AUDIO DE-THUMPING USING HUANG'S EMPIRICAL MODE DECOMPOSITION

Paulo A. A. Esquef[*]
Coordination of Systems and Control
National Lab. of Scientific Computing - MCT
Petrópolis, Brazil
pesquef@lncc.br

Guilherme S. Welter
Coordination of Systems and Control
National Lab. of Scientific Computing - MCT
Petrópolis, Brazil
gswelter@lncc.br

ABSTRACT

In the context of audio restoration, sound transfer of broken disks usually produces audio signals corrupted with long pulses of low-frequency content, also called thumps. This paper presents a method for audio de-thumping based on Huang's Empirical Mode Decomposition (EMD), provided the pulse locations are known beforehand. Thus, the EMD is used as a means to obtain pulse estimates to be subtracted from the degraded signals. Despite its simplicity, the method is demonstrated to tackle well the challenging problem of superimposed pulses. Performance assessment against selected competing solutions reveals that the proposed solution tends to produce superior de-thumping results.

1. INTRODUCTION

Severe damages or discontinuities to the grooves of a disk, such as those produced by deep scratches or breakages, may give rise to long-duration pulses of low-frequency content in the resulting audio signal, during disk playback [1, 2]. Being typically preceded by high-amplitude impulsive disturbances, long pulses are considered the impulse response of the stylus-arm system added to the waveform of interest [1, 2].

An illustration of a synthetically generated pulse is seen in Figure 1. As can be seen, the pulse waveform seems to have both an amplitude-modulated component (decaying exponential envelope) and a frequency-modulation component. Typically, the pulse oscillations start at about 150 Hz, right after the initial click, and decay exponentially down to about 10 Hz.

To the author's knowledge, apart from crude high-pass filtering, which is usually unsatisfactory, there are three available techniques for long pulse removal or audio de-thumping. Brief descriptions of these methods follow.

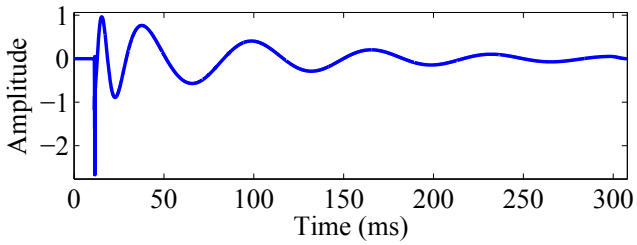

Figure 1: *Synthetic example of a long pulse. The initial click occurs at about 11 ms.*

The so-called template matching method has been first proposed by Vaseghi in [1] and [3]. Its main assumption is that long pulses are shape-invariant, for being the impulse response of a given stylus-arm system. Thus, if a clean version of the pulse (a template) is available, its time-reversed version can be used as a matched filter to detect other pulse occurrences in the audio signal. Furthermore, amplitude-scaled versions of the template can be used to suppress corrupting pulses from the signal by simple subtraction. The remaining initial click is supposed to be removed afterward by standard de-clicking techniques [2].

In [4, 2, 5] Godsill and colleagues have proposed a model-based signal separation technique for audio de-thumping. The first step of the method consists in estimating two distinct autoregressive (AR) models: one of high order for the signal of interest and another of low-order for the corrupting pulse. Then, pulse removal is achieved by separation of the two AR processes. In this approach, the initial click is taken as part of the long pulse, being modeled by the same AR model of the pulse, but with a much higher excitation variance. Therefore, suppression of the initial click and the pulse is taken care of at once by the method.

Simple non-linear filtering techniques for audio de-thumping have been proposed in [6] by Esquef and colleagues. In this solution, hereafter referred to as the TPSW method, an initial estimate for the long pulse is obtained from a non-linear filtering technique called two-pass split window (TPSW) [7, 6], which is capable of producing relatively smooth pulse estimates, despite the presence of the high-amplitude clicks that precede the pulses. Then, these pulse estimates are made even smoother by means of an overlap-and-add signal segmentation with low-order polynomial fitting.

Thorough comparisons among the aforementioned methods is beyond the scope of this paper. Nevertheless, in general terms, the main advantages of the template matching method are its simplicity and ability to detect long pulses even if the initial clicks are absent. However, the solution is less flexible to tackle more challenging situations, such as the occurrence of superimposed long pulses, i.e., when a second pulse appears within the duration of a preceding one. According to the results presented in [6] the AR-separation and TPSW methods perform equally well, being the latter less intense computationally. Moreover, both can treat superimposed pulses.

In this paper an alternative solution to audio de-thumping is proposed. More specifically, it makes use of the so-called Empirical Mode Decomposition (EMD) [8] and its improved version, the Complementary Ensemble EMD (CEEMD) [9, 10], as a means to obtain estimates of long pulses corrupting an audio signal of interest. In principle, the EMD is capable of decomposing a given time-domain waveform into a finite set of Intrinsic Mode Functions (IMFs) and a monotonic residue, each IMF being a single AM-FM component. Since a long pulse can be well characterized as a single AM-FM component, the choice of the EMD for the

[*] The work of Dr. Esquef was supported by CNPq-Brazil via grants no. 472856/2010-3 and 306607/2009-3.

problem at hand seems justified.

The experimental results reported in this paper reveal that the EMD and the CEEMD are effective and simple tools to provide adequate pulse estimates. Performance evaluation of the proposed method against the AR-separation and TPSW methods is carried out via the Perceptual Audio Quality Measure (PAQM) [11]. The attained results show that the CEEMD-based audio de-thumping performs comparably to the competing solutions.

The remainder of the paper is organized as follows. In Section 2 brief reviews of the EMD and the CEEMD are given. The proposed pulse estimation method is explained in Section 3. The experimental setup defined for the comparative tests is described in Section 4. In Section 4.3 the attained results are presented and discussed. Finally, conclusions are drawn in Section 5.

