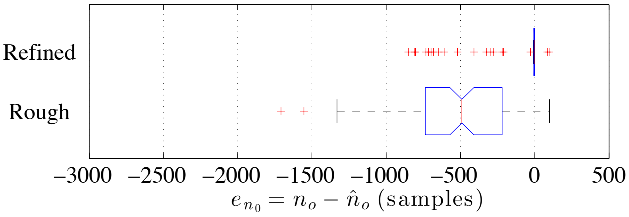


Figure 3: *Distribution of errors made on the estimation of the onset time, in the case where all partials have a phase of $\pi/2$. The labels "Rough" and "Refined" refer to the first and second pass of the onset detection.*

Fig.4 depicts the median value of the computational time[5], t_{comp}, versus the absolute value of the error on the onset,[6] $|e_{n_0}|$, for the different analysis scenarios. The two-pass onset estimation is represented by a white disk, while the time-based method is symbolized by a black triangle. Finally, the several instances of the STFT-based analysis are depicted using grey squares (one for each hop size used).

The ideal method would lie in the leftmost bottom corner of the plot as it would mean that this method is very quick and has no error. With this in mind, it is then clear that the two-pass onset estimation procedure studied here outperforms the time-based method in terms of speed, by a factor of 10. It also performs better

[5] in Matlab R2012b on a MacBook with a 2GHz processor, 4GB of RAM.

[6] As e_{n_0} can be negative.

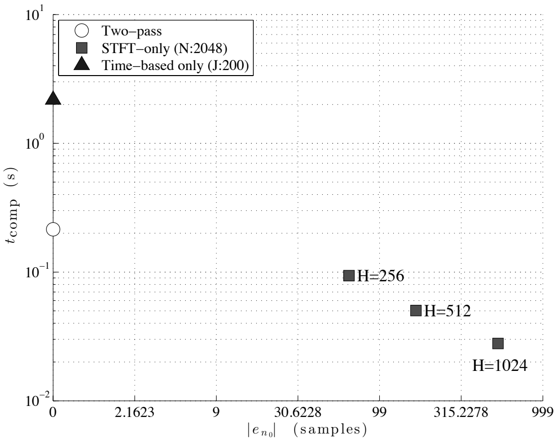


Figure 4: *Median value of t_{comp} the computation time vs. the median value of $|e_{n_0}| = |n_0 - \hat{n}_0|$ for different estimation scenarios: the two-pass procedure (white disk), the time-based method (black triangle), the STFT-based method with varying hop sizes, H (grey squares). Logarithmic scales are used on both axes.*

than the STFT-based method only in terms of error on the onset, since its median is 0 compared to a minimum error of about 50 for $H = 256$. The presence of errors smaller than H is due to the parabolic interpolation performed after smoothing of the detection function $d_f[l]$.

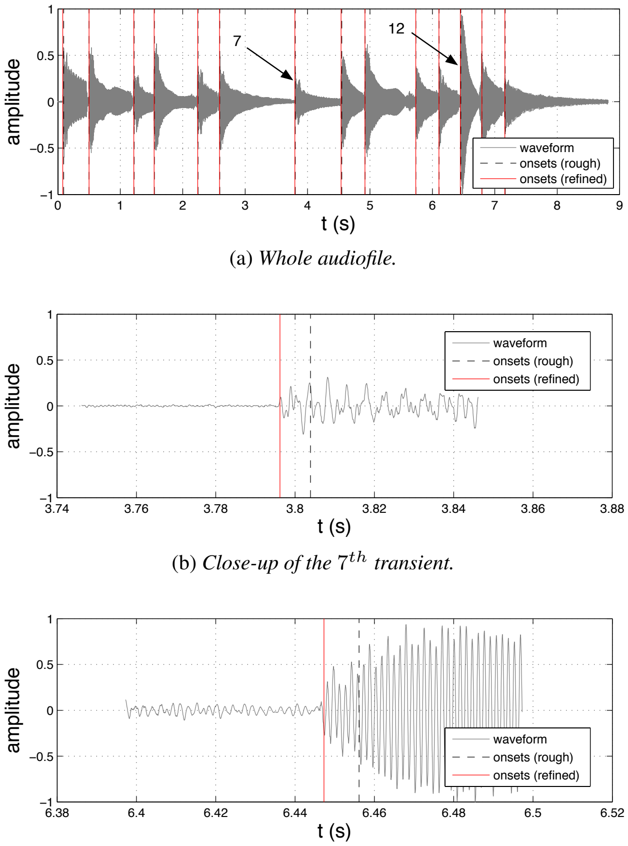
This experiment also confirms the importance of choosing H properly as it seems to significantly reduce the error on the onset estimation: if $H = 256$, the error is about 7 times less than if $H = 1024$ while it only takes about 4 times more to compute. Even for $H = 256$, however, the error is still several orders of magnitude greater than using the two-pass onset estimation procedure.

