Robust Digital Envelope Estimation Via Geometric Properties of an Arbitrary Real Signal

Carlos Tarjano   Valdecy Pereira

September 2, 2020

Abstract

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 modulates 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. Despite of that, a general approach to envelope detection of signals with rich spectral content seems to be lacking, and most methods involve manual intervention, like filter design or smoothing, based on prior knowledge about the wave under investigation. In this paper 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 general, computational and differential geometry to isolate the frontiers and thus estimate the temporal envelope of an arbitrary signal; to that end The concepts of alpha-shapes, concave hulls, and discrete curvature are explored. We also define a pulse in the context of a arbitrary digital signal as a means to reduce dimensionality and the complexity of the proposed algorithm. We find that the algorithm accurate in the identification of the frontiers of a range of digital sound signal with very diverse characteristics, while providing the dominant frequency of the wave as a by-product. The algorithm also compares favourably with a common detection technique based on the Hilbert Transform. While the results are satisfactory per se, we define several geometric entities in this work, including a new measure of discrete curvature; we hope to consolidate theory in the area, while providing a solid basis for the development of new techniques, specially in sound synthesis.

Keywords: DSP, alpha-shapes, envelope detection

1   Introduction

Despite being ubiquitous in both analog and digital signal processing, the literature about envelope detection is very fragmented (Lyons, 2017). Besides, most envelope detection techniques are designed to account for very specific settings, like pure sinusoids with moderate noise contents, in line with the most common usages of those algorithms with artificial signals in the context of analog telecommunications; that limitation arises in part by the lack of a mathematical definition for a temporal envelope (Meng et al., 2013).

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. (2017), for example, the envelope is at least as important as the fine structure of the soundwave in the context of the intelligibility of mandarin tones. That is also the case for English (Shannon et al., 1995), where even envelopes modulating mostly noise were still capable of conveying meaning. The envelope helps to convey emotion and identity to the human voice (Zhu et al., 2018), and envelope preserving characteristics of concert halls are associated with their pleasantness (Lokki et al., 2011).

When dealing with broadband signals, approaches tailored to specific applications are prevalent, such as the one presented by Yang et al. (2014) for the distributed monitoring of fibre optic or the one formulated by Assef et al. (2018) in the context of medical ultrasound imaging.

If one adresses the different units in the horizontal and vertical axes, one can transform the DSP problem of envelope detection in the geometric problem of defining the shape of a set of points in $\mathbb{R}^2$. For that purpose Edelsbrunner et al. (1983) introduced the concept of alpha-shapes, a mathematically well defined extension to the convex hull of a finite set of points, 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 (Varytimidis et al., 2016), reconstruction of surfaces from a cloud of points (Wu et al., 2015) and Spectroscopy (Xu et al., 2019), with the last work, that involves the estimation and removal of the Blaze function (a kind of envelope) of a echelle spectrograph, being particularly align with what is intended here.

Thus, in this paper we formulate a general approach to envelope detection, exploiting the intrinsic characteristics of a generic, spectrally complex, wave in order to avoid the need to 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 (Tarjano & Pereira, 2019), and a accurate description of the envelope would be useful for a cleaner spectral analysis. Moreover, the algorithm naturally divides a signal in pseudo-cycles, that could serve pragmatic building blocks for the reconstruction of the fine structure of the wave.

In the context of sound synthesis, 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.

2 Methodology

We start by defining a pulse as a series of consecutive samples in a discrete wave with 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 a subset of the natural numbers. 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 $P[j], j \in \mathbb{N}_0, j < q$ in W can then be defined as a sequence of samples indexed by i such that $\text{sign}(W[a-1]) \neq \text{sign}(W[a]) = \text{sign}(W[a+1]) = \cdots = \text{sign}(W[b-1]) = \text{sign}(W[b]) \neq \text{sign}(W[b+1]) \forall i, a \leq i \leq b | a, b \in \mathbb{N}_0$.

In other words, each pulse is a unique subset of W with its members being the maximum sequence of samples in W with the same signal and $P[0] \cup P[0] \cup \cdots \cup P[q-1] = W$. It is convenient to attribute to each pulse a pair of coordinates: those will be the coordinates of the sample with the maximum amplitude in each pulse.