2. THE EMPIRICAL MODE DECOMPOSITION

The Empirical Mode Decomposition was originally introduced by Huang and collaborators [8] as a way to decompose multicomponent signals into constituent functions from which meaningful instantaneous frequencies could be estimated via the Analytical Signal approach [12]. The EMD decomposition does not assume any basis function since it is an entirely data-driven iterative algorithm that operates over signal envelopes.

The EMD method decomposes a signal into components called *Intrinsic Mode Functions* (IMFs), which are typically characterized by zero-mean oscillations modulated by a slowly varying envelope. An IMF must have the following properties [8, 13, 14]:

1. The number of extrema and the number of zero-crossings must be either equal or differ at most by one;
2. The arithmetic mean between the upper and lower envelopes of an IMF must be zero at any point of its domain.

With reference to the item 2 above, the upper (lower) envelope is usually obtained via low-order polynomial fitting to the local maxima (minima) of the signal. Variations on the EMD algorithm exist [14, 15, 16] and are mainly concerned with two issues: different criteria to stop the intermediate iterative sifting procedure that culminate in an acceptable IMF; and alternative data extrapolation schemes to obtain the signal envelopes [17, 10].

2.1. EMD Implementation

For the experiments reported in the paper, a standard version of the EMD, *i.e.*, one that uses a Cauchy-type stopping criterion and natural cubic spline interpolation to compute the envelopes [8, 13], has been implemented in Matlab. Alternatively, these envelopes can also be obtained via piecewise cubic Hermite interpolating polynomials across local maxima (minima).

Considering a signal $x(t)$, the EMD is described as follows.

1. Let $j = 1$ and set $x_j(t) = x(t)$;
2. Identify all local maxima and minima of $x_j(t)$;
3. Obtain the upper envelope $e_{\text{upper}}(t)$ (respectively, lower envelope $e_{\text{lower}}(t)$) by polynomial interpolation across the local maxima (respectively, local minima) of $x_j(t)$;
4. Compute the mean envelope $m(t) = \big[e_{\text{upper}}(t)+e_{\text{lower}}(t)\big]/2$;
5. Obtain an IMF estimate $C_j(t) = x_j(t) - m(t)$;
6. If $m(t)$ is a non-monotonic function (or if it has enough extrema to allow envelope computation), make $x_j(t) = m(t)$; increment j by one unit; and go back to step 2 to obtain the subsequent IMFs. Otherwise, stop the iterations and set the residue of the decomposition as $r(t) = m(t)$.

In practice, step 5 above is insufficient to produce a proper IMF. To remedy this, step 5 is modified to include an inner loop to perform additional siftings to $x_j(t)$. More specifically, a so-called proto-IMF is defined as $C_{j,k}(t) = x_j(t) - m(t)$ for the k^{th} iteration of the sifting loop, which must continue until a stopping criterion (defined below) is satisfied. If an additional sifting is needed, then one sets $x_j(t) = C_{j,k}(t)$ and returns to step 2 going through step 5. The final IMF estimate $C_j(t)$ is then obtained as the last $C_{j,k}(t)$ of the sifting loop.

Here, the chosen stoping criterion is the same adopted in [13], i.e., the iterations stop when the quantity

$$S_d = \text{Var}\big\{C_{j,k-1}(t) - C_{j,k}(t)\big\}/\text{Var}\big\{C_{j,k-1}(t)\big\} \qquad (1)$$

becomes smaller than a pre-assigned value, typically within the range $[0.0001–0.0003]$. After obtaining a total of J IMFs and a residual trend $r_J(t)$, the original signal can be reconstructed by summing up all IMFs and the trend: $x(t) = \sum_{j=1}^{J} C_j(t) + r_J(t)$.

For broad spectrum signals such as those of fractal or Gaussian processes, the maximum number of IMFs is approximately $\log_2 L$, where L is the number of samples of the signal [18].

2.2. Complementary Ensemble EMD

EMD has been successfully used for analysis of diverse kinds of signals, mainly due its ability to tackle responses of non-linear and non-stationary systems. Nevertheless, the presence of intermittency in such signals often results in a phenomenon called *mode mixing* [13, 14], where coherent parts of the signal may end up in adjacent IMFs, thus devoid of physical meaning.

The original EMD algorithm is sensitive to the addition of small perturbations to the input signal, in the sense that it may produce a new set of IMFs in comparison with those of the noiseless version. Based on this fact, Wu and Huang [10] proposed that more reliable IMFs should be estimated from EMD of an ensemble formed by a given input signal artificially corrupted with several realizations of Gaussian noise.

The idea behind of the Ensemble EMD (EEMD) is to take a large number of noisy versions of the original signal

$$x^{(i)}(t) = x(t) + \varepsilon\, w^{(i)}(t), \qquad (2)$$

where $\varepsilon\, w^{(i)}(t)$ is the i-th realization of a zero-mean white Gaussian noise with standard deviation ε, which can be made a fraction of that of $x(t)$. After obtaining the IMFs $C_j^{(i)}(t)$ for each realization $x^{(i)}(t)$, the final result is obtained by averaging the IMFs across all realizations:

$$\tilde{C}_j(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} C_j^{(i)}(t). \qquad (3)$$

Given a non-infinitesimal ε and a large enough N, the residual noise amplitude will be proportional to $\varepsilon/\sqrt{N}$ and the resulting IMFs stable. More importantly, the EEMD largely reduces the mode mixing problem [10], thus improving the EMD performance at the expense of a much higher computational load.

