

Abstract

Keywords:

CONTENTS

1	INTRODUCTION	4
2	THEORY	5
2.1	WAVES	5
2.1.1	Continuous and discrete waves	5
2.1.2	Points of interest of periodic waves	6
2.2	LEAST MEAN SQUARES	6
2.3	CONSTRAINED LEAST MEAN SQUARES	7
2.4	SAVITZKY–GOLAY FILTER	7
3	CHARACTERIZATION OF A DIGITAL SIGNAL	8
3.1	ENVELOPE OF A DIGITAL SIGNAL	8
3.1.1	Envelope Detection	8
3.2	FREQUENCY CONTENT	10
3.3	REPRESENTATIONS	10
4	USING GEOMETRIC PROPERTIES OF A SIGNAL TO ESTIMATE THE EN- VELOPE	11
4.1	DISCRETE CURVATURE ESTIMATION - EQUIVALENT CIRCLE APPROACH	15
4.2	RESULTS	17
4.3	ROBUSTNESS	18
4.4	COMPARISON WITH TRADITIONAL ALGORITHMS	18
4.5	RETRIEVING INFORMATION ABOUT THE UNDERLYING WAVE	22
4.6	DISCUSSION	22

4.7	CONCLUSION	23
5	REFINING THE ENVELOPE	23
	GLOSSARY	28

1 INTRODUCTION

Humans are a creative bunch. Since the beginning of computing, even when mainframes were the only available medium, restricted to organizations with considerable budgets, people found ways and motivation to overcome the technical limitations and develop artistic applications, like video games, in their (and the machine's) spare time.

On a parallel note, also relevant to this thesis, those first machines, like the Electronic Numeric Integrator and Calculator (ENIAC) in 1946, brought with them the promise of the artificial intelligence as their ultimate goal [5].

Music, on the other hand, is another staple of humanity's creativity. Is only natural that, as soon as this technology became reasonably stable, musical applications started to emerge.

Ambitious as it may seem, it is the goal of this monograph to motivate a change of approach in domains that, despite their modest intersections, share more attributes than is apparent at a superficial glance: Both machine learning and digital musical acoustics are areas that, despite their current relevance, are fresh out of their respective infancies, and a scrutiny of both promptly reveals the difficulties that arise from this state of affairs. A lack of their own terminology, with frequent borrows from other, more established areas, can be readily cited, as the pronounced prominence of few individual contributions.

Those shortcomings have, however, a bright side to them, in that they still expose those areas to reinterpretation, rendering them somewhat open to the influence of outside ideas.

Digital musical acoustics borrows heavily from the digital signal processing terminology, the last having its roots in the analogic world. Many algorithms are conceived in terms of filters, circuits and other similar legacy constructs, a fact that burdens the conceptual, theoretical framework of the area. Besides the introduction of a nouveau envelope extraction method, we find that his formulation in terms of (differential) geometric concepts is, perhaps, as great a contribution as the algorithm per se.

Perhaps less daring, and owing more to the advanced age of the author than to his research prowess, is the presentation of this monograph in a format more akin to a book; In addition to turning this somewhat dense tome into a more palatable experience.

2 THEORY

2.1 WAVES

It can be insightful to analyse the behaviour of the critical points and the zeroes of an arbitrary, periodic wave over time. We will assume, for the sake of simplicity, that each cycle of said wave has a unique maximum and minimum, and crosses the horizontal axis one single time. Those assumption dont compromise the generality of the analysis, as in the case of multiple maxima, minima and/or zeroes, one of those can be chosen randomly to represent the global one, as long as consistency between cycles is maintained.

We are going to also assume that a wave with a zero phase has an maximum instantaneous amplitude at time zero, as is the case of the cosine wave, for example. Altough that assumption somewhat departs from the physical intuition, it is useful in making our description consistent with the theory behind the Fourier Transform.

For a continuous wave, if we were to plot the time at which any of those values occurred as a function of their ordinality, the result would be parallel straight lines, their slope somehow related to the frequency of the underlying wave. We will illustrated the point in the case of discrete waves, as this case will be of more immediate use in this work.

2.1.1 Continuous and discrete waves

First, it is necessary to establish a correspondence between a continuous and a discrete wave, as the latter will be the object of most of this work. To do that we will define a simple continuous sinusoidal wave of the form $S(t) = a \cos(\phi + 2\pi\mathcal{F}t)$, where $a \in \mathbb{R}^+$ is the amplitude; $\phi \in \mathbb{R}^+, 0 \leq \phi < 2\pi$ is the phase; $\mathcal{F} \in \mathbb{R}^+$ is the absolute frequency of the wave, in Hz and $t \in \mathbb{R}$ is time, in seconds.

The cosine function is used for a twofold reason: its use helps to underline the periodic nature of the wave in consideration and is the parameter of the phase when the FT is used. It is important to keep in mind, however, that this discussion applies any arbitrary function of the form $f(x) = f(x + P) \forall x \in D(f)$, where D is the domain of the function f .

If we were to sample $S(t)$ at regular intervals, that is, to measure the value of the function at a constant rate of h frames per second (FPS), neglecting errors of measurement, we would have a discrete representation of the continuous wave of the form $W[i] = a \cos(\phi + 2\pi f \frac{i}{n})$ where, alongside the quantities introduced before, we would have $i \in \mathbb{N}$ as the frame number; $n \in \mathbb{N}$ as the number of discrete samples and $f \in \mathbb{R}^+$ being the relative frequency of our discrete wave in cycles per n samples.

The relative, local frequency f of the discrete representation is related to the absolute frequency \mathcal{F} via the equation $f = \frac{n}{\text{fps}}\mathcal{F}$ while the absolute time, in the continuous case can be obtained from the frame number i of the discrete signal via $i = t \text{ fps}$.

2.1.2 Points of interest of periodic waves

2.2 LEAST MEAN SQUARES

The linear Least Mean Squares (LMS), also known as linear regression [21] is a data fitting technique.

We present here the theory for real numbers, noting that it can be easily extended for complex numbers [13], should the necessity arise. That, however, will not be the case in this work.

Given two vectors X and Y , such that:

$$X = [x_0, x_1, \dots, x_{m-1}]^T, x_i \in \mathbb{R} \forall i, 0 \leq i < m \in \mathbb{N}$$

$$Y = [y_0, y_1, \dots, y_{m-1}]^T, y_i \in \mathbb{R} \forall i, 0 \leq i < m \in \mathbb{N}$$

We ideally want to find the vector of coefficients $A = [a_0, a_1, \dots, a_k]^T, a_j \in \mathbb{R} \forall j, 0 \leq j \leq k \in \mathbb{N}$ that satisfies the equality $x_i^0 a_0 + x_i^1 a_1 + \dots + x_i^{k-1} a_{k-1} + x_i^k a_k = y_i \quad \forall i \in \{0, \dots, m-1\}$.

Put in that way, what we have is a system of linear equations that can be rewritten in matrix form as seen in equation 1.

$$\begin{bmatrix} Q_{m \times (k+1)} \\ 1 & x_0 & \cdots & x_0^k \\ 1 & x_1 & \cdots & x_1^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m-1} & \cdots & x_{m-1}^k \end{bmatrix} \begin{bmatrix} A_{(k+1) \times 1} \\ a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} Y_{m \times 1} \\ y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{bmatrix} \quad (1)$$

More compactly, we have that $QA = Y$. This system will generally be over-determined, only having an exact solution in the case where Y is a linear combination of the columns in Q [21].

Since an exact solution is often beyond reach, we can at least choose A in a way that makes QA and Y close. A way to do that is to minimize the absolute errors (or their square) between those two quantities; $\min_A \|QA - Y\|_2^2$

Using calculus, for example, one can arrive at the equation $\hat{A} = (Q^T Q)^{-1} Q^T Y$, where \hat{A}

is the vector of coefficients that minimizes the squared errors. A detailed derivation is available at [21].

2.3 CONSTRAINED LEAST MEAN SQUARES

In the constrained version of the LMS, known as the Constrained Least Mean Squares (CLMS) problem we are still interested in finding the coefficient vector \hat{A} that minimizes $\|QA - Y\|_2^2$, but respecting a restriction on this vector of the general form of $VA = B$.

