Energy-Based Effective Length of the Impulse Response of a Recursive Filter

Timo I. Laakso, Senior Member, IEEE, and Vesa Valimäki, Member, IEEE

Abstract—A measure for the effective length of the impulse response of a stable recursive digital filter based on accumulated energy is proposed. The new measure finds applications in several fields of digital signal processing, including estimation of the extent of attack transients for filters with dynamically varying inputs, elimination of transients in variable recursive filters, and design and implementation of linear-phase IIR systems. A general definition and a simple algorithm to evaluate it are introduced, and closed-form expressions are derived for first and second-order all-pole filters. The effect of zeros on the effective length is analyzed. An upper bound for the effective length of higher-order filters is derived using results for low-order filters, which is illustrated for classical digital lowpass filters. The use of the measure is demonstrated with examples of implementation of linear-phase IIR systems and estimation of transients in variable IIR filters.

Index Terms—All-pole filters, effective length of impulse response, IIR filters, recursive digital filters, transients in digital filters, waveform analysis.

I. INTRODUCTION

THE IMPULSE response of a stable recursive digital filter is infinitely long in principle, but due to exponential decay it eventually sinks below the quantization step or the noise in the system. Thus, for all practical purposes the impulse response of a stable recursive filter can be regarded as finite.

In several applications, a practical measure for the length of the impulse response of an IIR filter is needed. For example, when a finite-length signal is processed off-line with a recursive filter, the beginning of the resulting signal is disturbed by the attack transient of the filter and it is desirable to know how many samples are corrupted [1]. The length and the waveform of the transient naturally depend on the input signal, but it can be characterized by the filter’s impulse response or its effective length (EL). Furthermore, in any fixed-coefficient recursive filter that is used for processing a continuous flow of data, it is often crucial that the transient or “ringing” due to an abrupt change in the input signal (e.g., a step input) is limited. This ringing can also be quantified by the effective length of the impulse response.

Another application is found in inverse filtering, i.e., when it is desired to equalize linear distortion with the corresponding inverse filter. When the original filter is an FIR filter, the inverse filter is a linear-phase all-pole filter. The effective length of this inverse filter may be of interest if there is a finite length of data to be processed or if the inverse filter needs to be approximated by an FIR filter, which implies truncation of the impulse response of the inverse filter [2], [3].

A communication channel is characterized by its impulse response, and its length is a critical parameter. A technique for reducing the effective length of the channel impulse response in discrete multitone (DMT) transceivers using a pre-equalizer was proposed in [4].

Time-varying or tunable filters are filters whose coefficients are changed on-line to alter the filter properties, such as the magnitude response [5], [6] or the group delay [7]. When changing the coefficients of a recursive filter, transients will occur [8]–[11]. It is often essential to make changes in small enough steps so that the resulting transients stay within prescribed limits. Again, these transients also depend on the filter input, but an impulse-response-based measure can be used to guarantee satisfactory performance in the general case.

A special case of the transient problem is encountered when the transients are eliminated using the novel technique proposed in [12]–[15]. The resulting transient can be canceled within desired accuracy by updating the state variables of the filter. This accuracy depends on the effective length of the impulse response of the filter after the change of coefficients.

Previously, at least three different amplitude-based methods have been used for measuring the effective length of an infinitely long but decaying impulse response.

1) In [16], a general duration $d$ (or width) of a signal was defined via the expression

$$d^2 = \frac{1}{E} \sum_{n=-\infty}^{\infty} n^2 |x(n)|^2$$

where $E = \sum_{n=-\infty}^{\infty} |x(n)|^2$ is the total energy of the signal. (This is actually a discretized version of Papoulis’s continuous-time definition.) The duration $d$ is analogous to the concept of standard deviation of a probability distribution and gives a rough estimate of the length of the time-domain signal.

2) A traditional technique is based on the concept of a time constant. Typically, the time constant of the pole with
the largest radius $\tau_{\text{max}}$ is used for estimating the decay rate of the impulse response and an amplitude threshold is chosen to determine the effective length [17]. Smith has proposed to approximate this time constant as $1/(1-\tau_{\text{max}})$ which is obtained by truncating the Taylor series of the exact equation for the time constant [14], [18]. Based on merely one pole of the system, this measure is easy to use but only gives a crude estimate for the length of the impulse response.

3) Finally, an amplitude threshold (e.g., 1% of the maximum amplitude) can be set and the effective length can be determined as the sample index where the impulse response ultimately goes below this threshold [12], [13]. In principle, this technique gives a better approximation for a given recursive filter of arbitrary order. The drawbacks are the lack of analytical methods and the complication of the measure when the impulse response does not decay monotonically, which is the case in general. One cannot be sure how large sample values will be encountered later.