Naturally, each pulse can be classified as positive or negative, based on the signal of its samples, which must be the same. For a continuous function, then, a pulse would correspond to the pieces between it's roots, and would be equal to half the cycle of a sinusoid. Thus, in our analogy, two consecutive pulses, naturally with opposing signs, have the potential do 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 will be defined later.
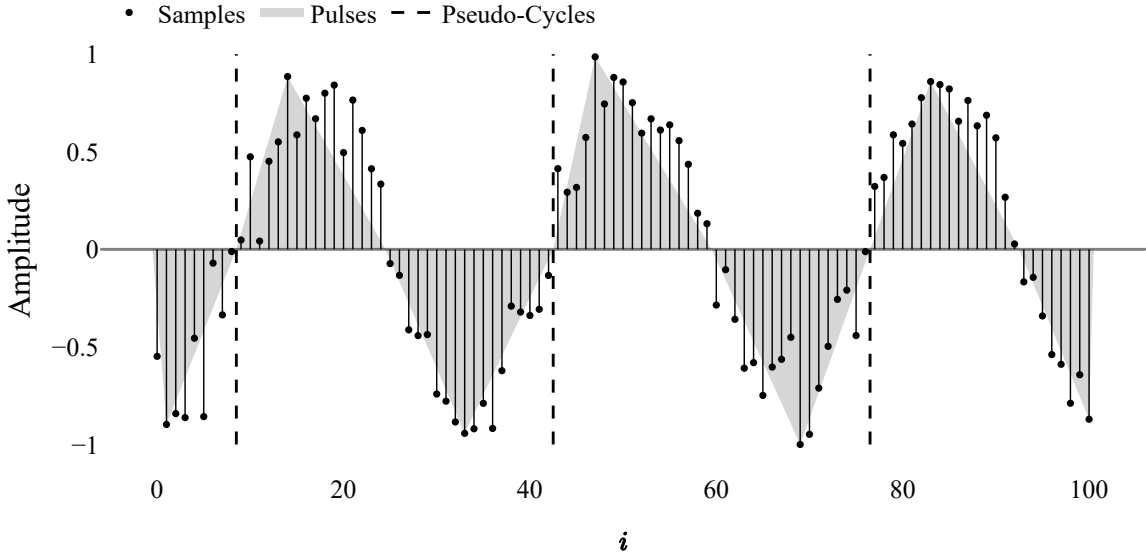

Figure 1: The grey areas encompass the simplified convex hulls of the pulses. The dashed lines mark the frontier between pseudo-cycles.

The most mathematically sound definition of an envelope involves its representation as an analytic signal, introduced by Gabor in 1946 (Hahn, 2007). In his work, Gabor (1946) applies the then relatively new mathematical machinery of the quantum mechanics to unify time and frequency representations of a wave, showing how the Hilbert transform could be applied to a real signal in order to obtain a complex signal that would become known as the analytic signal.

This analytic signal is of the form $a(t) = s(t) + \mathbb{H}(s(t))$ i (He & Dellwo, 2016) where $s(t)$ is the original real signal, that becomes the real part of the analytic signal. $\mathbb{H}(s(t))$, the Hilbert transform of the original signal becomes, then, 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 Hilbert transform, for a pure sinusoid of local frequency 20, modulated by a polynomial of the third degree, in the presence and absence of gaussian white noise with standard deviation of 1/10 of the wave's maximum amplitude. It is readily noticeable 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, specially in the case of sound, to present somewhat different
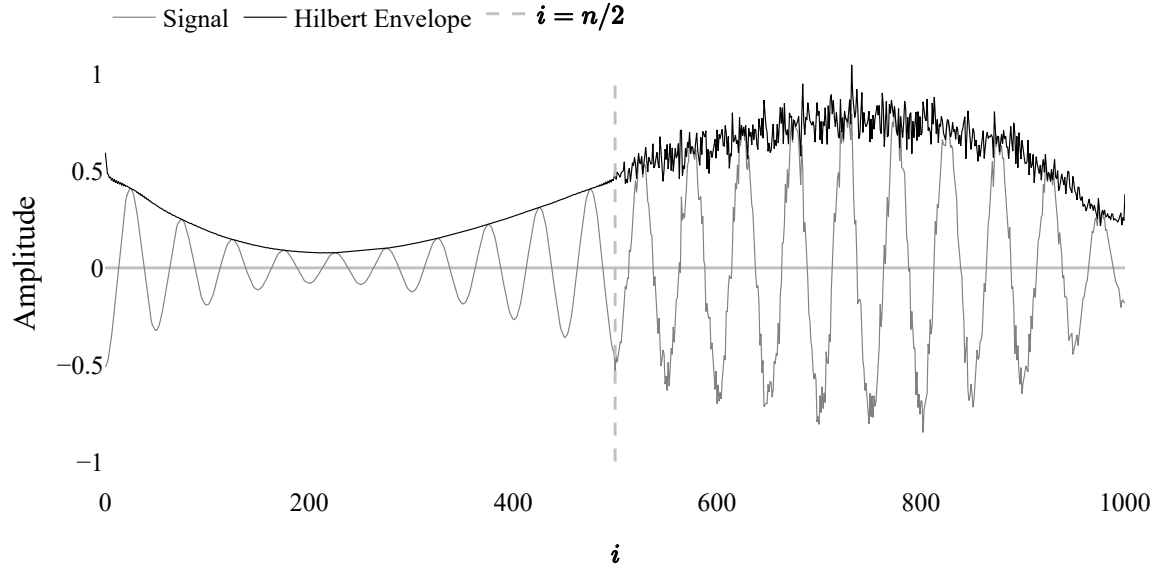
Figure 2: Envelope of a pure sinusoid modulated by a polynomial of degree 3, as obtained by the Hilbert Transform. In the first half the sinusoid is free of noise, while in the second half white gaussian noise with a standard deviation of 1/10 of the maximum amplitude of the wave was added to the base signal.

positive and negative contours; as our method will forcibly have to be applied separately to both sides, superior and inferior, of the discrete wave, it is only natural to introduce such separation as soon as possible.