A further improvement to the EEMD is the Complementary EEMD (CEEMD), in which the ensemble is formed by $N/2$ complementary pairs of noise realizations with symmetric amplitude. This way, Eq. (2) is modified to $x^{(i)}(t) = x(t)+(-1)^i \varepsilon\, w^{(i-\gamma)}(t)$, for $i = 1, 2, \ldots, N$, where $\gamma = i$ modulo 2. The IMFs of the thus constructed ensemble are then obtained as before via Eq. (3). This procedure ensures that the residual noise will be zero.

3. LONG PULSE ESTIMATION VIA EMD AND CEEMD

Similar to the AR-separation and TPSW methods, the proposed EMD-based de-thumping requires prior knowledge of pulse locations in time. This means that, for a particular pulse, estimates of its onset time and duration must be available. In practice, the former can be inferred from the location of the initial click, which can easily obtained by standard detection techniques [2].

From the auditory perception perspective, the most salient part of a long pulse is its beginning, for its higher amplitude and frequency. Therefore, pulse duration estimates can be obtained by visual inspection. In other words, underestimation of pulse durations is likely to produce no audible effects.

3.1. EMD-based Estimation of Single Pulses

For didactic reasons, EMD-based estimation of single pulses is presented first, being that of superimposed pulses left to later.

The main steps of the pulse estimation procedure (one pulse at a time) are listed below. More specific details of each step are given in the sequel.

1. Select a portion of the signal of interest containing one single long pulse (to be called input signal hereafter);
2. Extend the input signal backward in time;
3. Analyze the extended input signal via the EMD. The main parameter to be defined is the maximum number of IMFs.
4. Form the pulse estimate by mixing together partial reconstructions of the signal, with different levels of detail, via an overlap-and-add windowing scheme.

As regards step 1, the beginning of the input signal should coincide with that of the pulse, i.e., it should start right after the initial click. The duration of the segment should be approximately that of the observed long pulse. It is advisable though to add about 5 ms to the duration in order to overcome boundary effects that may affect IMF estimation [17, 10]. For that very reason, step 2 is taken. The idea here is to analyze an input signal a bit longer than necessary and then discard samples at the extremities of the ensuing IMFs and residue to get rid of possible boundary effects. Therefore, the backward signal extrapolation carried out in step 2 does not need to be much involved. It can be simple enough to just capture the tendency of the signal trajectory.

Signal extension backward in time should be made for at least the duration of the initial click. For that, extrapolation schemes based on AR modeling [2, 19] can be used. A simpler solution, which is employed here, consists in mirroring the beginning of the input signal with odd symmetry w.r.t. its first sample.

The EMD in step 3 uses a standard version of the algorithm (see Section 2.1). As for the treatment of envelope boundaries, the solution proposed in [10] is resorted to. More specifically, considering the upper envelope for didactic reasons, a straight line is first fitted to the two consecutive maxima nearest to the end (or beginning) of the signal. Then, an artificial new end (or beginning) point for the envelope is created at the end (or beginning) of the segment. This new point is taken as the largest value between the own signal and the linearly extrapolated envelope. A similar scheme can be employed to extend the lower envelope. In both cases, no extension of the input signal is carried out, just extrapolation of its lower and upper envelopes toward its boundaries.

The aforementioned procedure surely helps to reduce end effects observed in the IMFs, but do not completely eliminate them. Hence, input signal extrapolation performed in step 2 is still needed.

One of the known issues of the standard EMD is the so-called mode-mixing or intermittence [8], which consists of the split of an apparent single intrinsic mode between two adjacent IMFs. Intrinsic mode segregation is a complex question whose discussion is beyond the scope of this paper. As reported in [20] it depends on parameters such as the relative amplitude of the modes and their proximity in frequency.

As one could anticipate from the above discussion, although a single long pulse would qualify for being an IMF, it is not always true that one of the IMFs produced by EMD of the input signal constitutes alone an adequate pulse estimate. The main reason for that seems to be the overlap between the spectral range of the long pulse and the low-frequency content of the audio signal of interest.

An illustration of the mode-mixing problem in the context of EMD-based pulse estimation is depicted in Figure 2. As can be seen, while the tail of the pulse is well captured by the residue obtained after extracting seven IMFs, adequate representation of the initial faster oscillations only happens if the 7th IMF is added to the residue.

Similar to the strategy employed in [6], a practical way to obtain a useful pulse estimate is to predefine three temporal regions for the pulse and assign to each region pulse estimates with different degrees of detail (or frequency ranges). One pulse partition that typically works in practice is the following:

p_{F}: about half oscillation cycle from the beginning of the pulse;

p_{M}: about one and half oscillation cycles from the end of p_{F};

p_{L}: the rest of the pulse from the end of p_{M}.

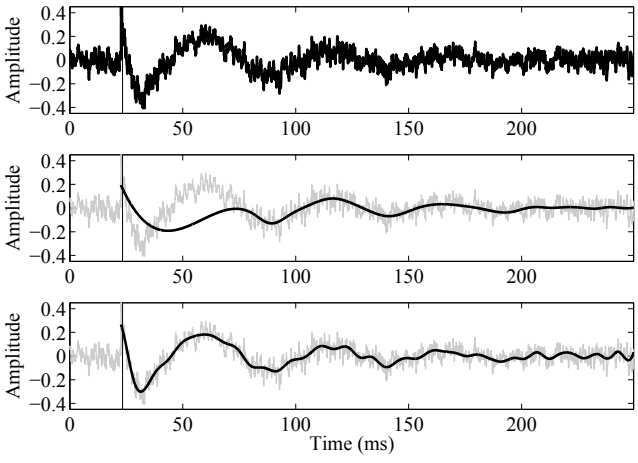

Figure 2: *Top: Signal corrupted with a long pulse. The thin vertical line at about 25 ms indicates the beginning of the pulse. Middle: corrupted signal (thin faded gray line) and residue after extracting the first 7 IMFs (thick black line). Bottom: corrupted signal (thin faded gray line) and the sum of the residue with the 7th IMF (thick black line).*