Still another application for the effective length of an infinite impulse response is a realization technique for linear-phase IIR filters based on cascading a minimum-phase IIR filter $H(z)$ and its maximum-phase (unstable or noncausal) counterpart $H(z^{-1})$, resulting in exactly linear phase in theory [19]–[24]. For a causal and finite-delay implementation, the filtering must be based on processing the input signal in finite-length blocks of $L$ samples. Details of different approaches vary, but the basic constraint is to choose the block length $L$ so that the impulse response of $H(z)$ has decayed to a small enough level. On the other hand, $L$ should be chosen as small as possible to minimize the latency (processing delay) of the technique. Although the block length $L$ is an essential system parameter, the techniques to determine its value proposed in the referred papers are rather heuristic and do not attempt to define or find an optimal value. In [19], it was suggested that the filter be implemented in parallel form employing second-order filter sections and using a rough time-constant-based measure for the length of the impulse response of each section. An upper bound for estimating the resulting errors for a given $L$ was derived in [20] and in [23] but no explicit measure for defining or determining $L$ was given.

From the above, it is apparent that several ways to measure the effective length of the impulse response of recursive filters have been suggested before, but none of them seems to have gained wide acceptance. In this paper we want to introduce a meaningful yet simple and practical definition. We define the effective length of the impulse response of a general recursive filter based on the accumulated percentage of the total energy. This concept has several advantages.

1) In many applications, the energy or power of an additive disturbance that is sufficiently uncorrelated with the signal itself is a natural measure. It can be used, e.g., to (roughly) estimate auditive effects of a disturbing signal or its impact on the bit error rate in a data transmission system. Hence, the definition of the effective length of the impulse response based on how much of the total energy has been received (or, equivalently, how much is still to come) yields more useful information about the system than, e.g., knowledge of the instantaneous amplitude.

2) The total energy of a given filter is easy to determine either in the time domain (using the impulse response) or in the frequency domain (using the frequency response), thanks to Parseval’s theorem. Fast algorithms have been proposed in [25]–[27]; and closed-form expressions are known for low-order filters [25], [28].

3) The measure is parametric and thus flexible. The length of an impulse response corresponding to an arbitrary percentage of total energy can be defined, and the length of any stable recursive filter can be determined with a simple recursive algorithm. Furthermore, efficient closed-form expressions can be derived for low-order all-pole filters. These can not only be used for estimation of the total impulse response of high-order filters, but they also yield valuable insight into the effect of poles on the length of the impulse response.

The paper is organized as follows. Section II introduces the general energy-based measure and a simple algorithm for estimating the effective length of a general recursive filter. In Section III, closed-form formulas are derived for first and second-order all-pole filter sections. The effect of zeros on the effective length of the impulse response is discussed in Section IV while higher-order filters are tackled in Section V, with examples of estimating the length of classical digital lowpass filters using closed-form formulas for first and second-order filters. Section VI illustrates the use of the measure in the implementation of linear-phase IIR filters and in elimination of transients in variable recursive filters, and compares results to previous work. Finally, conclusions are drawn in Section VII.

II. EFFECTIVE LENGTH OF A GENERAL RECURSIVE FILTER

A. Definitions

Let us consider a general $N$th-order recursive filter with the $z$ transfer function

$$ H(z) = \sum_{k=0}^{N} b_k z^{-k} \over 1 + \sum_{k=1}^{N} a_k z^{-k} \quad (2) $$

where the filter coefficients $a_k$ and $b_k$ are real-valued. The numerator or the denominator may be of order lower than $N$, i.e., some of the numerator or denominator coefficients with the largest indices may be zero. Special cases are nonrecursive (FIR) filters for which all $a_k$ are zero (which are not of great interest here since their impulse responses are of finite length anyway) and all-pole filters for which all $b_k$ except one are zero. Assuming a stable and causal implementation, the recursive filter (2) can also be described via an equivalent difference equation as

$$ y(n) = \sum_{k=0}^{N} b_k x(n-k) - \sum_{k=1}^{N} a_k y(n-k) \quad (3) $$

where $x(n)$ and $y(n)$ are the input and output of the filter, respectively. When the input signal is a unit impulse $x(n) =$
TABLE I
ALGORITHM FOR COMPUTING THE EFFECTIVE LENGTH OF A GENERAL RECURSIVE FILTER

---

Step 0: Compute $E$ and $E_P$ for the chosen $P$. Initialize: $x(n) = \delta(n)$,

\[ h(-1) = h(-2) = \ldots = h(-N) = 0, \quad E_N = 0, \quad n = 0. \]

Step 1: \( h(n) = \sum_{k=0}^{N} h_k \delta(n-k) - \sum_{k=1}^{N} a_k h(n-k) \)

Step 2: \( E_N = E_N + |h(n)|^2 \)

Step 3: If \( E_N \geq E_P = \frac{P}{100} E \) then $N_{\text{EL}} = n$ and stop;
else $n = n + 1$; go to Step 1.