To do so we will define the positive (upper) frontier and negative (lower) frontier as the set of points $(i, |W[i]|)$ that are, necessarily but not sufficiently, the point of maximum amplitude of their respective pulses. 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 to address the different units in which the amplitude and horizontal distance of our wave are expressed.
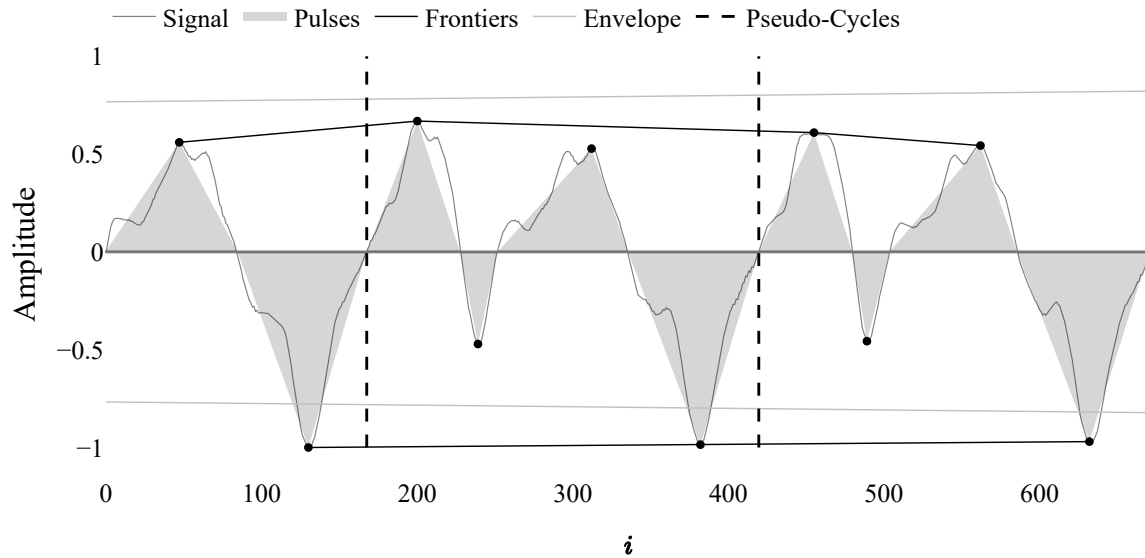

Figure 3: Positive and negative frontiers, envelope and pseudo-cycles for a fragment of piano sound. While we have only one (positive) envelope, here mirrored in the inferior space, we have two frontiers, the superior and inferior.

To make abscissa and ordinate units consistent for the computation of the frontier, and proceed with a geometric treatment of the subject, we consider the set of positive and negative pulses separately. For each set, we multiply all pulses 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 frame related unit i, related to time by the equation $i = t\ fps$.

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

3

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 $\max(P[j]) \approx \max(P[j+2])$. That justifies the approximation of the convex hull of a pulse by the triangle defined by the points where the pulse crosses the horizontal the point of maximum amplitude; this will allow a considerable dimension reduction as this triangular pulse can become the atomic unit for the rest of the method.

This definition also 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.

Translating the intuitive explanation of alpha-shapes in Edelsbrunner and Mücke (1994) to our context, we can say that the points in the frontier are those touched by a circle 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 therefore 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 belong to their respective frontiers, including eventual pulses between them, as can be seen in figure 3.

It is now necessary to infer the appropriate radius 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 (Fleischmann et al., 2010) for which no default definition exists; The two possible approaches are the derivation of direct methods that use characteristics of the discrete wave to calculate its curvature at each point, or the calculation of the curvature of a curve fitted to the discrete wave (Coeurjolly et al., 2001).

To fit a polynomial to the a generic wave two approaches are readily available: the least mean squares approach, that seeks to minimize the dependant variable errors and the total least squares problem, that treats both variables symmetrically (Golub & van Loan, 1980).

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 (Markovsky & Huffel, 2007); this leads us to turn our attention to direct approaches of curvature estimation.

2.1 Discrete Curvature Estimation - The Equivalent Circle Approach

As previously stated, no default discrete curvature definition exists; Carroll et al. (2014), for example, derives 3 such different definitions based on the approximation of a circle by an inscribed, centred e circumscribed polygon. In the context of 3d meshes, Váša, Kühnert, et al. (2016) evaluate a range of estimators from a multivariate point o view.

For our needs, however, it is more convenient to define our own discrete curvature estimator; the approach here presented differs from most others found in the literature in that the curvature is defined over the edges linking the points that form the discrete wave, instead of the vertices.