We can then define the Lagrangian and set its derivatives to zero to obtain, as described in Selesnick [19], the closed form solution

$$\hat{A} = (Q^T Q)^{-1} \left(Q^T Y - V^T \left(V (Q^T Q)^{-1} V^T \right)^{-1} V (Q^T Q)^{-1} Q^T Y - B \right)$$

2.4 SAVITZKY–GOLAY FILTER

3 CHARACTERIZATION OF A DIGITAL SIGNAL

3.1 ENVELOPE OF A DIGITAL SIGNAL

Despite being an elusive concept, the temporal amplitude envelope of a signal is essential for its complete characterization, being the primary information-carrying medium in spoken voice and telecommunications, for example. Intuitively, the temporal envelope can be understood as a smooth function that multiplies the signal, being responsible for its outer shape. Envelope detection techniques have applications in areas like health, sound classification and synthesis, seismology and speech recognition. Nevertheless, a general method to digital envelope detection of signals with rich spectral content doesn't exist, as most methods involve manual intervention, in the form of filter design, smoothing, as well as other specific design choices, based on a priori knowledge about the nature of the specific waves under investigation. To address this problem, we propose a framework that uses intrinsic characteristics of a signal to estimate its envelope, completely eliminating the necessity of parameter tuning. The approach here described draws inspiration from geometric concepts to isolate the frontiers and thus estimate the temporal envelope of an arbitrary signal; to that end, alpha-shapes, concave hulls, and discrete curvature are explored. We also define entities, such as a pulse and frontiers, in the context of an arbitrary digital signal, as a means to reduce dimensionality and the complexity of the proposed algorithm. Specifically, a new measure of discrete curvature is used to obtain the average radius of curvature of a discrete wave, serving as a threshold to identify the wave's frontier points. We find the algorithm accurate in the identification of the frontiers of a wide range of digital sound waves with very diverse characteristics, while localizing each pseudo-cycle of the wave in the time domain. The algorithm also compares favourably with classic envelope detection techniques based on filtering and the Hilbert Transform. Besides the most direct applications of this work to audio classification and synthesis, we foresee impact in compression techniques and machine learning approaches to audio. The discrete curvature definition presented could also be extended to three-dimensional settings, to improve shape detection algorithms based on alpha-shapes.

3.1.1 Envelope Detection

Envelope detection is ubiquitous in both analogue and digital signal processing [2]. Nevertheless, the literature in this area is very fragmented [15]. Besides, most envelope detection techniques are designed to account for very specific settings, like pure sinusoids with moderate noise content, a limitation that excludes most physical signals, as is the case of recorded sound,

for example; that limitation arises in part due to the lack of a strict mathematical definition of a temporal envelope [17].

In many natural signals, however, the temporal amplitude envelope of a signal plays a prominent role in the characteristics exhibited: according to Qi et al. [18], for example, the envelope is at least as important as the fine structure of a sound wave in the context of the intelligibility of mandarin tones. This is also the case for English [20], where even envelopes modulating mostly noise were still capable of conveying meaning.

The envelope helps to impart emotion and identity to the human voice [31], and envelope preserving characteristics in concert halls are associated with their pleasantness [14].

When dealing with broadband signals, approaches tailored to specific applications are prevalent, such as the one presented by Yang, Yang, and Yang [30] for the distributed monitoring of fibre optic or the one formulated by Assef et al. [1] in the context of medical ultrasound imaging.

Addressing the different units in the horizontal and vertical axes, one can transform the DSP problem of envelope detection into the geometric problem of defining the shape of a set of points in \mathbb{R}^2 . For that purpose, Edelsbrunner, Kirkpatrick, and Seidel [6] introduced the concept of alpha-shapes, a mathematically well-defined extension to the convex hull of a finite set of points, closely related to the Delaunay triangulation and Voronoi diagrams of those points.

This approach is used in areas such as the detection of features in images [23], reconstruction of surfaces from a cloud of points [26] and Spectroscopy [27], with the last work, which involves the estimation and removal of the Blaze function (a kind of envelope) of an echelle spectrograph, being particularly illustrative of the potential of cross-fertilization.

Steps in the direction of translating geometric algorithms to the context of envelope detection were also made by Yang et al. [29] via an algorithm based on the construction of a skeleton underlying the wave, and also via the direct translation of computer vision methods to the task [28].

Following this path, we present a general approach to envelope detection, exploiting the intrinsic characteristics of a generic, spectrally complex wave, in order to avoid the need for manual intervention or parameter tuning.

While the robustness of the proposed approach allows it to be used as a plug-in replacement for many methods encountered in the literature, we feel that it would be particularly useful for sound synthesis.

The envelope is shown to add complexity to the spectral representation of a wave [22], and an accurate description of the envelope, while describing the evolution of the instantaneous

amplitude of a signal in time, would also greatly simplify further spectral analysis. Moreover, the algorithm developed in this work naturally divides a signal into its pseudo-cycles, pinpointing them in the time domain, providing the building blocks for the reconstruction of the fine structure of the wave.

We then proceed to illustrate the results of the application of the algorithm in a set of six diverse sound waves, ranging from voice to unpitched instruments comparing, afterwards, its performance with that of classic digital envelope estimation algorithms, both from a visual and numerical standpoint.

We then discuss the implications of the work, emphasizing the focus on the construction of new interpretations of the relationship between time domain and frequency domain sound descriptions, and our vision of the impact of this methodology in the field of sound synthesis, providing directions for future developments.

3.2 FREQUENCY CONTENT

3.3 REPRESENTATIONS

4 USING GEOMETRIC PROPERTIES OF A SIGNAL TO ESTIMATE THE ENVELOPE

In this work it is assumed, without loss of generality, that one is interested in retrieving the mathematically ill-defined [17] temporal envelope of an arbitrary physical sound wave, of which only a digital representation is available.

Naturally, due to the necessary truncations both in time and amplitude, as well as measurement errors introduced by the instruments used to record the signal, it is not possible to exactly recover the original, continuous wave, from its digital counterpart.

The best representation one can hope to get of the original wave is, then, the recording itself. It is sensible, thus, to characterize the envelope of the wave in terms of a subset of the exact samples that comprise its digital representation.

To that end, this work will emphasize the concept of frontier, as the points of the digital wave that mark the wave's upper and lower boundaries; from those points an envelope, conforming to the vague definitions of uniqueness and smoothness, can be trivially constructed, as will be shown.

We start by defining a pulse as a series of consecutive samples of a discrete wave that have the same sign. More formally, let $W[i] \in \mathbb{R} \forall i \in \mathbb{N}_0; i < n \in \mathbb{N}_0$ be a real, discrete and finite signal indexed, without loss of generality, over an adjacent subset of the natural numbers starting at zero. This definition relates closely to the concept of an array or vector in most programming languages, and is used in the interest of simplicity.

A pulse $P[j]; \forall j \in \mathbb{N}_0; j < q \in \mathbb{N}_0; q \leq n$ is, then, a subset of W with cardinality $\#P[j] = b - a$ such that $\text{sign}(W[a - 1]) \neq \text{sign}(W[a]) = \text{sign}(W[a + 1]) = \dots = \text{sign}(W[b - 1]) = \text{sign}(W[b]) \neq \text{sign}(W[b + 1]) \forall i; a \leq i \leq b; a, b \in \mathbb{N}_0$.

In other words, each pulse is a unique subset of W with its members being the maximum sequence of samples of W having the same signal. Thus $P[0] \cup P[1] \cup \dots \cup P[q - 1] = W$ and $P[i_1] \cap P[i_2] = \emptyset \forall i_1, i_2 \in \mathbb{N}_0; i_1 \neq i_2; i_1, i_2 < q$.

Naturally, each pulse can be classified as positive ($P[j]^+$) or negative ($P[j]^-$), based on the sign of its samples, which must be the same. It is also convenient to attribute to each pulse a pair of coordinates: those will be the coordinates of the sample with the maximum absolute amplitude in each pulse, and will be denoted by $P[j]_x^\pm$ and $P[j]_y^\pm$. Additionally, the length, or cardinality, of a pulse will be denoted by $\#P[j]^\pm$.

For an arbitrary continuous function, then, a pulse would correspond to the regions between its roots; for a sinusoid, a pulse would be equal to half of its cycle. In analogy, two consec-

utive pulses, naturally with opposing signs, have the potential to delimitate a pseudo-cycle, as illustrated in figure 1. For that to be true, however, the pulses need to make part of the frontiers of the signal, as yet to be defined.