---

\( \delta(n) \) which equals unity at \( n = 0 \) and zero elsewhere, the output \( y(n) = h(n) \) is the impulse response of the filter. The impulse response can also be obtained formally via the inverse $z$ transform from (2).

The total energy of the causal impulse response $h(n)$ is defined as

\[
E = \sum_{n=0}^{\infty} h^2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega
\]

\[ = \frac{1}{2\pi j} \int H(z)H(z^{-1})z^{-1} dz \quad (4) \]

where the frequency-domain expression follows from the Parseval relation. The determination of the integral in the $z$-domain has been addressed in [25]–[27]. For low-order filters, the closed-form expressions of [25] and [28] can be used.

We define the energy-based effective length (EL) of the impulse response as the smallest nonnegative integer time index $N_P$ by which at least $P\%$ of the total energy of the impulse response has arrived. The corresponding accumulated energy $E_A(N_P)$ can be expressed as

\[
E_A(N_P) = \sum_{n=0}^{N_P} h^2(n) \geq E_P = \frac{P}{100} E. \quad (5)
\]

Hence, we always require $E_A(N_P) \geq E_P$ since the effective length $N_P$ must be an integer. Note that this differs slightly from the usual definition of length of the corresponding FIR filter: the truncated part contains $N_P + 1$ samples but the effective length (5) is one less, $N_P$. The energy-based length (for any percentage) of a filter having an impulse response $h(n) = \delta(n)$ is thus zero, and that of a two-point averager $h(n) = \frac{1}{2}[\delta(n) + \delta(n-1)]$ is unity, which is in accordance with common sense.

B. General Algorithm

The most straightforward way to compute the impulse response of a given causal and stable recursive filter is to use the difference equation (3). When the total energy $E$ is pre-computed, the corresponding accumulated energy $E_A(N_P) \geq E_P$ for the chosen percentage $P$ can be determined recursively via the algorithm presented in Table I.

The algorithm is simple and can be used for many recursive filters having an impulse response length of some tens of samples. However, for very narrowband filters the length can be several hundreds of samples. For first and second-order all-pole filters more practical closed-form expressions can be derived, as will be shown in the following. These results also give valuable insight into the effect of the pole parameters on the EL of the filter.

III. EFFECTIVE LENGTH OF LOW-ORDER ALL-POLE FILTERS

A. First-Order All-Pole Filter

Consider a first-order all-pole filter with the transfer function

\[
H(z) = \frac{1}{1 - az^{-1}} \quad (6)
\]

where $a$ is real-valued and the pole radius $|a| = r < 1$ for stability. Its causal impulse response is simply $h(n) = a^n$ for nonnegative $n$. The accumulated energy $E_A(N_P)$ can be expressed as

\[
E_A(N_P) = \sum_{n=0}^{N_P} (a^2)^n = \frac{1 - (r^2)^{N_P+1}}{1 - r^2} \quad (7a)
\]

from which the total energy is also obtained as a limit ($N_P \to \infty$) as $E = 1/(1 - r^2)$. The requirement (5) now becomes

\[
E_A(N_P) = \frac{1 - (r^2)^{N_P+1}}{1 - r^2} \geq \frac{P}{100} \frac{1}{1 - r^2} = \frac{P}{100} E \quad (7b)
\]

and the EL can be solved as

\[
N_P = \left\lceil \frac{\log(1 - P/100)}{\log(r^2)} - 1 \right\rceil \quad (8)
\]

where the logarithm can have any (positive) base and $\lceil \cdot \rceil$ denotes the ceiling operation (i.e., rounding upwards). Note that quantization is necessary because $N_P$ must be an integer.

Fig. 1 presents the EL $N_P$ for $P = 90\%$, 95\%, and 99\% as a function of the pole radius $r$ computed according to (8). These curves show the expected phenomenon that the EL of the impulse response increases rapidly as the pole radius $r$
approaches the value 1. Furthermore, it is seen that the EL is fairly insensitive to the percentage value so that the lengths corresponding to 90%–99% energy do not differ much except for when the pole radius \( r \) is larger than 0.9.