4. EXPERIMENTS ON MUSICAL SOUNDS

In this section the two-pass onset detection scheme is used on more realistic signals, that do not exactly meet the model of Eq. 1. Sec. 4.1 presents a qualitative experiment that illustrates how the method performs on pitched percussive sounds (a monophonic recording of classical guitar) and on non-pitched percussive sounds (a monophonic recording of castanets). Sec. 4.2 discusses a quantitative evaluation of the method introduced in this paper on a small set of annotated guitar, piano and castanet sounds.

4.1. Qualitative experiment

An example of this two-pass onset detection method applied to a monophonic guitar recording is depicted in Fig. 5. In Fig. 5a, the waveform of the whole audio file is represented and includes both “rough” and “refined” onset times (dashed black lines and solid red lines, respectively). Fig. 5b and Fig. 5c feature close-ups of the 7th and 12th notes in Fig. 5a, respectively. Theses two examples were chosen to show the type of refinement on the “rough" onset detection that this method offers. One may note that, although on the whole signal the refinement may seem minor, the cases shown in Fig. 5b and Fig. 5c represent in fact a substantial improvement especially for analysis tasks requiring precise knowledge of the start of the transient. These two examples also indicate that the two-pass onset detection scheme is relevant for both isolated notes (Fig. 5b) and notes played closer together (Fig. 5c). The parameters used to obtain these results are listed in Table 2.



(a) *Whole audiofile.*

(b) *Close-up of the 7^{th} transient.*

(c) *Close-up of the 12^{th} transient.*

Figure 5: *Whole audiofile with the 7^{th} and 12^{th} transients singled out. The solid red lines are the “refined” onsets, while the dashed black lines are the onsets found at the “rough” onset detection step.* Close-ups on transients indicated by arrows in a) are found in b) and c).

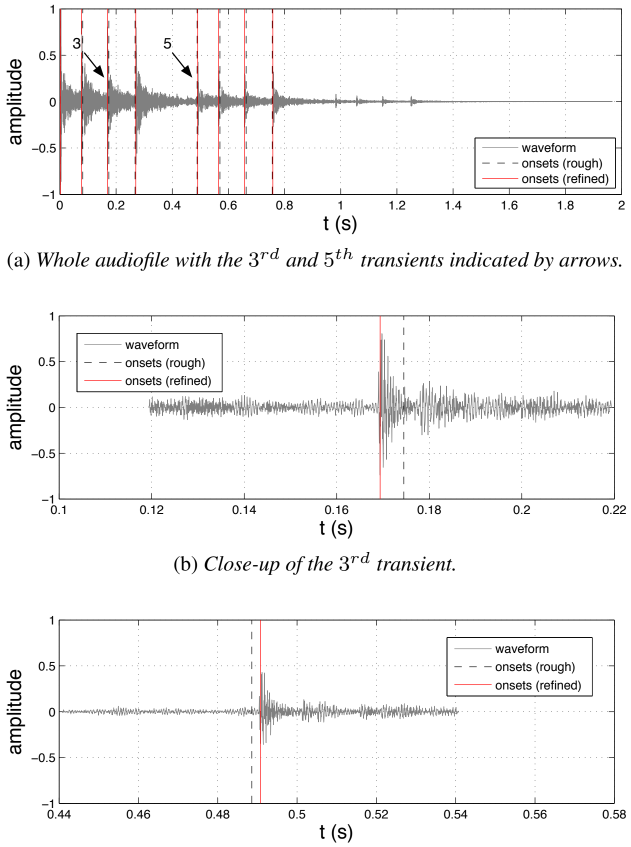
In order to demonstrate that the onset time estimation approach discussed here can also be applied to non-pitched percussive signals another test was conducted. It uses castanet sounds, often chosen as a test case in audio encoding experiments (*e.g.* [1]) as they are quite short with wide energy variations. Fig. 6 depicts both the rough and refined onsets for a castanet recording analyzed using the parameters in Table 3. Fig. 6 depicts two different examples of refinement where the rough onset was late (Fig. 6b) and where the rough onset was slightly too early (Fig. 6c). This shows that the method we propose is also well adapted to signals with very short and sharp transients and could be valuable in audio coding applications based on EDS modeling or other applications requiring very accurate estimations of transient times.