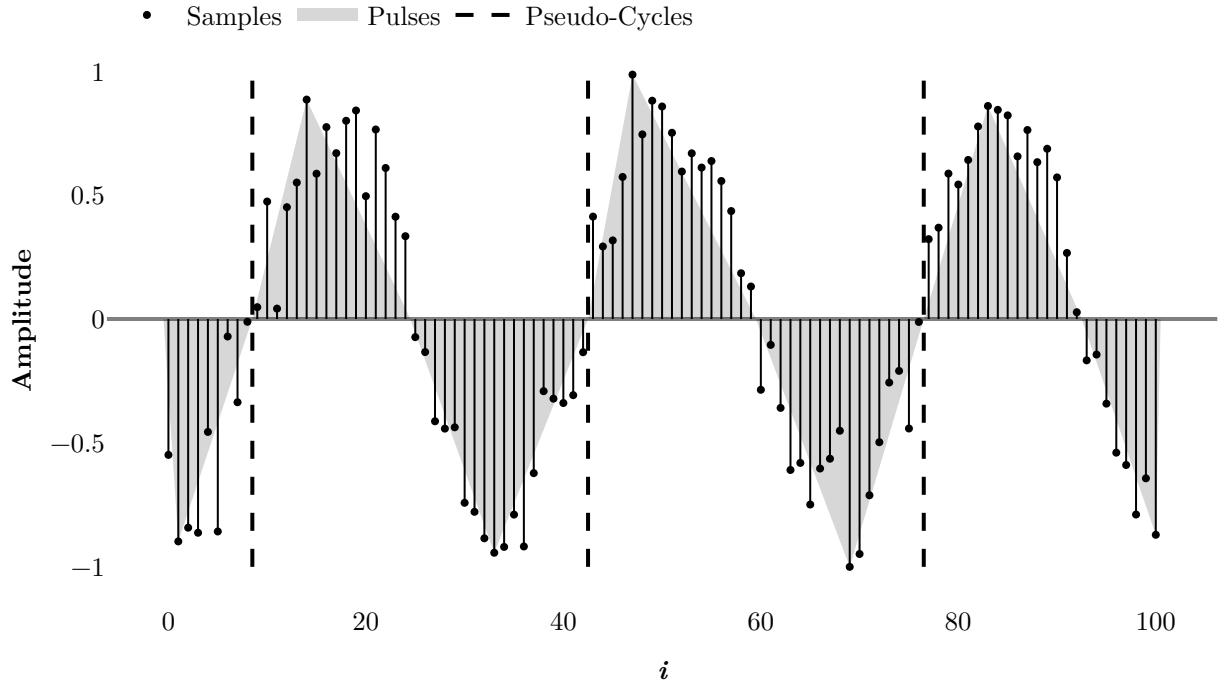


Figure 1: Example of a discrete wave. The black stems represent the individual samples of the signal. The grey areas encompass the simplified convex hulls of the pulses. The dashed lines mark the frontiers between pseudo-cycles.

The most mathematically sound definition of an envelope involves the representation of its underlying wave as an analytic signal, introduced by Gabor in 1946 [11]. In his work, Gabor [9] applies the then relatively new mathematical machinery of the quantum mechanics to unify the time and frequency representations of a wave, showing how the Hilbert transform could be applied to a real signal in order to obtain an equivalent complex signal, later known as the analytic signal.

This analytic signal has the form $a(t) = s(t) + \mathbb{H}(s(t)) i$ [12] where $s(t)$ is the original real signal. $\mathbb{H}(s(t))$, the Hilbert transform of the original signal, becomes the imaginary part of the analytic signal; the envelope of a signal thus represented can be straightforwardly obtained by the computation of its complex modulus.

Figure 2 shows the envelope obtained via the analytic signal, for a pure sinusoid with a local frequency of 20 cycles, modulated by a polynomial of the third degree, illustrating the difference in the shape of the envelope in the presence and absence of gaussian white noise with standard deviation of 1/10 of the wave's maximum amplitude.

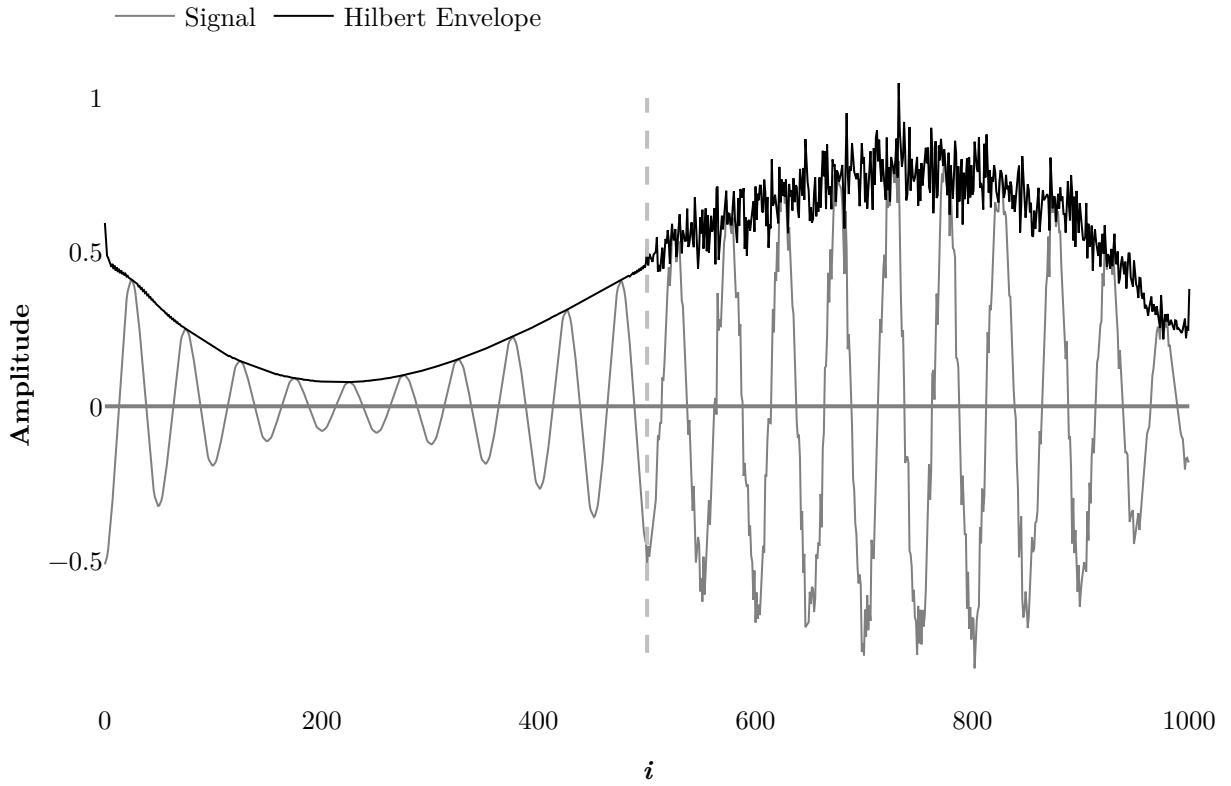


Figure 2: Envelope of a pure sinusoid modulated by a polynomial of degree 3, as obtained by the Hilbert Transform. In the first half, the sinusoid is free of noise, while in the second half white gaussian noise with a standard deviation of $1/10$ of the maximum amplitude of the wave was added to the base signal.

It is readily noticeable from figure 2 that, as soon as noise is introduced in the original signal, it is reflected in the envelope, breaking our assumption of smoothness. Other than that, this definition yields a unique, positive envelope.

In practice, however, is not uncommon for a discrete wave, especially in the case of sound, to present somewhat different positive and negative contours; to account for this discrepancy, we will define a positive (upper) and a negative (lower) frontier of a discrete wave. They will be composed by the points $(i, W[i])$ that are, necessarily but not sufficiently, the points of respectively maximum and minimum amplitude of their pulses, i.e. $(P[j]_x^\pm, P[j]_y^\pm)$.

The set of all these points such that $W[i] > 0$ will be called the positive, piecewise linear frontier and will be denoted by F^+ . Similarly, F^- will denote the negative frontier of W , defined by the subset of points $(i, W[i])$ where $W[i] < 0$, as is shown in figure 3.

To identify such points, we need first to address the different units in which the amplitude and horizontal distance of our wave are expressed. To make abscissa and ordinate units consistent for the computation of the frontiers, and proceed with a geometric treatment of the subject, we consider the set of positive and negative pulses separately.