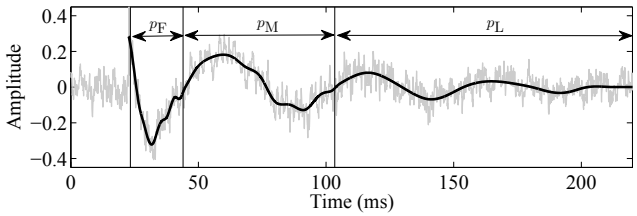

Figure 3: *Signal corrupted with a long pulse (thin faded gray line) and composite pulse estimate after EMD with 7 IMFs (thick black line). The thin vertical lines delimit the partitions p_F, p_M, and p_L. The residue is attributed to p_L. The sum of the residue and the 7th IMF is attributed to p_M. The sum of the residue with the 7th and 6th IMFs is attributed to p_F.*

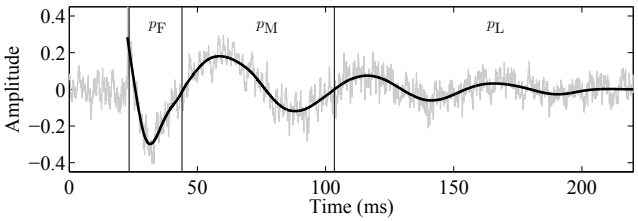

Figure 4: *Signal corrupted with a long pulse (thin faded gray line) and composite pulse estimate after applying the piece-wise polynomial smoothing to the estimate shown in Figure 3.*

The residue of the EMD analysis can be assigned to p_{L}. Since the maximum number of IMFs can be forcefully limited, the residue is not a monotonic function and could be further decomposed into more IMFs. However, the idea here is to set the maximum number of IMFs so as to produce a residue that, besides being smooth, captures well the oscillations present in the pulse tail p_{L}. Continuing the decomposition process until obtaining a monotonic residue is then unnecessary and would undesirably increase the computational costs involved.

To the portion p_{M} one can assign the sum of the residue and the last IMF observed in that region. In a similar fashion, to the portion p_{F} one can assign the sum of the residue and the two last IMFs observed in that region. In practice, the partitions p_{F}, p_{M}, and p_{L} must overlap a bit in time. This way, it is possible to merge their waveforms together seamlessly via straightforward cross-fading schemes. An example of the proposed partition and assignment scheme is seen in Figure 3, where the cross-fading among adjacent partitions lasts about 4.5 ms.

As reported in [18], EMD of white noise tends to produce IMFs that could be considered as sub-band signals of a dyadic octave filterbank analysis with poor selectivity channels. Thus, the higher the IMF number the lowest the mean frequency of its power spectral density. Assuming this behavior holds for spectrally rich audio signals, one may speculate that the estimates assigned p_{L}, p_{M}, and p_{F} would consist of lowpass versions of the input signal with progressively increasing cutoff frequencies.

From the above discussion, an advantage of the EMD is that the resulting IMFs are pulse estimates with different frequency bandwidths that can be readily combined in the partition and assignment scheme. However, from Figure 3 one perceives that, especially in regions p_{F} and p_{M}, part of the low-frequency content of the signal is present in the pulse estimate.

A practical solution to gain more control over the smoothness of the pulse estimate is to post-process the partial pulse estimates obtained in each partition prior to their merging. An effective post-processing is the piece-wise polynomial fitting described in [6]. In brief terms, this signal smoothing scheme consists of fitting a low-order polynomial to short-duration frames of the signal of interest, in an overlap-and-add signal segmentation, e.g., Hanning windows with 50% temporal superposition. If the polynomial order is fixed to 2, as adopted in the conducted experiments, the degree of smoothness is solely controlled by the length of the overlapping windows.

Here, a different window length can be chosen for each partition p_{L}, p_{M}, and p_{F} . A rule of thumb is to set the window length to some value (in units of time or number of samples) between a quarter and a half of the smallest period of the pulse oscillation observed in the considered partition. The pulse estimate that results from applying the piece-wise polynomial fitting is seen in Figure 4, where windows of sizes 10 ms, 20 ms, and 30 ms have been used, respectively, to smooth out the pulse estimates in partitions p_{F}, p_{M}, and p_{L}. In order to avoid boundary effects during the smoothing procedure, the bumpy pulse estimate in partition p_{F} was also extended backward by about 10 ms via odd symmetry reflection w.r.t. its first sample.

Once an adequate pulse estimate is obtained, de-thumping is simply achieved by subtracting the pulse from the signal. Removal of the initial click can be easily accomplished via standard model-based de-clicking techniques [2]. In practice, it may be desirable to artificially overestimate the click duration toward the beginning of the long pulse.

At this stage it seems appropriate to comment on the strong and weak points of the EMD-based pulse estimation. An obvious weakness lies in its inability to obtain a pulse estimate at once as a single IMF. Furthermore, the user is left with the task of choosing several additional parameters for the post-processing stage. This burden, however, can be alleviated through a graphical user interface, similar to that designed and proposed in [6]. On the other hand, the EMD can be seen as a computationally cheap way of obtaining lowpass and bandpass filtered versions of the input signal.

3.2. CEEMD-based Estimation of Single Pulses

CEEMD-based estimation of single pulses follows the first three processing steps defined in the beginning of Section 3.1, the latter carried out with the CEEMD instead of the EMD. The fourth step, however, turns out to be unnecessary, as it will be demonstrated.