### B. Second-Order All-Pole Filter: Complex-Conjugate Poles

A second-order all-pole transfer function with complex-conjugate poles can be expressed as

\[
H(z) = \frac{1}{1 - 2r \cos(\theta)z^{-1} + r^2 z^{-2}}
\]

(9)

with the poles \( z = re^{\pm j\theta} \) having radius \( 0 < r < 1 \) and angle \( 0 < \theta < \pi \). The impulse response of the filter can be given as

\[
h(n) = \begin{cases} 0, & n < 0 \\ \frac{r^n}{\sin\theta} \sin[(n + 1)\theta], & n \geq 0. \end{cases}
\]

(10)

The total energy of this impulse response is known to be [25], [28]

\[
E = \frac{1 + r^2}{(1 - r^2)(1 + r^4 - 2r^2 \cos(2\theta))}.
\]

(11)

For the chosen percentage \( P \) of the total energy \( E \), the nominal residual energy is simply

\[
\bar{E}_P = E(100 - P)/100.
\]

(12)

On the other hand, the instantaneous residual energy is the complement of the accumulated energy and can be expressed using (10) as

\[
\bar{E}_A(N_P) = E - E_A(N_P) = \frac{1}{\sin^2\theta} \sum_{n=N_P+1}^{\infty} r^{2n} \sin^2[(n + 1)\theta].
\]

(13)

A loose upper bound for this can be obtained by using \( \sin^2[(n + 1)\theta] \leq 1 \), which yields

\[
\bar{E}_A(N_P) \leq \frac{1}{\sin^2\theta} \sum_{n=N_P+1}^{\infty} r^{2n} = \frac{r^{2(N_P+1)}}{(1 - r^2)^2 \sin^2\theta}.
\]

(14)

Requiring the right-hand side to be smaller than the nominal residual energy \( \bar{E}_P \) results in a simple upper bound for the EL as

\[
N_{P\text{sim}} = \left[ \log\left( \frac{\bar{E}_P(1 - r^2)\sin^2\theta}{\log(r)} \right) - 1 \right].
\]

(15)

where the subscript “sim” stands for “simple”. Note that \( N_{P\text{sim}} \) is always larger than or equal to the true \( N_P \), since (14) overestimates the instantaneous residual energy. For filters with poles close to the unit circle this estimate can be grossly pessimistic so that a more accurate estimate would be desirable.

A more accurate but complicated estimate can be constructed from (13) by first using \( \sin^2 x = \frac{1}{2}(1 - \cos 2x) \) to get

\[
\bar{E}_A(N_P) = \frac{1}{2\sin^2\theta} \sum_{n=N_P+1}^{\infty} r^{2n}(1 - \cos[2(n + 1)\theta])
\]

\[
= \frac{r^2(N_P+1)}{2\sin^2\theta} \sum_{m=0}^{\infty} r^{2m}(1 - \cos[2(m + N_P + 2)\theta]).
\]

(16)

This can further be expressed in the form

\[
\bar{E}_A(N_P) = \frac{r^2(N_P+1)}{2\sin^2\theta} \left[ G_{r1} - G_{r2} \right]
\]

(17)

where

\[
G_{r1} = \frac{1}{1 - r^2}
\]

(18a)

\[
G_{r2} = \sum_{n=0}^{\infty} r^{2n} \cos(2n\theta + \theta_P)
\]

(18b)

and

\[
\theta_P = 2(N_P + 2)\theta,
\]

(19)

Further employing \( \cos(x + y) = \cos x \cos y - \sin x \sin y \) and the following sum formulas from [29, Eq. (1.447)]

\[
\sum_{k=0}^{\infty} k^p \sin(kx) = \frac{p \sin x}{1 - 2p \cos x + p^2}
\]

(20a)

\[
\sum_{k=0}^{\infty} k^p \cos(kx) = \frac{1 - p \cos x}{1 - 2p \cos x + p^2}
\]

(20b)

we obtain (18b) in the form

\[
G_{r2} = \frac{F(\theta_P)}{1 + r^4 - 2r^2 \cos(2\theta)}
\]

(21)

with

\[
F(\theta_P) = [1 - r^2 \cos(2\theta)] \cos \theta_P - r^2 \sin(2\theta) \sin \theta_P.
\]

(22)