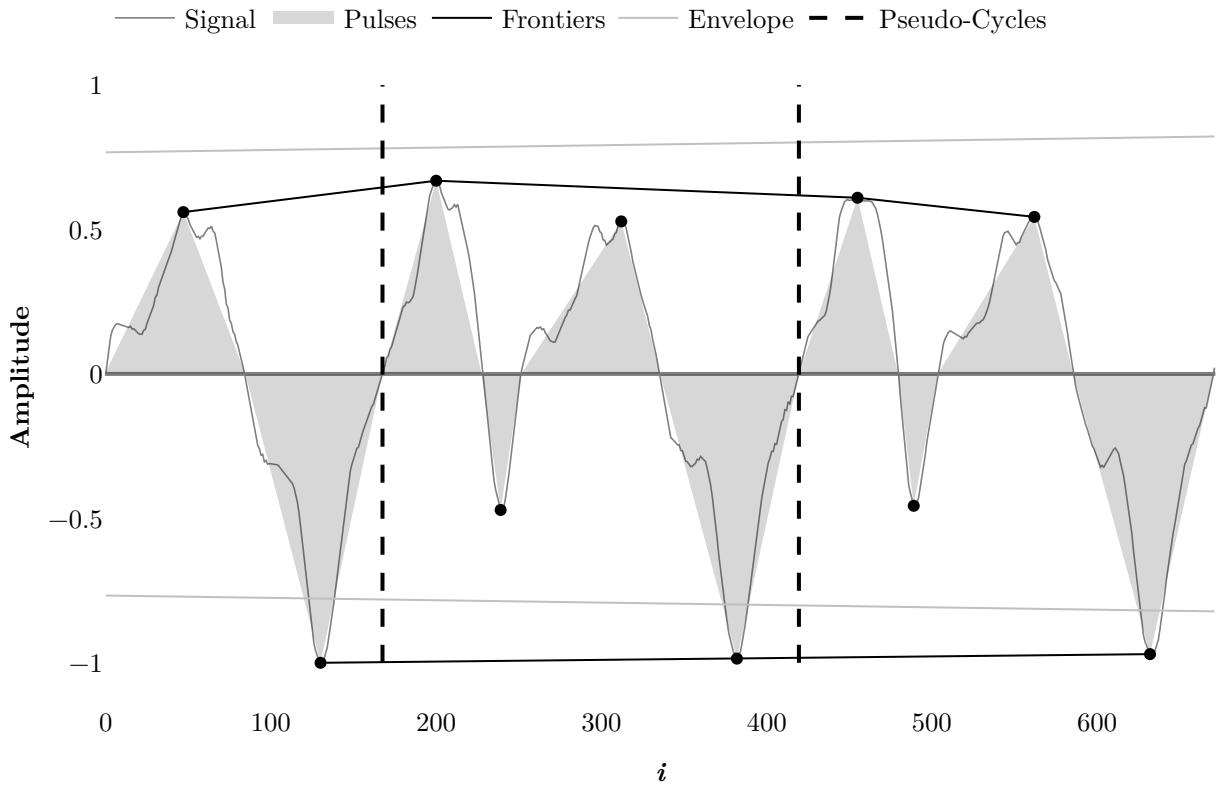


Figure 3: Positive and negative frontiers, envelope and pseudo-cycles for a fragment of piano sound. While we have only one (positive) envelope, here mirrored in the inferior space, we have two frontiers, the superior and inferior, that are obtained independently.

For each set, we multiply all pulses vertical coordinates by the average length of the pulses in the set, after having divided by the average of the pulses' amplitudes, putting both axes in the same unit i , related to time by the equation $i = t \text{ FPS}$.

Formally, supposing we have m pulses of the same sign, we scale each pulse's vertical coordinate $P[j]^\pm$ by s^\pm , where:

$$s^\pm = \frac{\sum_{j=1}^m \#P[j]^\pm}{\sum_{j=1}^m P[j]^\pm} \quad (2)$$

Because we are interested in approximating a function that is smooth over the whole domain of the wave, we can assume that, in the scale of a pulse, this function will be very close to a horizontal straight line and, thus, won't touch any point inside the convex hull defined by the pulse.

By the same token, as the change in amplitude from one pulse to the next pulse with the same sign is motivated by changes in this same envelope, one can reasonably suppose that $\max(P[j]) \approx \max(P[j+2])$.

This makes clear that only one point of an arbitrary pulse can be in the frontier, namely,

its point of maximum amplitude; those definitions will allow us to derive the frontiers of a wave using an alpha-shapes inspired algorithm without the need to compute the Delaunay triangulation first.

That also justifies the approximation of the convex hull of a pulse by the triangle defined by the points where the pulse crosses the horizontal axis and the point of maximum amplitude ($P[j]_x^\pm, P[j]_y^\pm$); this triangular pulse and, more specifically, its point of maximum, will become the atomic unit for the rest of the method.

Translating the intuitive explanation of alpha-shapes in Edelsbrunner and Mücke [7] to our context, we can say that the points in the frontier are those touched by a circle of a given radius r outside the signal that is not allowed to contain any point of the signal; the pulses that contain said points can be seen as frontier pulses.

Intuitively, one can picture a circle being rolled above (or below, in the case of the negative frontier) the signal, and marking the points it touches as frontier points.

We can now revisit the concept of a pseudo-cycle, only hinted at before, defining it as the samples of the original wave that belong to two successive pulses that, in turn, belong to their respective frontiers, including eventual pulses between them, as can be seen in figure 3.

It is necessary to infer the appropriate radius r of such a circle and, to that end, a measure of the curvature of a discrete function is needed. Discrete curvature estimation is an important task in image processing [8] for which no default definition exists.

The two possible general approaches are the derivation of direct methods that use characteristics of the discrete wave to calculate the curvature, or the use of the smooth curvature of a curve fitted to the discrete wave [4].

Concerning the last method, two approaches to fit a polynomial to a generic wave are readily available: the least mean squares approach, that seeks to minimize the dependent variable errors and the total least squares problem, that treats both variables symmetrically [10].

The geometric, symmetrical nature of the problem excludes the more computationally economic least mean squares approach, however, leaving us with the total least squares, for which no general closed-form solutions are available [16]; this leads us to turn our attention to direct approaches of curvature estimation.

4.1 DISCRETE CURVATURE ESTIMATION - EQUIVALENT CIRCLE APPROACH

As previously noted, no default discrete curvature definition exists; Carroll et al. [3], for example, derive three such different definitions based on the approximation of a circle by an inscribed, centred and circumscribed polygon. In the context of 3d meshes, Váša, Kühnert, and

Brunnett [25] evaluate a range of existing estimators from a multivariate point of view.

None of those approaches, especially for being defined over the vertices of a discrete wave, fit seamlessly into our algorithm; we thus proceed to define a discrete curvature measure over the edges of a discrete wave. To that end we are going to apply the definition of smooth curvature as the rate of change of the unit tangent to a curve, noting that this is equivalent to that of the osculating circle [24].

The idea is to find the equivalent circle whose tangent presents the same change in direction, in the same horizontal distance, as the edge of interest, as shown in figure 4.

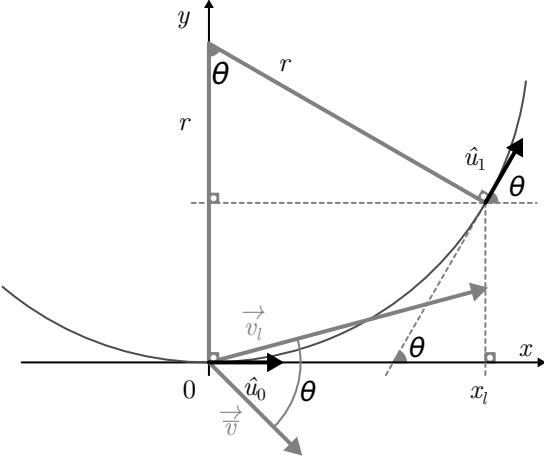


Figure 4: The tangent unit vector of the circle changes from the horizontal direction in \hat{u}_0 to an inclination of θ in \hat{u}_1 . θ is also the angle between the average direction \vec{v}_l and the vector \vec{v}_l that represents the edge of interest, shown in grey.

For each frontier, the angle θ between the average direction and each of the vectors can be obtained via the relation between the slopes d of the respective lines, as in the expression $\theta = \arctan((d - \bar{d})/(1 + d\bar{d}))$, with $\bar{d} = \sum_{l=1}^{m-1} (P[l+1]_y - P[l]_y)/(P[l+1]_x - P[l]_x)$ and $d_l = (P[l+1]_y - P[l]_y)/(P[l+1]_x - P[l]_x)$

The radii of the equivalent circles are thus given by $r_l = x_l / \sin(\theta_l)$ with the curvature given by $\kappa_l = 1/r_l = \sin(\theta_l)/x_l$, where $x_l = P[l+1]_x - P[l]_x$ and $y_l = P[l+1]_y - P[l]_y$. Joining both equations, one can write:

$$r_l = \left| \frac{x_l \sqrt{\bar{d}^2 + 1} \sqrt{x_l^2 + y_l^2}}{\bar{d} x_l - y_l} \right| \quad (3)$$

After this procedure, illustrated in figure 5, two sets of points in \mathbb{R}^2 , corresponding to the positive and negative frontiers, are obtained, where the ordinates represent the instantaneous radius, with its localization in the time domain given by the abscissas.