Besides the number of IMFs, signal analysis via CEEMD requires the choice of two other parameters: the standard deviation of the additive noise realizations and the number of their pairs. Fortunately, in the context of long pulse estimation tackled here, these two parameters have minor impact to the final results. For all simulations presented in this paper involving CEEMD, the standard deviation of the additive noise has been set to 20% of that of the input signal, whereas the number of noise realization pairs was set to 4, mainly to reduce computational costs.

Once the maximum number of IMFs is defined, the pulse estimate is simply taken as the ensuing residue, after discarding the initial samples that are due to the artificial signal extension. As before, this residue is not a monotonic function and could be further decomposed into more IMFs.

At this point it is worth mentioning that the maximum number of IMFs required for the CEEMD to produce a smooth pulse estimate is about twice as that of EMD. This slower convergence to a target-residue may come from a much higher number of signal extrema in the beginning of the decomposition, due to the addition of noise to the input signal.

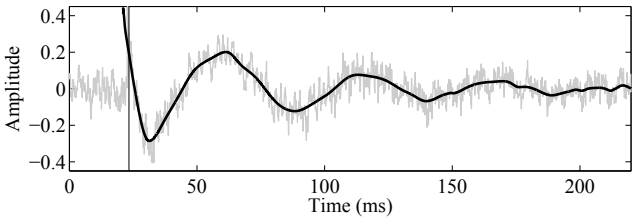

Figure 5: *Signal corrupted with a long pulse (thin faded gray line) and pulse estimate (thick black line) as the CEEMD residue after extracting the first 12 IMFs. The thin vertical line at about 23 ms indicates the actual beginning of the pulse. Pulse samples before that limit should be discarded.*

Figure 5 displays an example of pulse estimate obtained via the proposed CEEMD-based method, where the pulse is taken as the residue after extracting 12 IMFs. Here, for illustration purposes, the pulse is plotted including the samples related to a backward signal extension of about 2.5 ms. As can be seen, the attained pulse estimate exhibits an adequate level of smoothness and is capable of following the pulse trajectory in both fast and slow oscillation regions.

In comparison with the pulse estimate shown in Figure 4, the one yielded by the CEEMD solution is a bit bumpier. However, since the amplitudes of these faster oscillations are quite small, their subtraction from the corrupted signal is bound to produce inaudible effects. Therefore, post-processing for further smoothing and windowing schemes for pulse composition are no longer needed.

3.2.1. Speeding up the CEEMD-Based Pulse Estimation

As seen in the previous section, the CEEMD-based pulse estimation is effective, yet simpler than the EMD-based counterpart, from the algorithm implementation and calibration perspectives. Its main drawback is a higher computational cost that slows down the process of obtaining the desired pulse estimates.

In the context of EMD of white noise, findings reported in [18] suggest that the number of IMF zero-crossings, which holds relation with the number of IMF extrema, tends to decrease on average by half from a given IMF to the subsequent one. Thus, the larger the number of extrema in the beginning, the longer the decomposition takes to converge to a monotonic residue.

With the previous information in mind, the following modification, which affects the computation of signal envelopes within the EEMD processing chain, has been found operative to speed up the CEEMD-based pulse estimation method:

1. Detect all peaks and valleys of the input signal as usual;
2. Select from the previous set of peaks and valleys only the peak (valley) with maximum (minimum) amplitude inside juxtaposed observation windows of 3.4 ms (about 150 samples at 44.1 kHz sampling rate);
3. Obtain the signal envelopes as usual, but using only the peaks and valleys selected in step 2;
4. Apply item 2 only for extraction of the first two IMFs.

The selection performed in step 2 above is an attempt to retain only the most prominent peaks and valleys of the input signal, which is artificially corrupted with noise in the CEEMD. The proposed peak and valley pruning produces upper and lower envelopes are way smoother than those of the noisy input signal. As a consequence, the frequency bandwidths of the first two IMFs are larger than those of the corresponding IMFs computed via conventional CEEMD, forcing the modified CEEMD iterations to converge faster to slowly varying IMFs, which are of interest to the present application.

Together with the maximum number of IMFs, the length of the observation window in step 2 above can also be changed by the user as a means to control the degree of smoothness of the pulse estimate. Considering a practical range from 1 to 10 ms, the longer the window length, the smaller the maximum number of extracted IMFs required for the residue to become an adequate pulse estimate.

To illustrate the combined role of the two previously discussed parameters in the final CEEMD-based pulse estimate, outcomes of two different yet equally effective configurations of the CEEMD method are presented in Figure 6 and Figure 7. In connection with these results, Table 1 summarizes the processing parameters adopted in each configuration and the average savings in computational time w.r.t the conventional CEEMD-based pulse estimation.

Visual assessment among the plots shown in Figures 5 to 7 reveals similar pulse estimates. Hence, from Table 1, adoption of the proposed peak (valley) picking for envelope computation within CEEMD is advantageous, for it can produce up to a tenfold reduction in computational time.

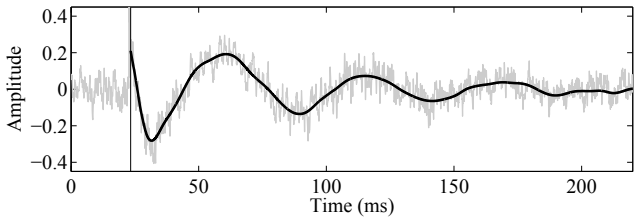

Figure 6: *Signal corrupted with a long pulse (thin faded gray line) and pulse estimate (thick black line) as the CEEMD residue after extracting 4 IMFs. Peak and valley selection was carried out within juxtaposed windows of 3.4 ms.*