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 (Váša, Vaněček, et al., 2016). 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. This approach presents the convenience of estimating the radius directly.
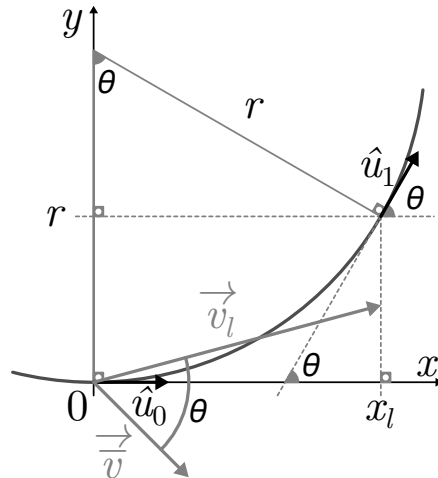

Figure 4: Change of direction from horizontal, in $\hat{u}_0$, until an angle of θ, in $\hat{u}_1$, that is equivalent to the angle between the average direction and the direction of the vector $\vec{v}_l$ whose radius of curvature is being calculated, shown in the picture in grey.

Formally, for each frontier, positive or negative, we define $m - 1$ vectors, m being the number of positive or negative pulses respectively, as the coordinates of the posterior pulse with the relevant signal minus the preceding pulse: $\vec{v}_l^{\pm} = (P[l + 1]_x^{\pm} - P[l]_x^{\pm}, P[l + 1]_y^{\pm} - P[l]_y^{\pm}), \forall 0 \leq l < m - 1$, recalling that the x coordinate of a pulse was defined as the x coordinate of its sample with maximum amplitude, while the y coordinate of a pulse is this amplitude itself.

Since an edge defines only one direction, there remains the need for a reference. As we are interested in the relative curvature, it is natural to use the average direction of all vectors, $\vec{\overline{v}}^{\pm} = \frac{1}{m-1} \sum_{l=0}^{m-1} \vec{v}_l^{\pm}$, so as to measure the relative change of direction of each edge. This procedure, as well as the circles generated by it, are illustrated in figure 5.

The angle θ between the average direction and each of the vectors can be obtained via the relation of the slope of the respective lines, as in the expression $\theta = \arctan((m - \overline{m})/(1 + m\overline{m}))$. The radii of the equivalent circles are thus given by $r_l = x_l/\sin(\theta_l)$ with the curvature given by $\kappa_l = 1/r_l = \sin(\theta_l)/x_l$.

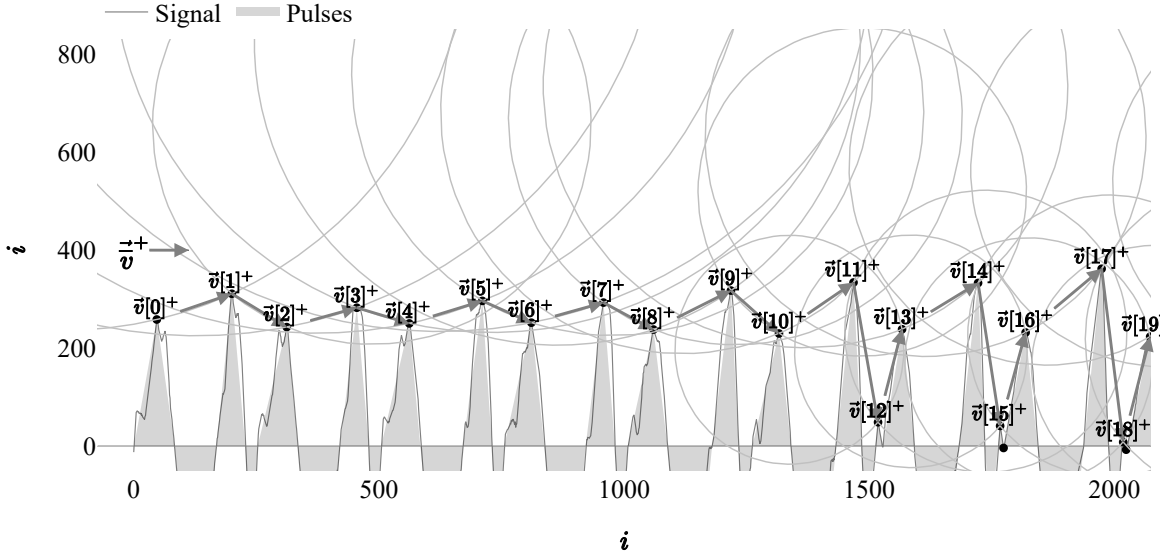$$r_l = \left| \frac{x_l \sqrt{\overline{m}^2 + 1} \sqrt{x_l^2 + y^2}}{\overline{m} x_l - y} \right| \tag{1}$$

Figure 5: The equivalent circles obtained via the procedure described for the positive frontier. θ, the angle between $\vec{\overline{v}}^+$ 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.