Given the robustness of the method, for an average wave of a few seconds at an ordinary

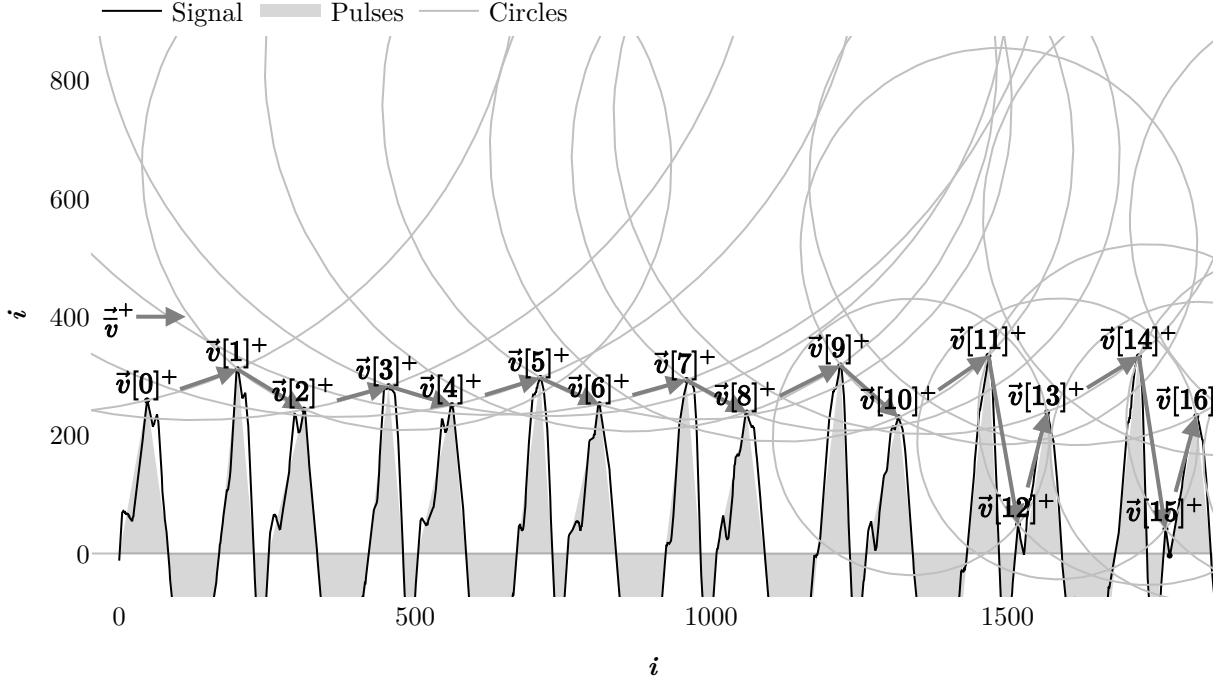


Figure 5: The equivalent circles obtained via the procedure described for the positive frontier. θ , the angle between \vec{v}^+ and each of the \vec{v}_l^+ is used, as is the horizontal component of each \vec{v}_l^+ to obtain the radii of the circles.

rate of 44100 FPS the average of the radius values is sufficient to assure the accurate identification of the frontier. However, a curve fitting method can be used in the case of longer or ill-behaved waves. An example of the use of the average equivalent circle can be seen in figure 6.

4.2 RESULTS

Figure 7 illustrates the frontiers of six diverse discrete sound waves, as well as an in-detail view of the highlighted segment for each signal. All waves are records of physical sounds, chosen to represent the applicability of the algorithm in real-world scenarios.

We can see that the frontier is satisfactorily detected in the case of harmonic and inharmonic sounds, and is robust in relation to the number of samples and the frequencies of the waves.

The impact of the normalization of a wave using the frontiers obtained can be seen in figure 8, which shows the original and normalized wave in the case of a recording of the key 33 of a piano both in the time and frequency domains. We can see that the Fourier power spectra is greatly simplified in the case of the normalized wave.

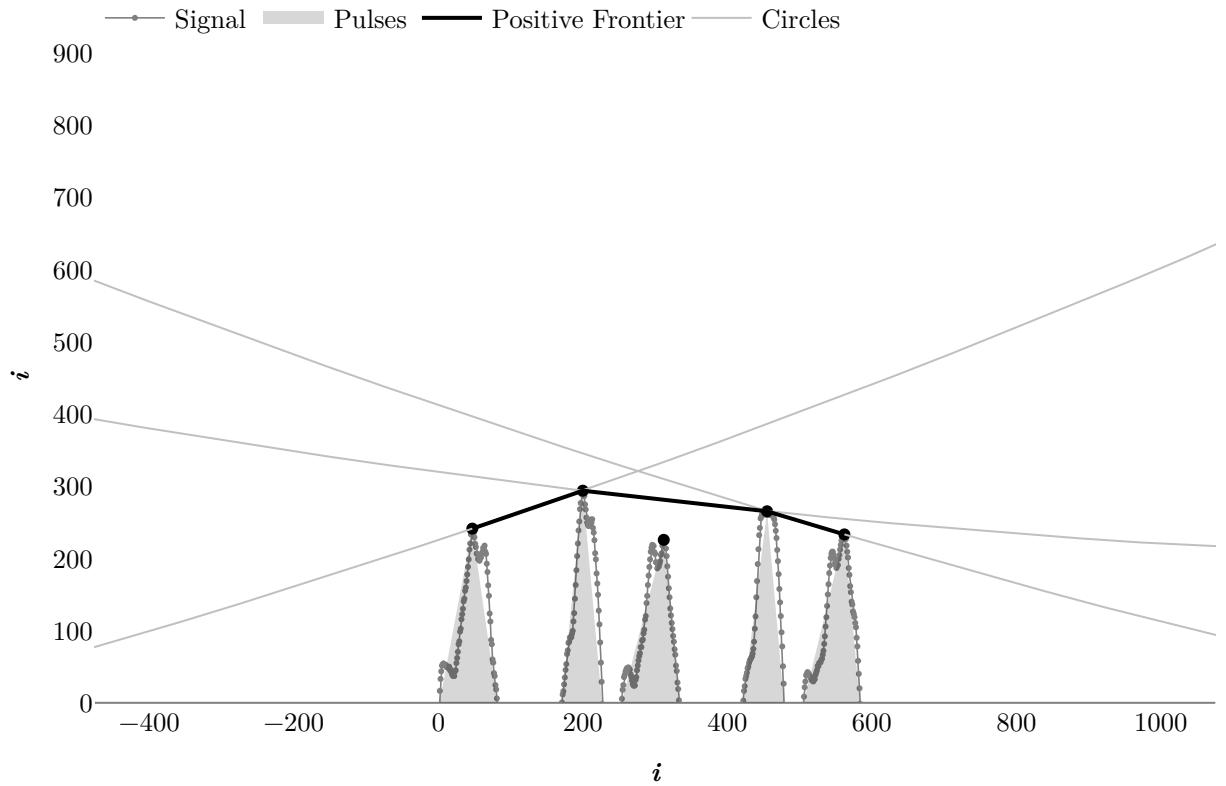


Figure 6: The positive frontier as obtained by the use of the circle representing the average curvature of the discrete wave illustrated.

4.3 ROBUSTNESS

Figures 9 and 10 illustrate the computational resilience of the algorithm for scaling in both the abscissa and ordinate axes, respectively. This is due to the normalization scheme presented in equation 2.

Although important from a theoretical standpoint, this step has an especial computational relevance in the case of digital sound waves, as they tend to be composed of a high number of samples with relatively low amplitudes, generally normalized between -1 and 1, leading to a numerical disparity between axes.

4.4 COMPARISON WITH TRADITIONAL ALGORITHMS

Direct comparison with many of the more recent algorithms is made difficult by the unavailability of digital implementations of such works, many designed to process analogue signals [e.g., 1].

Nevertheless, insight can be gained from a comparison of the results of the method here proposed with some of the most common envelope extraction algorithms, as can be seen in figure 11.