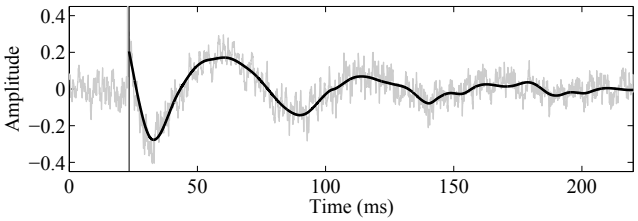

Figure 7: *Signal corrupted with a long pulse (thin faded gray line) and pulse estimate (thick black line) as the CEEMD residue after extracting only 2 IMFs. Peak and valley selection was carried out within juxtaposed windows of 6.8 ms.*

Table 1: *Configuration of the CEEMD-based pulse estimation and corresponding results. The value of T is about 10 s when running the CEEMD analysis on a Core i7 870 2.93 GHz Quad Core CPU. All simulations ran on the same machine.*

CEEMD Configuration		Results	
No. IMFs	Window Size	Avg. Proc. Time	Visual Output
13	–	T	Figure 5
4	3.4 ms	T/3.8	Figure 6
2	6.8 ms	T/10	Figure 7

3.2.2. A Real-World Example of CEEMD-based De-Thumping

As a real-world example, one of the long pulse occurrences in the signal available from [21] has been subjected to the proposed CEEMD-based de-thumping. The signal in question, which is sampled at 22.05 kHz, contains pulses with initial oscillations of about 180 Hz, thus faster than in the previously considered pulse. In order for the CEEMD method to capture those fast pulse variations, the window size in the peak/valley selection scheme has been experimentally set to 1.4 ms, whereas the maximum number of IMFs was limited to 2.

The attained pulse estimate is seen in the top panel of Figure 8, where one can notice undesirable fast oscillations after about 40 ms of the beginning of the pulse. To improve the estimate, they were flattened out via the overlap-and-add polynomial smoothing [6] with windows of 13.6 ms. The final pulse estimate, which is shown in the bottom panel of Figure 8, is then composed by seamlessly merging the first approximately 16 ms of the original CEEMD-based estimate with its smoothed out version from about 40 ms onward.

Concerning the corrupted signal and related pulse estimate depicted in the bottom panel of Figure 8, the corresponding de-thumped version is seen in the top plot of Figure 9. The remaining click of about 680 μs (about 15 samples at 22.05 kHz sampling rate) was suppressed by LSAR signal reconstruction with model order 65 to produce the signal shown in the middle plot of Figure 9. A detailed vision of the signal reconstruction around the click location is seen in the bottom plot.

3.3. Estimation of Superimposed Pulses

As regards estimation of superimposed pulses, a strategy that has been found effective was to treat independently the parts that form the pulse. In other words, the signal part that follows the last driving initial click is subjected for instance to the CEEMD-based pulse estimator as if it were a single pulse occurrence. Separately, pulse estimation in the intermediate part between two consecutive initial clicks is carried out using the same processing parameters. For this part, it is desirable to extent the signal backward and forward in time for about the duration of the delimiting clicks.

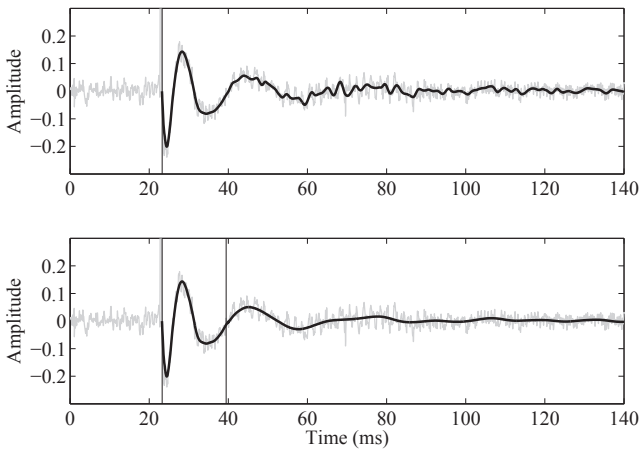

Figure 8: *Top: real-world example of a signal corrupted with a long pulse [21] (faded gray line) and corresponding CEEMD pulse estimate (thick black line). Bottom: same corrupted signal (faded gray line) and improved estimate via polynomial smoothing from 40 ms onward (thick black line).*

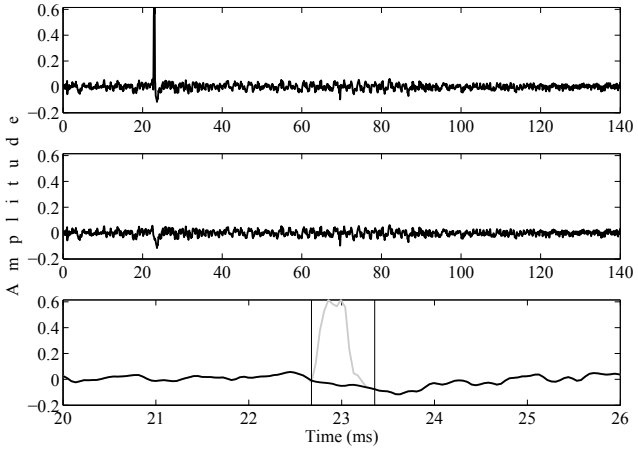

Figure 9: *Top: de-thumped signal related to the corrupted signal and related pulse estimate shown in the bottom plot of Figure 8. Middle: de-clicked signal via LSAR signal reconstruction. Bottom: detail of the signal reconstruction around the click location, with the original click painted in faded gray line. The thin vertical lines around 23 ms delimit the audio region subjected to LSAR interpolation.*

Figure 10 shows an example of CEEMD-based estimation of superimposed pulses in which the same processing setup that generated the result seen in Figure 6 has been used.

4. PERFORMANCE ASSESSMENT

In this paper the performance assessment methodology for audio de-thumping methods defined in [6] is employed. The same set of test signals and one of the quantitative metrics considered in [6] are also used as a means to allow direct comparisons with those previous results. A brief overview of the experimental setup is given in the sequel. The reader is referred to [6] for a more detailed description.