After this procedure one obtains a set of points in $\mathbb{R}^2$ where the ordinate represents the instantaneous radius, with its localization in the space of the wave given by the ordinates. Given the robustness of the method, for an average wave of a couple of seconds at an ordinary rate of 44100 *fps* the average of the radius values is sufficient to assure the identification of the frontier. However, a curve fitting method can be used in the case of longer waves. An example of the use of the average equivalent circle can be seen in figure 6.

3 Results

Figure 7 illustrates the frontiers obtained to six diverse discrete sound waves, as well as a detail view of a shorter segment for each wave. 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 harmonic an inharmonic sounds, and presents robustness in relation to the number of samples and the frequencies of the waves.

The impact of the normalization of the wave with 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.

3.1 Robustness

While a theoretical basis for the normalization of the amplitude is given, and is justified in practice by the computational stability issues that might arise with arbitrary values, the algorithm is, from a pure mathematical point of view, impervious to amplitude scaling, as we now prove.
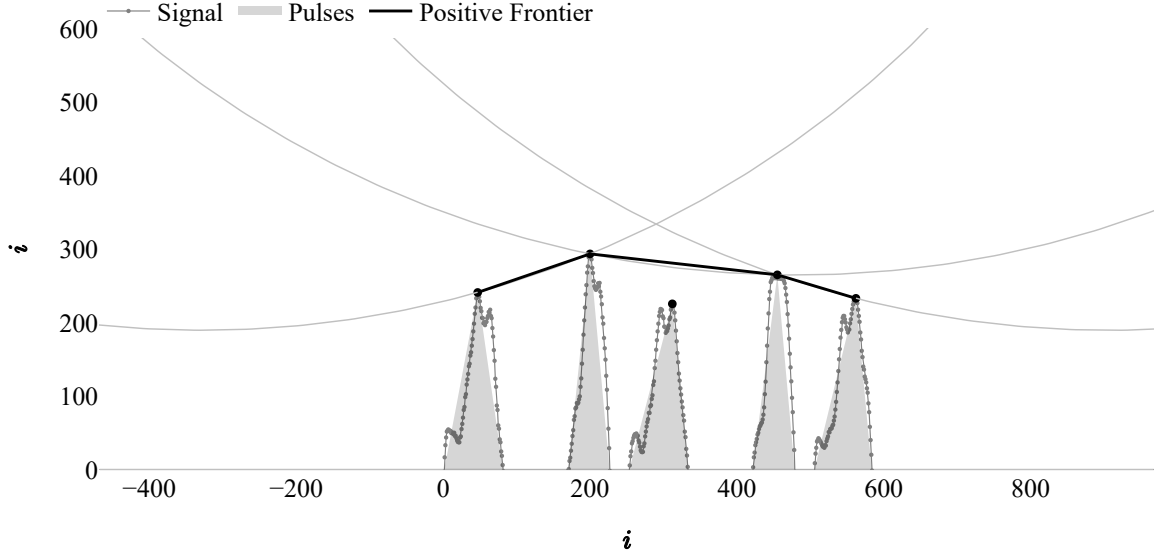
Figure 6: The positive frontier as obtained by the use of the circle representing the average curvature of the discrete wave illustrated.
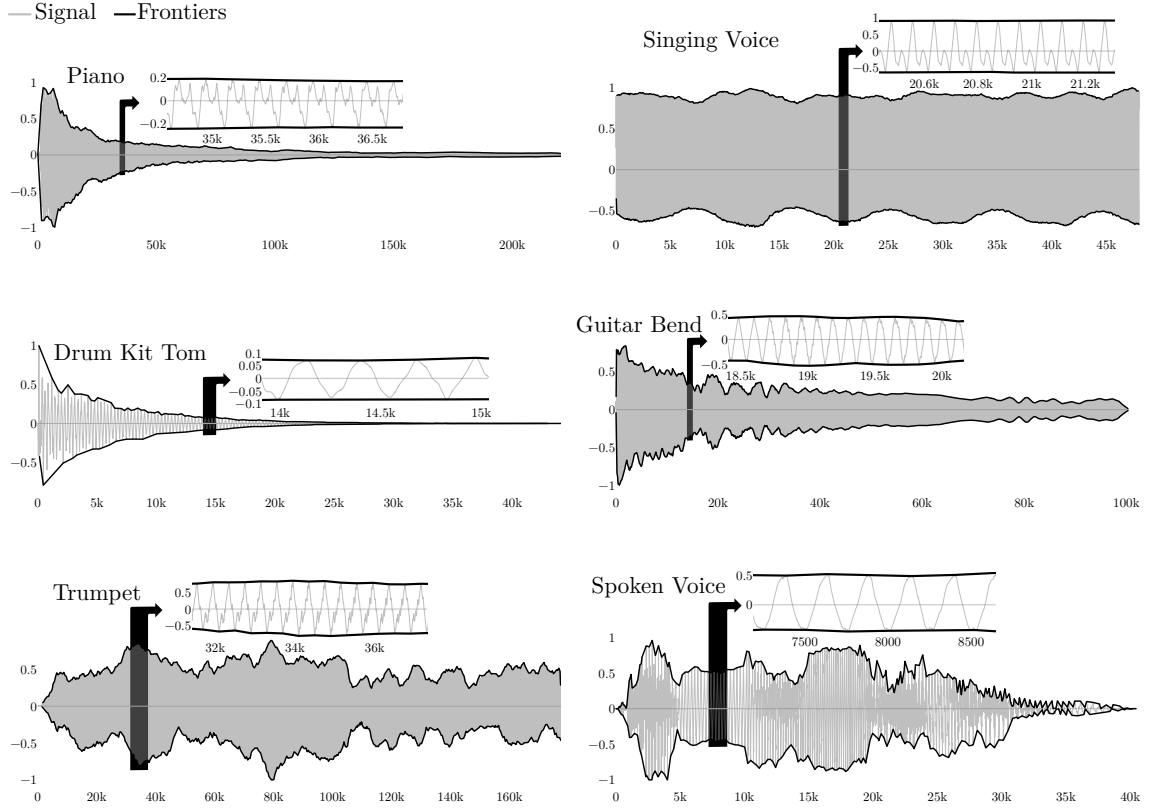