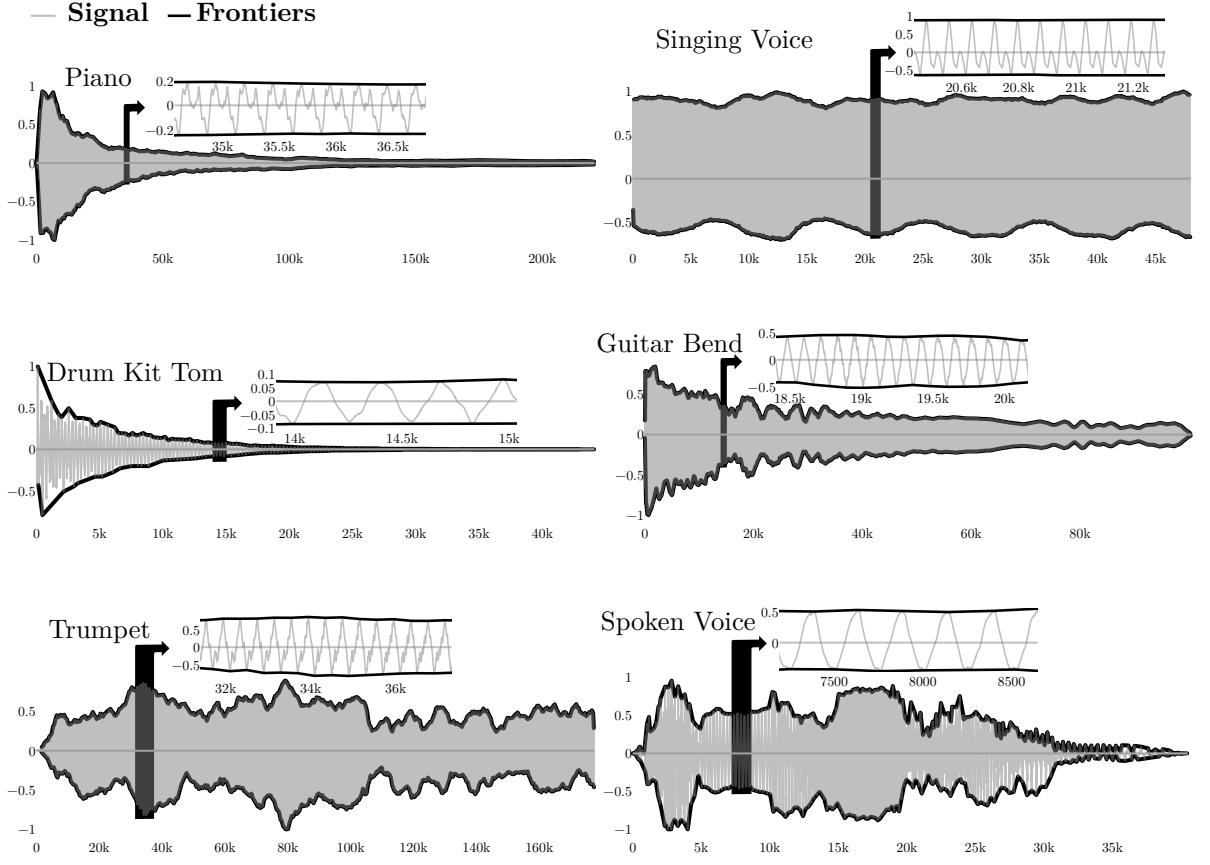


Figure 7: Positive and negative frontiers of six digital waves, as extracted by the algorithm here presented. For each wave, the region highlighted in black is shown in detail besides the whole wave. For each wave the horizontal axis is the sample number i , while the vertical axis is the normalized amplitude. All 6 waves were recorded at 44100 FPS.

For a recording of a guitar sound, in which a bend is performed, we contrast the proposed algorithm with four common approaches of varying complexity: simple smoothing of the rectified original wave, low pass filtering followed by rectification, peak identification and subsequent smoothing, and the extraction of the envelope via Hilbert transform, as presented earlier in the paper, with a previous smoothing of the original wave to avoid the artifacts shown in figure 2.

All those methods demanded careful choice of parameters: in the case of filtering, a cut-off frequency of approximately 88 Hz was used, consisting of one additional parameter to be tuned. For all the other traditional approaches, the Savitzky-Golay smoothing algorithm was used, and the window size was the parameter to be defined.

In the case of the proposed algorithm, the smooth temporal envelope was obtained merging both positive and negative frontiers, the result being smoothed with the same procedure described. Figure 12 illustrates the results.

Table 1 presents a computational comparison between the techniques. The times are for

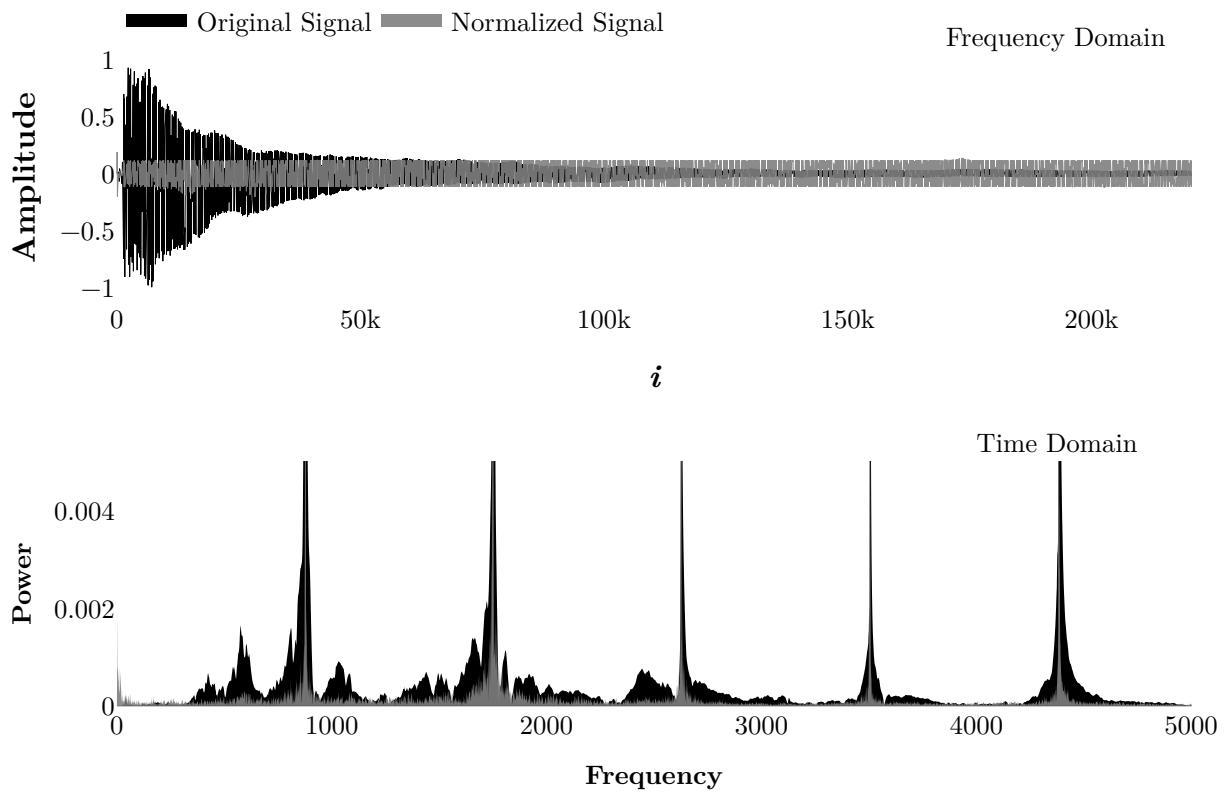


Figure 8: Original and normalized views of a recording of key 33 (F3, 175 Hz) from a grand piano both in time and frequency domains.

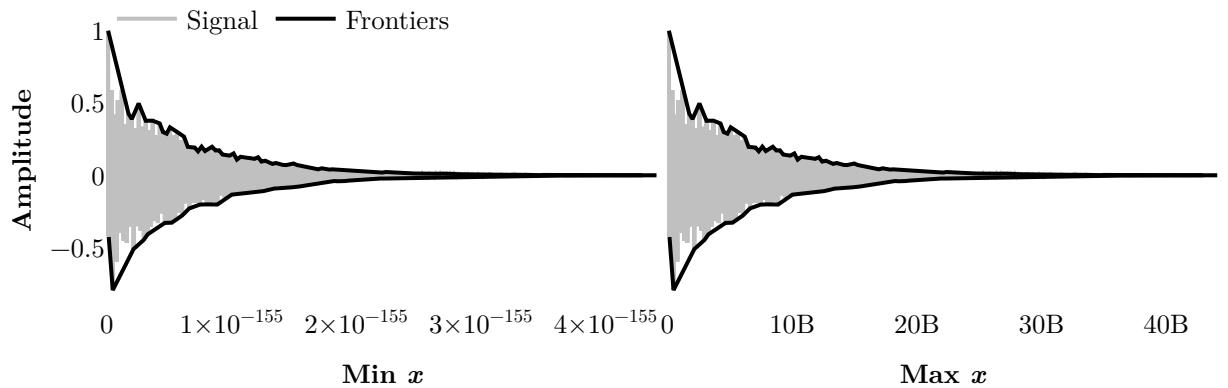


Figure 9: Minimum and maximum scales of the abscissa before computational stability issues arose, for a recording of a drum kit tom. $B = 10^6$.

custom implementations of those methods in the Python programming language. The least mean squares result was calculated by halving the envelopes presented in figure 11 and comparing them with the rectified absolute wave, so as to avoid the potential influence of an estimated frequency.

We can see that, while the algorithm here presented is the most accurate, it's also the most computationally intensive. The connection between alpha-shapes and Delaunay triangulation can be explored in future works as a means to improve the efficiency of the algorithm.