4.1. Test Signals

The test signals comprise a set of reference uncorrupted signals and a corresponding set of artificially corrupted versions. The reference set is composed of 6 CD-quality short-duration (11 to 20 s) excerpts of audio including diverse musical genres such as pop, jazz, classic, and ethnic, as well as solo of drums and acoustic bass.

The corrupted set was produced by adding several single long pulses (with initial click) to the reference signals. Successive pulses were placed approximately 769 ms apart from each other. Here, only the set of strong pulses [6, 22] will be considered for performance evaluation.

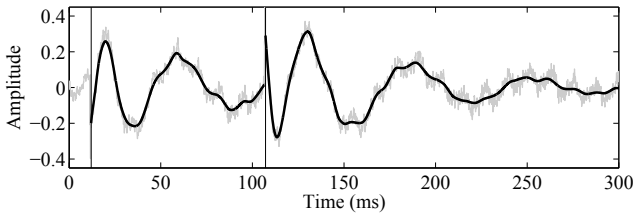

Figure 10: *Signal corrupted with a superimposed long pulse (thin faded gray line) and CEEMD-based pulse estimates (thick black lines).*

4.2. Experimental Setup and Performance Metric

As in [6], the evaluation methodology adopted here consists in first obtaining restored versions (de-thumped and de-clicked) of the corrupted set via a selection of de-thumping methods with predefined configurations. Then, the reference and restored sets are compared by objective means.

Here, the Perceptual Audio Quality Measure (PAQM) [11] is used as the performance metric. In a nutshell, the PAQM compares a processed signal w.r.t. a reference and outputs a dissimilarity index that takes into account several properties of the human auditory system, such as masking in time and frequency. The closer to zero the PAQM, the more similar perceptually are the processed and reference signals.

The aim here is to run a direct comparison among the PAQM values associated with the restored signals produced by the following de-thumping methods: AR Separation (ARS), TPSW-based (TPSW), EMD-based pulse estimation (EMD), and CEEMD-based pulse estimation (CEEMD). Therefore, only the processing setup of the EMD and the CEEMD will be defined. The processing parameters employed in the ARS and the TPSW can be found in [6].

4.2.1. EMD Configuration

The following configuration was employed to obtain the EMD-related results:

- Backward signal extension: 100 samples with odd symmetry w.r.t. the first sample;
- Maximum number of IMFs: 6;
- Envelope computation: piecewise cubic Hermite interpolating polynomials across local maxima (minima);
- p_{F}: 44 ms from the beginning of the pulse;
- p_{M}: 60 ms from the end of p_{F};
- p_{L}: 100 ms from the end of p_{M};
- Pulse estimate inside p_{F}: (residue + 6th IMF + 5th IMF) smoothed out via the overlap-and-add polynomial scheme with windows of 10 ms and 2nd-order polynomials;
- Pulse estimate inside p_{M}: (residue + 6th IMF) smoothed out via the overlap-and-add polynomial scheme with windows of 20 ms and 2nd-order polynomials;
- Pulse estimate inside p_{L}: (residue) smoothed out via the overlap-and-add polynomial scheme with windows of 30 ms and 2nd-order polynomials;
- Click removal: 75th-order LSAR signal reconstruction of 50 samples from the click onset.

4.2.2. CEEMD Configuration

The following configuration was employed to obtain the CEEMD-related results:

- Number of noise realization pairs: $N/2 = 4$;
- Standard deviation of each realization: $\varepsilon = 0.2\,\mathrm{std}\{x(t)\}$;
- Backward signal extension: 100 samples with odd symmetry w.r.t. the first sample;
- Maximum number of IMFs: 4;
- Envelope computation (for first two IMFs): piecewise cubic Hermite interpolating polynomials across local maxima (minima), after selection of one maximum (minimum) per juxtaposed windows of 3.4 ms;
- Envelope computation (third and fourth IMFs): piecewise cubic Hermite interpolating polynomials across local maxima (minima);
- Click removal: 75th-order LSAR signal reconstruction of 50 samples from the click onset.

4.3. Results and Discussion

Table 2 summarizes the attained PAQM of the restored (de-thumped and de-clicked) test signals for each of the considered methods and predefined configurations given in the previous sections and [6].

Analysis of the results suggests a tendency of the EMD to outperform the competing methods. For instance, for all test signals restored by EMD, the associated PAQMs are smaller than those of TPSW. ARS offers the most effective restoration for signals *Drums* and *Singing*, whereas CEEMD is the least successful in these cases. Signal *Ethnic* as restored by CEEMD yields the smallest PAQM. It is worth mentioning that too small PAQM differences may not necessarily imply a noticeable perceptual difference. Audio examples can be found in [23].

In summary, for the considered experimental scenario, the attained PAQM results suggest EMD and CEEMD are effective and competitive tools for audio de-thumping.

Table 2: Comparative performance evaluation of the EMD, CEEMD, TPSW, and ARS de-thumping methods using PAQM. The best results (lowest PAQM) are highlighted.

Test Signal	EMD	CEEMD	TPSW	ARS
Pop	**0.024**	0.028	0.036	0.046
Jazz	**0.012**	0.031	0.028	0.061
Classic	**0.010**	0.023	0.017	0.033
Ethnic	0.032	**0.030**	0.051	0.048
Drums	0.029	0.097	0.049	**0.014**
Bass	**0.015**	0.030	0.031	0.188
Singing	0.092	0.282	0.110	**0.066**

5. CONCLUSIONS

This paper addressed the problem of long pulse removal from audio signals (de-thumping). Two methods for long pulse estimation based on Huang's Empirical Mode Decomposition (EMD) were proposed. After an overview of the EMD and its related complementary ensemble version (CEEMD), their use as tools to obtain adequate pulse estimates from corrupted audio signals was investigated by means of practical examples.