Figure 7: Positive and negative frontiers of 6 digital waves. For each wave, the region highlighted in black is shown in detail. For each wave the horizontal axis is the sample number i, while the vertical axis is the normalized amplitude. All 6 waves were recorded at 44100 fps.

let s be a positive real number that multiplies every sample of a discrete wave W. The subdivision in pulses of such a wave is unchanged by the vertical scaling operation, and we can write the slope of a line segment from a pulse to the next pulse with the same sign as $m_l = (sy_{l+1} - sy_l)/(x_{l+1} - x_l)$.

Assuming, for the sake of simplicity, $\overline{m} = 0$, the radius of the equivalent circle of such a segment, as per equation 1, is
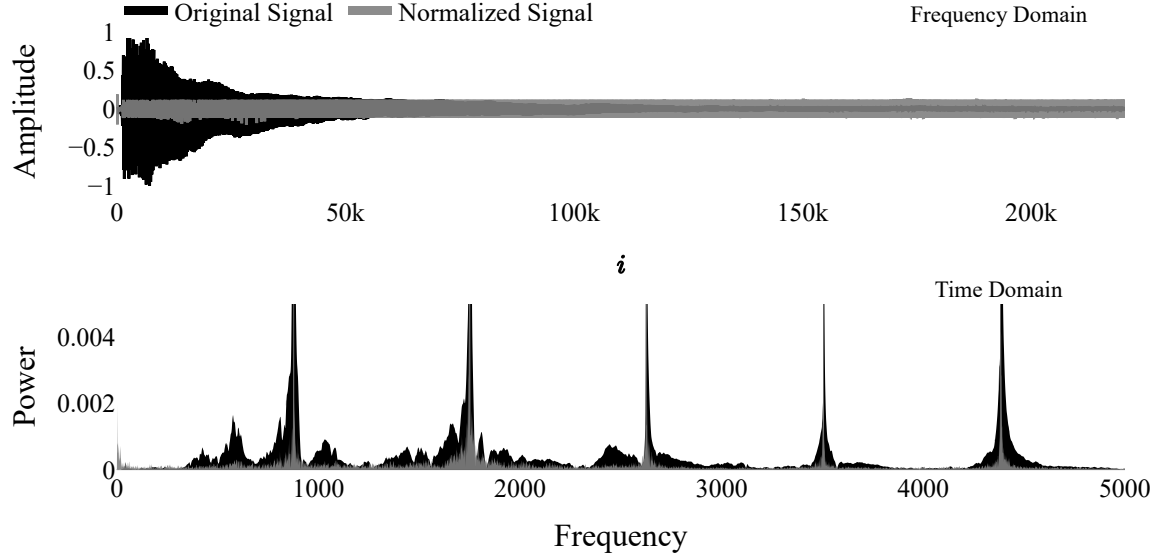
Figure 8: Original and normalized views of a recording of key 33 (F3, 175 Hz) from a grand piano both in time and frequency domains.

$$r_l = \left| \frac{\sqrt{(sy_{l+1} - sy_l)^2 + (x_{l+1} - x_l)^2} \, (x_{l+1} - x_l)}{sy_{l+1} - sy_l} \right|$$

And the square of the distance d_l from the center of a circle C with that radius passing through the points defined by the coordinates of both pulses to the median point $\bar{p} = \big((x_l + x_{l+1})/2, (y_l + y_{l+1})/2\big)$ of the line segment is

$$d_l^2 = -\frac{(sy_{l+1} - sy_l - 2x_{l+1} + 2x_l)(sy_{l+1} - sy_l + 2x_{l+1} - 2x_l)(s^2 y_{l+1}^2 - 2s^2 y_{l+1} y_l + s^2 y_l^2 + x_{l+1}^2 - 2x_{l+1} x_l + x_l^2)}{4s^2 (y_{l+1} - y_l)^2}$$