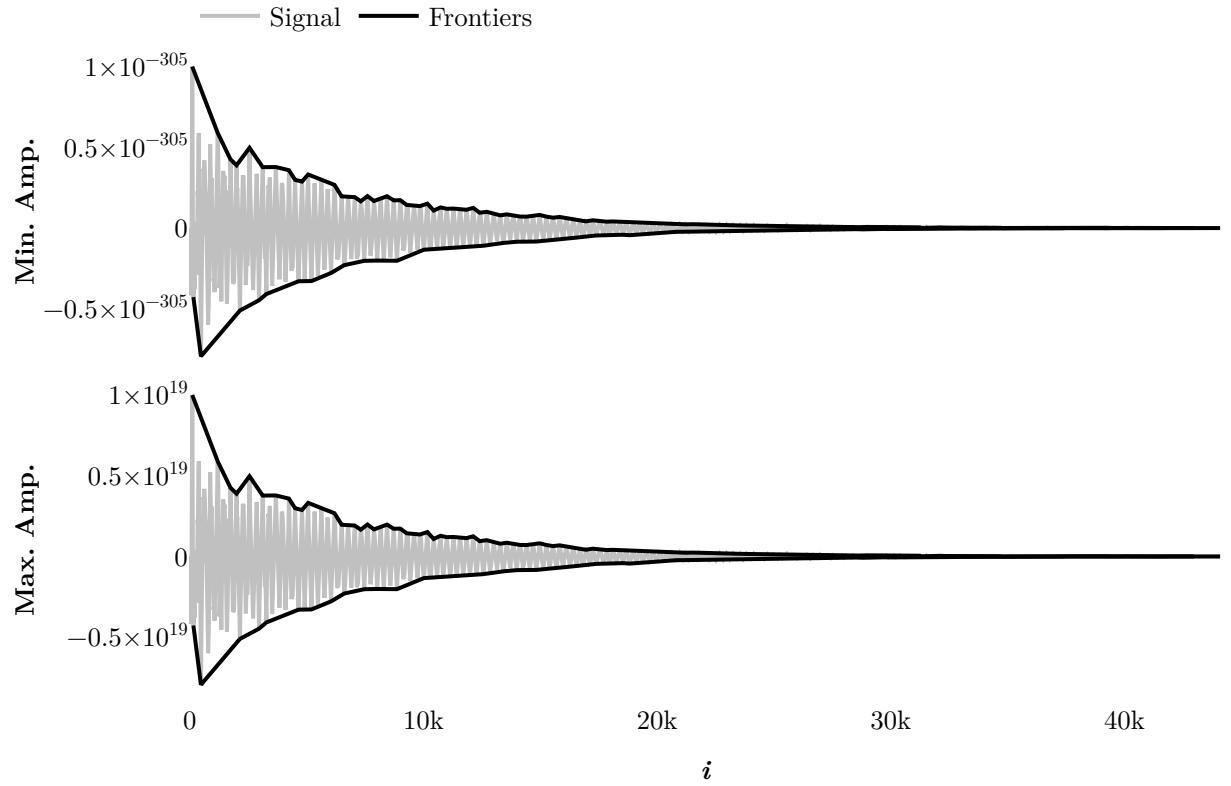


Figure 10: Minimum and maximum values of the amplitude before computational stability issues arose, for a recording of a drum kit tom.

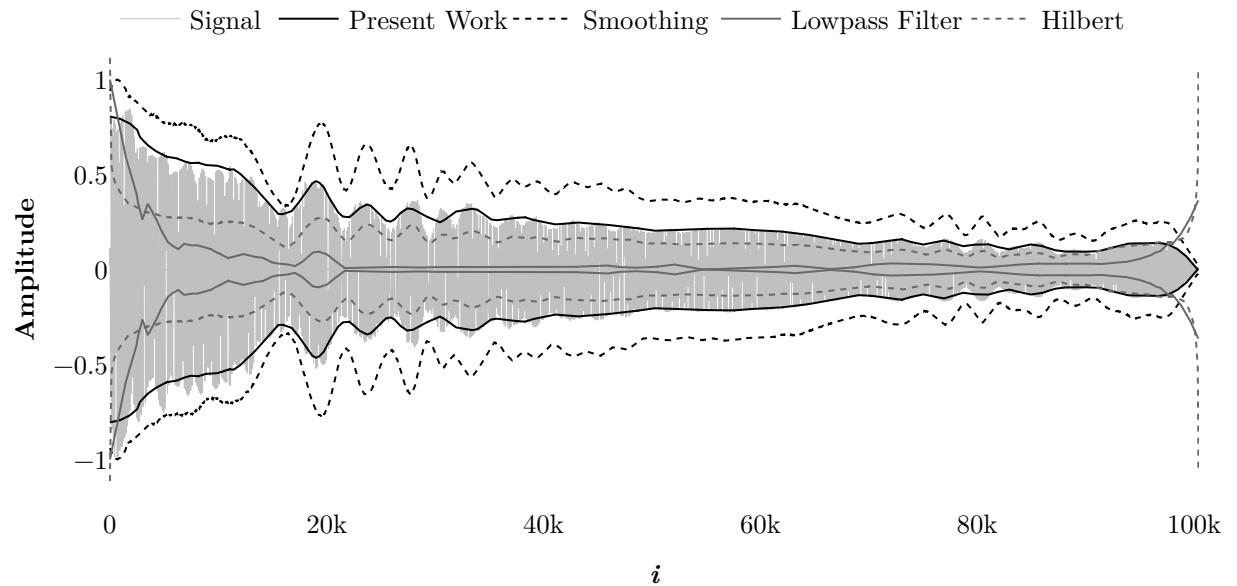


Figure 11: Comparison of the algorithm present in this work with the most common methods of digital envelope identification.

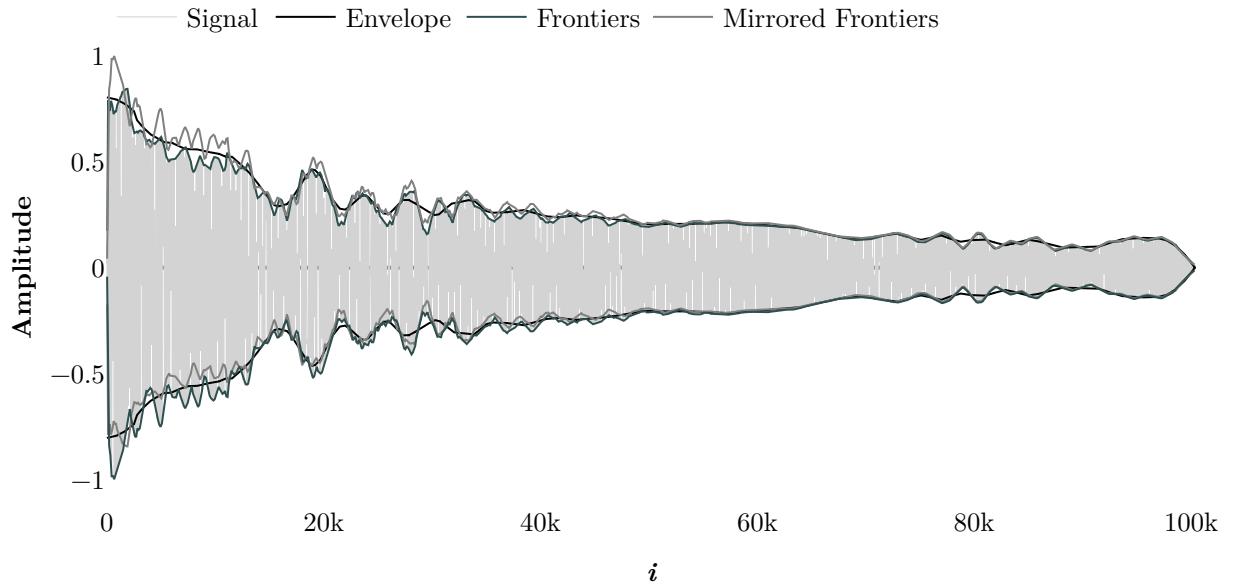


Figure 12: Original positive and negative frontiers in dark grey and their mirrored versions. The temporal envelope, in black, is obtained by merging and approximating both frontiers.

Method	LMS	time(s)
Presented Algorithm	0.0088	4.1507
Smoothing	0.0123	0.1911
Low pass Filter	0.0258	0.5601
Hilbert Transform	0.0141	0.2599

Table 1: Comparison of the algorithm present in this work with the most common methods of digital envelope identification.

4.5 RETRIEVING INFORMATION ABOUT THE UNDERLYING WAVE

The algorithm provides the position of the pseudo-cycles of a wave, with which one can, for example, derive its fundamental frequency and phase: using one of the two frontiers, we have $f = (\#F^\pm - 1)n/(F^\pm[-1] - F^\pm[0])$, where $\#F^\pm$ denotes the cardinality of the frontier, while $F^\pm[0]$ and $F^\pm[-1]$ stand for the frontier's first and last item, respectively.

Similarly, each point of each of the frontiers can be used to estimate the local phase at that location. For the positive frontier, it is necessary to note that, for each point, the following equation holds: $\cos(\phi + 2\pi fi/n) = 1$, and thus $\phi = 2\pi k - 2\pi fi/n$, with $k \in \mathbb{N}$, while for the negative frontier $\cos(\phi + 2\pi fi/n) = -1$ and $\phi = 2\pi k + \pi - 2\pi fi/n$.