Table 2: *Values of the analysis for the guitar recording in Fig. 5a. The analysis parameters are those presented in Sec. 2 with a sampling rate of 44.1kHz.*

Rough onsets	**Refined onsets**
N : 1024	J : 400
H : 512	$\upsilon : 10^{-4}$
γ : 0.3	γ : 0.1
τ : 0.15	τ : 0.5
p : 5	p : 5
ℓ : 0.5	ℓ : 0.5
α : 6dB	α : 6dB
I : 2205	I : 900

Table 3: *Values of the analysis for the castanet recording in Fig. 6. The analysis parameters are those presented in Sec. 2.*

Rough onsets	**Refined onsets**
N : 1024	J : 400
H : 512	$\upsilon : 10^{-4}$
γ : 0.3	γ : 0.1
τ : 0.1	τ : 0.5
p : 5	p : 5
ℓ : 0.5	ℓ : 0.5
α : 3dB	α : 6dB
I : 2205	I : 900



(a) *Whole audiofile with the 3^{rd} and 5^{th} transients indicated by arrows.*

(b) *Close-up of the 3^{rd} transient.*

(c) *Close-up of the 5^{th} transient.*

Figure 6: *Onset detection on a castanet signal. Solid red lines indicate "refined" onsets, while dashed black lines mark onsets found at the "rough" onset detection step. In a), the results over the whole file are depicted, while b) and c) illustrate how the initial onset time was refined in two particular instances.*

4.2. Quantitative experiment

A preliminary quantitative evaluation has also been carried out to complement the qualitative observations made on the two previous examples. The goal of this experiment is to evaluate the improvement resulting from adding a refining stage. The two-pass onset detection method is thus evaluated on a small set of annotated sounds. This set is comprised of the guitar and castanet sounds previously studied, as well as two non distorted guitar sounds and one piano sound from a small database used for the MIREX2005 Onset Detection task[7]; they are referred to as guitar2, guitar3 and piano1 in the rest of this paper. The annotation of the guitar and castanet sounds was done using the software accompanying [15].

The approach outlined in the instructions of the MIREX Audio Onset Detection task was implemented[8] to perform the evaluation. In particular, the F-measure [16] was used as the main evaluation metric:

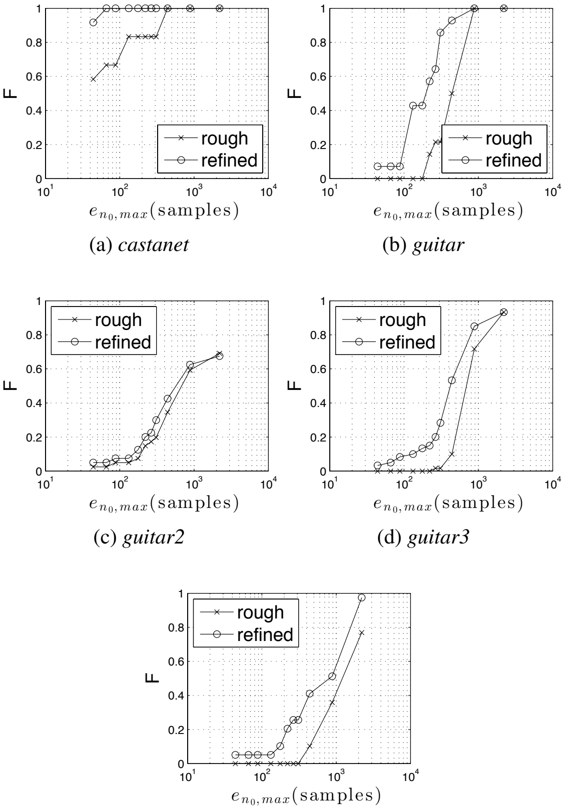
$$\begin{aligned} F &= 2\frac{P.R}{P+R} \qquad (7)\\ \text{with} \quad P &= \frac{n_{TP}}{n_{TP}+n_{FP}} \\ \text{and} \quad R &= \frac{n_{TP}}{n_{TP}+n_{FN}} \end{aligned}$$

where P is the *precision*, R is the *recall* and n_{TP}, n_{FP}, n_{FN} are the numbers of *true positive*, *false positive* and *false negative* detections, respectively. In this paper, for each labelled onset, a *true positive* detection is counted if at least one detected onset is within a certain onset time tolerance, $e_{n_0,max}$, of that onset. When there are no *false positive* or *false negative* detections, $P = 1$ and $R = 1$, so $F = 1$ (see Eq. 7). Conversely, when there are no *true positive* detections, $F = 0$. Thus, for a given sound and a given set of detected onsets, the value of F will change depending on the chosen $e_{n_0,max}$. It is often taken to be equivalent to 50 ms [6] for musical tasks. Since the goal of the method presented in this paper is to provide a fine time resolution for the estimation of EDS parameters, $e_{n_0,max}$ values ranging between 44 samples (≃1 ms at 44.1 kHz) and 2205 samples (50 ms at 44.1 kHz) were used.

Fig. 7 depicts the evolution of the F-measure versus $e_{n_0,max}$, the onset time tolerance. Analysis parameters were found manually for all sounds so that they ensured most (if not all) onsets were found at the rough estimation stage. The tuning of analysis parameters for each sound is appropriate here as this experiment aims at quantifying the effect of the refinement of the onset detection, from the best possible rough onset estimation. The complete list of parameters used can be found on this paper's companion website. There are 5 subplots, one for each recording studied. In

[7] http://www.tsi.telecom-paristech.fr/aao/en/2011/07/13/onset_leveau-a-database-for-onset-detection/

[8] http://www.music-ir.org/mirex/wiki/2014:Audio_Onset_Detection



(a) *castanet*

(b) *guitar*

(c) *guitar2*

(d) *guitar3*

(e) *piano1*

Figure 7: *F-measure versus* $e_{n_0,max}$ *for various annotated sounds.*

each subplot, the F-measure obtained after rough (crosses) and refined estimation (circles) is represented as a function of $e_{n_0,max}$. As expected, the general trend for both rough and refined estimation is that F increases as the tolerance increases. In all plots of Fig. 7, it is clear that adding the refinement stage helps increase F as $e_{n_0,max}$ becomes smaller. The refinement stage does not seem to increase F significantly for a $e_{n_0,max}$ of 882 samples (20 ms) or 2205 samples (50 ms), except for the piano1 recording in Fig. 7e. For all recordings, however, it appears that one benefits most from the refinement stage for tolerances between 176 samples (4 ms) and 441 samples (10 ms).