Assuming invariance to vertical scaling, $\bar{p}$ lying inside or outside the circle must not depend on s; $|d_l| \le |r_l| \iff d_l^2/r_l^2 \le 1 \iff \bar{p} \in C$.

$$\frac{d_l^2}{r_l^2} = \frac{-\frac{s^2\left(y_{i+1}^2 + y_i^2\right)}{4} + \frac{s^2 y_{i+1} y_i}{2} + (x_{i+1} - x_i)^2}{(x_{i+1} - x_i)^2} \le 1$$

Reorganizing:

$$y_{i+1} y_i \le \frac{y_{i+1}^2 + y_i^2}{2} \iff \bar{p} \in C$$

Thus, the circle C containing the point $\bar{p}$ doesn't depend on s nor on the horizontal parametrization of W *quad*∎
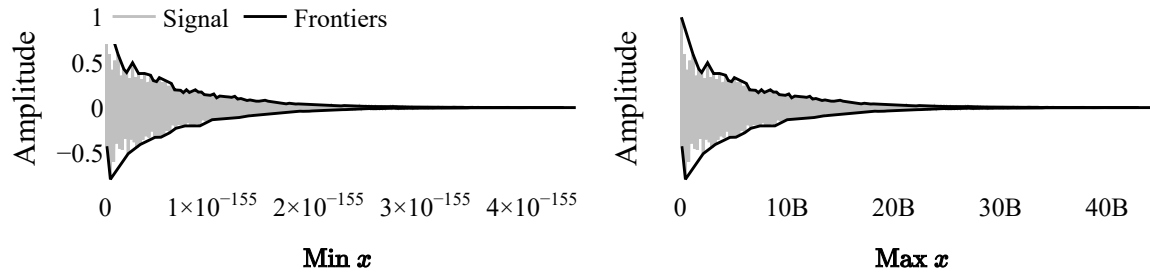

Figure 9: Minimum and maximum scales of the abscissa before computational stability issues arose, for a recording of a drumkit tom.

Figures 9 and 10 illustrate the computational resilience of the algorithm for scaling in both the abscissa and ordinate axes, respectively. Although mathematically independent of such parametrizations, in practice care must be taken, for different reasons, in the choice of values for both axes.
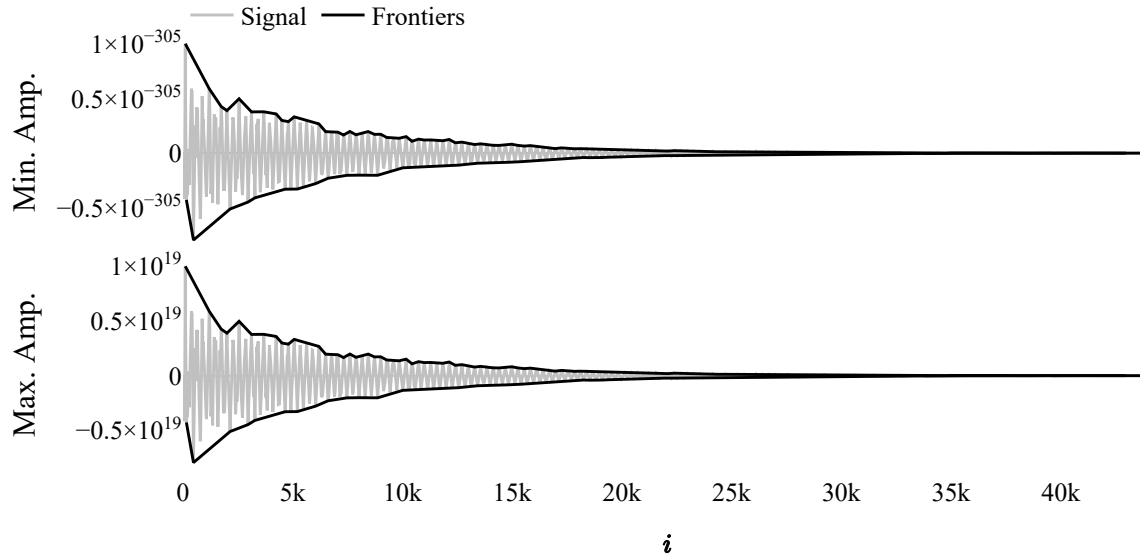
Figure 10: Minimum and maximum values of the amplitude before computational stability issues arose, for a recording of a drumkit tom.