It was found out experimentally that, possibly due to mode mixing issues afflicting signal analysis via EMD, an intrinsic mode function (IMF) or a non-monotonic residue (after successively extracting a given number IMFs from the original signal) may not serve alone as an adequate pulse estimate. To overcome the problem, the adopted solution [6] was to define three temporal partitions to which pulse estimates with different frequency bandwidths or levels of smoothness were attributed. The final composite pulse estimate was then assembled together by merging seamlessly the estimates from each partition.

As regards signal analysis via the CEEMD, it was discovered that adequate pulse estimates could be obtained at once as a non-monotonic residue after successively extracting a given number IMFs from the original signal. However, as compared with the EMD, about twice the number of initial IMFs needed to be extracted for producing a smooth enough pulse estimate. Other CEEMD parameters such as the variance of the additive noise and the number of ensemble pairs had negligible impact on the final results. As a means to decrease the computational cost of the CEEMD for the studied application, a modified scheme for signal envelope computation was devised: piecewise cubic Hermite interpolating polynomials were fitted across local maxima (minima), after selection of one maximum (minimum) per juxtaposed short-duration windows. Average reductions in computation time up to ten times were reported.

Objective performance evaluation of the proposed EMD- and CEEMD-based methods for audio de-thumping was carried out using the same methodology and test data of [6]. Indirect comparative results, in terms of the Perceptual Audio Quality Measure [11] of the restored versions of the corrupted data, suggest the proposed CEEMD-based method tends to perform as effectively as the competing TPSW-based procedure [6] and outperform the AR separation method [2]. As regards the EMD-based method the observed tendency is of a more favorably performance in comparison with the TPSW-based solution.

6. REFERENCES

[1] S. V. Vaseghi, *Algorithms for Restoration of Archived Gramophone Recordings*, Ph.D. thesis, Cambridge Univ., UK, 1988.

[2] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration — A Statistical Model Based Approach*, Springer-Verlag, London, UK, 1998.

[3] S. V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings," *J. Audio Eng. Soc.*, vol. 40, no. 10, pp. 791–801, Oct. 1992.

[4] S. J. Godsill, *The Restoration of Degraded Audio Signals*, Ph.D. thesis, Cambridge Univ., UK, 1993.

[5] S. J. Godsill and C. H. Tan, "Removal of Low Frequency Transient Noise from Old Recordings Using Model-Based Signal Separation Techniques," in *Proc. IEEE WASPAA*, 1997.

[6] P. A. A. Esquef, L. W. P. Biscainho, and V. Välimäki, "An Efficient Algorithm for the Restoration of Audio Signals Corrupted with Low-Frequency Pulses," *J. Audio Eng. Soc.*, vol. 51, no. 6, pp. 502–517, June 2003.

[7] W. A. Struzinski and E. D. Lowe, "A Performance Comparison of Four Noise Background Normalization Schemes Proposed for Signal Detection Systems," *J. Acoust. Soc. Am.*, vol. 76, no. 6, pp. 1738–1742, Dec. 1984.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, E. H. Shih, Q. Zheng, C. C. Tung, and H. H. Liu, "The Empirical Mode Decomposition Method and the Hilbert Spectrum for Nonstationary Time Series Analysis," in *Proc. Roy. Soc. London*, 1998, vol. 454A, pp. 903–995.

[9] N. E. Huang and Z. Wu, "An Adaptive Data Analysis Method for Nonlinear and Nonstationary Time Series: the Empirical Mode Decomposition and Hilbert Spectral Analysis," in *Proc. 4th Int. Conf. Wavelet Analysis and Its Applications*, China, 2005.

[10] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[11] J. G. Beerends and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation," *J. Audio Eng. Soc.*, vol. 40, no. 12, pp. 963–978, Dec. 1992.

[12] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. of IEEE – Part III: Radio and Communication Engineering,*, vol. 93, no. 26, pp. 429–441, 1946.

[13] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: The Hilbert Spectrum 1," *Annual Review of Fluid Mechanics*, vol. 31, no. 1, pp. 417–457, 1999.

[14] N. E. Huang, M. L. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proc. R. Soc. of Lond. A*, vol. 459, pp. 2317–2345, 2003.

[15] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," in *Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, 2003.

[16] Q. Chen, N. Huang, S. Riemenschneider, and Y. Xu, "A B-spline approach for empirical mode decompositions," *Advances in Computational Mathematics*, vol. 24, no. 1, pp. 171–195, 2006.

[17] M. Dätig and T. Schlurmann, "Performance and Limitations of the Hilbert-Huang Transformation (HHT) with an Application to Irregular Water Waves," *Ocean Engineering*, vol. 31, no. 14, pp. 1783–1834, Oct. 2004.

[18] P. Flandrin, G. Rilling, and P. Gonçalves, "Empirical Mode Decomposition as a Filter Bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, Feb. 2004.

[19] P. A. A. Esquef and L. W. P. Biscainho, "An Efficient Model-Based Multirate Method for Reconstruction of Audio Signals Across Long Gaps," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1391–1400, July 2006.

[20] G. Rilling and P. Flandrin, "One or Two frequencies? The Empirical Mode Decomposition Answers," *IEEE Trans. Signal Processing*, vol. 56, no. 1, pp. 85–95, 2008.

[21] `http://dea.brunel.ac.uk/cmsp/Home_Saeed_Vaseghi/Home.html`.

[22] `http://www.acoustics.hut.fi/publications/papers/jaes-LP/`.

[23] `www.lncc.br/~pesquef/dafx11/`.