The results obtained for the castanet recording (Fig. 7a) are much better for both stages of the estimation than all the other sounds. This is most likely due to the fact that this sound was comprised of well defined bursts of energy that do not really overlap in time whereas all the other sounds involve a fair amount of polyphony, where loud notes can overshadow softer new notes. In Fig. 7c, the refinement stage and the rough detections yield approximately similar F values, with a slight advantage to the refined detection. The reason for this is not entirely clear for now but after inspecting the results of the rough onset time estimations for all 5 sounds, it appears that the results obtained for that particular sound are the least satisfying of all five.

5. CONCLUSION

This paper presents and evaluates an onset time estimation procedure specifically designed for applications when a *very* precise onset time estimation is required (of the order of a few tenth of a millisecond). First, a frequency-domain onset estimation is performed. Then around each of those onsets a time-domain onset estimation is used in order to "refine" the onset time estimation.

Experiments on synthetic signals mimicking the structure of guitar sounds show that with this two-pass onset estimation procedure it is possible to obtain onset estimates with errors of at most 0.1ms, 75% of the time. It is also demonstrated that using this two-step method outperforms using either the STFT-based method or the time-based method in isolation. Indeed, Fig.4 shows that the two-pass procedure allows very small error with a small computation time.

Qualitative and quantitative tests on musical recordings help evaluate the performance of the method in more realistic conditions. It is indeed shown that adding the refining stage after the rough estimation helps improve the onset time estimation (increase the F-measure of that detection) when the time tolerance is between 4 ms and 10 ms. The difference with the synthetic case (where onsets were found within 0.1 ms) most likely comes from a combination of factors: real sounds do not conform exactly to the signal model of the synthetic sounds; onsets were manually annotated; the sounds used for testing were polyphonic.

As future work, we plan to investigate the nature of the outliers in the experiment of Sec. 3.2, itself designed to study the outliers of the study in Sec.3.1. Also, as hinted by the results in Fig.4, the choice of parameters of the STFT-based method can have important consequences on the error on the onset and on the computation time. A more systematic evaluation of the effect of N and H on the performances of the STFT-based method would help explain those observations. Another avenue for future work would be to try other methods for the "rough" detection that may be more robust to polyphony. Also, if one were to use this onset detection method on large datasets, and as with all onset time estimation methods, the automatic determination of parameters would be an interesting avenue for improvement.

6. REFERENCES

[1] R. Boyer and K. Abed-Meraim, "Audio modeling based on delayed sinusoids," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 2, pp. 110–120, March 2004.

[2] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automatic Control*, vol. 10, pp. 461–4, 1965.

[3] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT – a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–2, October 1986.

[4] R. Badeau, *Méthodes à Haute Résolution pour l'Estimation et le Suivi de Sinusoïdes Modulées. Application aux Signaux de Musique*, Ph.D. thesis, ENST, Paris, 2005.

[5] B. Scherrer and P. Depalle, "Extracting the angle of release from guitar tones: preliminary results," in *Proc. Acoustics 2012*, Nantes, France, 2012.

[6] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music

signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, September 2005.

[7] N. P. Donaldson, "Extending the phase vocoder with damped sinusoid atomic decomposition of transients," M.S. thesis, McGill University, June 2011.

[8] F. Eyben, S. Böck, and B. Schuller, "Universal onset detection with bidirectional long short-term memory neural networks," in *Proc. 11th ISMIR*, 2010.

[9] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, Upper Saddle River, NJ, 1978.

[10] S. Dixon, "Onset detection revisited," in *Proc. 9th DAFx*, 2006.

[11] Frederic J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–84, January 1978.

[12] X. Serra, *A system for sound analysis / transformation / synthesis based on a deterministic plus stochastic decomposition*, Ph.D. thesis, Stanford University, October 1989.

[13] G. Weinreich, "Coupled piano strings," *J. Acoust. Soc. Am.*, vol. 62, no. 6, pp. 1474–84, 1977.

[14] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*, Springer, 2nd edition, 1998.

[15] Pierre Leveau, Laurent Daudet, and Gaël Richard, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," in *Proc.4th ISMIR*, 2004.

[16] C. J. van Rijsbergen, *Information Retrieval*, Butterworth, London, 1979.