3.2 Fundamental Frequency as a Welcome Side-effect

4 Discussion + Conclusion

5 References

Assef, A. A., Ferreira, B. M., Maia, J. M., & Costa, E. T. (2018). Modeling and FPGA-based implementation of an efficient and simple envelope detector using a hilbert transform FIR filter for ultrasound imaging applications. *Research on Biomedical Engineering*, *34*(1), 87–92. https://doi.org/10.1590/2446-4740.02417

Carroll, D., Hankins, E., Kose, E., & Sterling, I. (2014). A survey of the differential geometry of discrete curves. *The Mathematical Intelligencer*, *36*(4), 28–35. https://doi.org/10.1007/s00283-014-9472-2

Coeurjolly, D., Miguet, S., & Tougne, L. (2001). Discrete curvature based on osculating circle estimation. In *Lecture notes in computer science* (pp. 303–312). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45129-3_27

Edelsbrunner, H., Kirkpatrick, D., & Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, *29*(4), 551–559. https://doi.org/10.1109/tit.1983.1056714

Edelsbrunner, H., & Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics (TOG)*, *13*(1), 43–72. https://doi.org/10.1145/174462.156635

Fleischmann, O., Wietzke, L., & Sommer, G. (2010). A novel curvature estimator for digital curves and images. In *Lecture notes in computer science* (pp. 442–451). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15986-2_45

Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, *93*(26), 429–441. https://doi.org/10.1049/ji-3-2.1946.0074

Golub, G. H., & van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, *17*(6), 883–893. https://doi.org/10.1137/0717073

Hahn, S. L. (2007). The history of applications of analytic signals in electrical and radio engineering, In *EUROCON 2007 - the international conference on "computer as a tool"*, IEEE. https://doi.org/10.1109/eurcon.2007.4400463

He, L., & Dellwo, V. (2016, September). A praat-based algorithm to extract the amplitude envelope and temporal fine structure using the hilbert transform, In *Interspeech 2016*, ISCA. https://doi.org/10.21437/interspeech.2016-1447

Lokki, T., Pätynen, J., Tervo, S., Siltanen, S., & Savioja, L. (2011). Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *The Journal of the Acoustical Society of America*, *129*(6), EL223–EL228. https://doi.org/10.1121/1.3579145

Lyons, R. (2017). Digital envelope detection: The good, the bad, and the ugly [tips and tricks]. *IEEE Signal Processing Magazine*, *34*(4), 183–187. https://doi.org/10.1109/msp.2017.2690438

Markovsky, I., & Huffel, S. V. (2007). Overview of total least-squares methods. *Signal Processing*, *87*(10), 2283–2302. https://doi.org/10.1016/j.sigpro.2007.04.004

Meng, Q., Yuan, M., Yang, Z., & Feng, H. (2013). An empirical envelope estimation algorithm, In *2013 6th international congress on image and signal processing (CISP)*, IEEE. https://doi.org/10.1109/cisp.2013.6745226

Qi, B., Mao, Y., Liu, J., Liu, B., & Xu, L. (2017). Relative contributions of acoustic temporal fine structure and envelope cues for lexical tone perception in noise. *The Journal of the Acoustical Society of America*, *141*(5), 3022–3029. https://doi.org/10.1121/1.4982247

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Tarjano, C., & Pereira, V. (2019). 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* (pp. 362–375). Springer International Publishing. https://doi.org/10.1007/978-3-030-30490-4_30

Varytimidis, C., Rapantzikos, K., Avrithis, Y., & Kollias, S. (2016). Alpha-shapes for local feature detection. *Pattern Recognition*, *50*, 56–73. https://doi.org/10.1016/j.patcog.2015.08.016

Váša, L., Vaněček, P., Prantl, M., Skorkovská, V., Martınek, P., & Kolingerová, I. (2016). Mesh statistics for robust curvature estimation. *Computer Graphics Forum*, *35*(5), 271–280. https://doi.org/10.1111/cgf.12982

Váša, L., Kühnert, T., & Brunnett, G. (2016). Multivariate analysis of curvature estimators. *Computer-Aided Design and Applications*, *14*(1), 58–69. https://doi.org/10.1080/16864360.2016.1199756

Wu, B., Yu, B., Huang, C., Wu, Q., & Wu, J. (2015). Automated extraction of ground surface along urban roads from mobile laser scanning point clouds. *Remote Sensing Letters*, *7*(2), 170–179. https://doi.org/10.1080/2150704x.2015.1117156

Xu, X., Cisewski-Kehe, J., Davis, A. B., Fischer, D. A., & Brewer, J. M. (2019). Modeling the echelle spectra continuum with alpha shapes and local regression fitting. *The Astronomical Journal*, *157*(6), 243. https://doi.org/10.3847/1538-3881/ab1b47

Yang, W., Yang, Y., & Yang, M. (2014). Fast digital envelope detector based on generalized harmonic wavelet transform for BOTDR performance improvement. *Measurement Science and Technology*, *25*(6), 065103. https://doi.org/10.1088/0957-0233/25/6/065103

Zhu, Z., Miyauchi, R., Araki, Y., & Unoki, M. (2018). Contributions of temporal cue on the perception of speaker individuality and vocal emotion for noise-vocoded speech. *Acoustical Science and Technology*, *39*(3), 234–242. https://doi.org/10.1250/ast.39.234