4.6 DISCUSSION

This work hopes to fill a gap identified by the authors in the context of sound synthesis, where the lack of procedures for the accurate identification of the temporal envelope of arbitrary

waves consists of an obstacle to the complete description and eventual manipulation of signals.

While relevant in its own accord, the procedure here presented isolates the individual pseudo-cycles of a wave, pinpointing them in the time domain. Due to that, a rich set of theoretical advancements, of whose an example was given in the end of the preceding section, can be built upon this initial algorithm.

With the knowledge of the exact position and characteristics of the pseudo-cycles of a wave, one can investigate its evolution in time, in terms of waveform, for example, free from the potential interference of a modulating wave; in other words, one can have a very satisfactory view of the instantaneous shape and phase of an arbitrary wave and its evolution in time, providing a way to bridge the gap between time domain and frequency domain interpretations.

In the context of sound synthesis, for example, one would be able to apply specific methods for the recreation of the envelope, more in line with its smooth, slow varying and non-periodic nature and a different approach to the periodic and relatively fast changes of the temporal fine structure.

4.7 CONCLUSION

It is our opinion that such theoretical advances can be used in the formulation of better representations of discrete sounds, that in turn can improve sound compression techniques, while providing more adequate representation for the application of machine learning techniques for both sound classification and synthesis.

5 REFINING THE ENVELOPE

REFERENCES

- [1] Amauri Amorin Assef et al. “Modeling and FPGA-based implementation of an efficient and simple envelope detector using a Hilbert Transform FIR filter for ultrasound imaging applications”. In: *Research on Biomedical Engineering* 34.1 (Jan. 2018), pp. 87–92. doi: 10.1590/2446-4740.02417.
- [2] Marcelo Caetano and Xavier Rodet. “Improved estimation of the amplitude envelope of time-domain signals using true envelope cepstral smoothing”. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2011. doi: 10.1109/icassp.2011.5947290.
- [3] Daniel Carroll et al. “A Survey of the Differential Geometry of Discrete Curves”. In: *The Mathematical Intelligencer* 36.4 (Oct. 2014), pp. 28–35. doi: 10.1007/s00283-014-9472-2.
- [4] David Coeurjolly, Serge Miguet, and Laure Tougne. “Discrete Curvature Based on Osculating Circle Estimation”. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 303–312. doi: 10.1007/3-540-45129-3_27.
- [5] Tristan Donovan. *Replay: the history of video games*. East Sussex, England: Yellow Ant, 2010. ISBN: 9780956507228.
- [6] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. “On the shape of a set of points in the plane”. In: *IEEE Transactions on Information Theory* 29.4 (July 1983), pp. 551–559. doi: 10.1109/tit.1983.1056714.
- [7] Herbert Edelsbrunner and Ernst P. Mücke. “Three-dimensional alpha shapes”. In: *ACM Transactions on Graphics (TOG)* 13.1 (Jan. 1994), pp. 43–72. doi: 10.1145/174462.156635.
- [8] Oliver Fleischmann, Lennart Wietzke, and Gerald Sommer. “A Novel Curvature Estimator for Digital Curves and Images”. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 442–451. doi: 10.1007/978-3-642-15986-2_45.
- [9] D. Gabor. “Theory of communication. Part 1: The analysis of information”. In: *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering* 93.26 (Nov. 1946), pp. 429–441. doi: 10.1049/ji-3-2.1946.0074.

- [10] Gene H. Golub and Charles F. van Loan. “An Analysis of the Total Least Squares Problem”. In: *SIAM Journal on Numerical Analysis* 17.6 (Dec. 1980), pp. 883–893. doi: 10.1137/0717073.
- [11] Stefan L. Hahn. “The History of Applications of Analytic Signals in Electrical and Radio Engineering”. In: *EUROCON 2007 - The International Conference on "Computer as a Tool"*. IEEE, 2007. doi: 10.1109/eurcon.2007.4400463.
- [12] Lei He and Volker Dellwo. “A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform”. In: *Interspeech 2016*. ISCA, Sept. 2016. doi: 10.21437/interspeech.2016-1447.
- [13] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, Jan. 1995. doi: 10.1137/1.9781611971217.
- [14] Tapani Lokki et al. “Engaging concert hall acoustics is made up of temporal envelope preserving reflections”. In: *The Journal of the Acoustical Society of America* 129.6 (June 2011), EL223–EL228. doi: 10.1121/1.3579145.
- [15] Richard Lyons. “Digital Envelope Detection: The Good, the Bad, and the Ugly [Tips and Tricks]”. In: *IEEE Signal Processing Magazine* 34.4 (July 2017), pp. 183–187. doi: 10.1109/msp.2017.2690438.
- [16] Ivan Markovsky and Sabine Van Huffel. “Overview of total least-squares methods”. In: *Signal Processing* 87.10 (Oct. 2007), pp. 2283–2302. doi: 10.1016/j.sigpro.2007.04.004.
- [17] Qinglin Meng et al. “An empirical envelope estimation algorithm”. In: *2013 6th International Congress on Image and Signal Processing (CISP)*. IEEE, Dec. 2013. doi: 10.1109/cisp.2013.6745226.
- [18] Beier Qi et al. “Relative contributions of acoustic temporal fine structure and envelope cues for lexical tone perception in noise”. In: *The Journal of the Acoustical Society of America* 141.5 (May 2017), pp. 3022–3029. doi: 10.1121/1.4982247.
- [19] Ivan Selesnick. *Least Squares with Examples in Signal Processing*. OpenStax CNX. Apr. 17, 2013. URL: <http://cnx.org/contents/5d13ca71-5821-41d6-a40a-d5278652fc11@1.>
- [20] R. V. Shannon et al. “Speech Recognition with Primarily Temporal Cues”. In: *Science* 270.5234 (Oct. 1995), pp. 303–304. doi: 10.1126/science.270.5234.303.

- [21] Lieven Vandenberghe Stephen Boyd. *Introduction to Applied Linear Algebra*. Cambridge University Pr., June 7, 2018. 500 pp. ISBN: 1316518965. URL: https://www.ebook.de/de/product/32105301/stephen_boyd_lieven_vandenberghe_introduction_to_applied_linear_algebra.html.
- [22] Carlos Tarjano and Valdecy Pereira. “Neuro-Spectral Audio Synthesis: Exploiting Characteristics of the Discrete Fourier Transform in the Real-Time Simulation of Musical Instruments Using Parallel Neural Networks”. In: *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*. Springer International Publishing, 2019, pp. 362–375. doi: 10.1007/978-3-030-30490-4_30.
- [23] Christos Varytimidis et al. “Alpha-shapes for local feature detection”. In: *Pattern Recognition* 50 (Feb. 2016), pp. 56–73. doi: 10.1016/j.patcog.2015.08.016.
- [24] L. Váša et al. “Mesh Statistics for Robust Curvature Estimation”. In: *Computer Graphics Forum* 35.5 (Aug. 2016), pp. 271–280. doi: 10.1111/cgf.12982.
- [25] Libor Váša, Tom Kühnert, and Guido Brunnett. “Multivariate analysis of curvature estimators”. In: *Computer-Aided Design and Applications* 14.1 (June 2016), pp. 58–69. doi: 10.1080/16864360.2016.1199756.
- [26] Bin Wu et al. “Automated extraction of ground surface along urban roads from mobile laser scanning point clouds”. In: *Remote Sensing Letters* 7.2 (Dec. 2015), pp. 170–179. doi: 10.1080/2150704x.2015.1117156.
- [27] Xin Xu et al. “Modeling the Echelle Spectra Continuum with Alpha Shapes and Local Regression Fitting”. In: *The Astronomical Journal* 157.6 (May 2019), p. 243. doi: 10.3847/1538-3881/ab1b47.
- [28] Cong Yang, Marcin Grzegorzek, and Ewa Lukasik. “Representing the evolving temporal envelope of musical instruments sounds using Computer Vision methods”. In: *2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE, Sept. 2015. doi: 10.1109/spa.2015.7365137.
- [29] Cong Yang et al. “Skeleton-based audio envelope shape analysis”. In: *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, Nov. 2015. doi: 10.1109/acpr.2015.7486556.

- [30] Wei Yang, Yuanhong Yang, and Mingwei Yang. “Fast digital envelope detector based on generalized harmonic wavelet transform for BOTDR performance improvement”. In: *Measurement Science and Technology* 25.6 (Apr. 2014), p. 065103. doi: 10.1088/0957-0233/25/6/065103.
- [31] Zhi Zhu et al. “Contributions of temporal cue on the perception of speaker individuality and vocal emotion for noise-vocoded speech”. In: *Acoustical Science and Technology* 39.3 (May 2018), pp. 234–242. doi: 10.1250/ast.39.234.

GLOSSARY

CLMS Constrained Least Mean Squares. 7

ENIAC Electronic Numeric Integrator and Calculator. 4

FPS frames per second. 5, 17, 19

LMS Least Mean Squares. 6, 7