The form (21) is exact but, unfortunately, \( F(\theta_P) \) still depends on \( \theta_P = 2(N_P + 2)\theta \) and thus on \( N_P \), which prevents the closed-form solution for \( N_P \) from (17). However, this dependence can be eliminated by considering \( \theta_P \) as a continuous variable in the range \([0,2\pi]\) and solving for bounds for \( F(\theta_P) \). The locations of these extrema are obtained by standard techniques (by solving for the zeros of the derivative of \( F(\theta_P) \)) to be

\[
\theta_{P1} = \arctan\left[ \frac{r^2 \sin(2\theta)}{r^2 \cos(2\theta) - 1} \right] + 2k\pi \quad \text{and} \quad \theta_{P2} = \theta_{P1} + (2k + 1)\pi, \quad k = 0, 1, 2, \ldots
\]

(23)

one of which yields the maximum \( F_{\text{max}} \) and the other the minimum \( F_{\text{min}} \). Hence, we obtain bounds for \( G_{r2} \) as

\[
G_{r2,\text{max}} = \frac{F_{\text{max}}}{1 + r^4 - 2r^2 \cos(2\theta)}
\]

(24a)

\[
G_{r2,\text{min}} = \frac{F_{\text{min}}}{1 + r^4 - 2r^2 \cos(2\theta)}
\]

(24b)
Fig. 2. Effective length of a second-order all-pole filter for \( P = 90\% \) as a function of the pole angle for the pole radius (a) \( r = 0.9 \) and (b) \( r = 0.95 \). The exact effective length computed according to the algorithm of Table I (solid line), the simple approximation of (15) (dotted line, top curve), and the upper and lower bounds of (25) (dashed lines).

which only depend on the filter pole parameters \( r \) and \( \theta \). Now we can solve for bounds of the effective length \( N_P \) from (17) as

\[
N_{P,\text{max}} = \left[ \frac{\log[2E_P\sin^2\theta] - \log(G_{P1} - G_{P2,\text{max}})}{\log(r^2)} - 1 \right]
\]

\[ (25a) \]

\[
N_{P,\text{min}} = \left[ \frac{\log[2E_P\sin^2\theta] - \log(G_{P1} - G_{P2,\text{min}})}{\log(r^2)} - 1 \right],
\]

\[ (25b) \]

Note that both bounds must be rounded upwards since the deviation of the unquantized value from the true one may be less than one quantization step.

Fig. 2(a) and (b) show the simple bound (15) and the upper and lower bounds (25) for the EL for \( P = 90\% \) as a function of the pole angle for the pole radii \( r = 0.9 \) and \( r = 0.95 \), respectively. It is seen that the upper and lower bound are very tight in the range where \( \theta_P = 2(N_P + 2)\theta \) is large enough, that is, when \( \theta_P > 2\pi \) or equivalently \( \theta > \pi/(N_P + 2) \) (and correspondingly for pole angles close to \( \pi \)). For small (and large) enough pole angles the upper bound (25a) yields a close estimate but the lower bound is increasingly inaccurate as there is no variation in \( \theta_P \). Fortunately, it is usually the upper bound for \( N_P \) that is of interest when it is desired to guarantee that at least a given percentage of the impulse response energy has arrived.

In the case \( r = 0.9 \), the simple bound (15) gives at most a four samples larger estimate than the tight upper bound (25a), so that for low values of \( r \) the accuracy of the simple bound may be quite satisfactory. For larger values of \( r \) the difference will be greater [see Fig. 2(b)] and the use of the tight upper bound (25a) is recommended.

C. Second-Order All-Pole Filter: Double Real Pole

A second-order all-pole transfer function with a double real pole \( P \) is

\[
H(z) = \frac{1}{1 - 2az^{-1} + a^2z^{-2}} = \frac{1}{(1 - az^{-1})^2}.
\]

The impulse response is now of the form

\[
h(n) = \begin{cases} 0, & n < 0 \\ (n+1)a^n, & n \geq 0. \end{cases}
\]

The accumulated energy can be expressed as

\[
E_A(N_P) = \sum_{n=0}^{N_P} (n+1)^2a^{2n} = (N_P + 1)^2r^{2N_P} + \frac{1}{r^2} \sum_{n=0}^{N_P} n^2r^{-2n}
\]

\[ (28) \]

where the second sum is obtained by applying [29, Eq. (0.114)] (note however the printing error in that reference: the second-last term \( n^2 \) in the numerator of 0.114 should be canceled). The total energy is obtained from (28) by zeroing the terms with \( N_P \) in the exponent, resulting in

\[
E = \frac{1 + r^2}{(1 - r^2)^2}.
\]

Unfortunately, using (28) and (29) to solve for \( N_P \) such that \( E_A(N_P) \geq E_P \) does not lead to an easy closed-form solution, but they can be used to efficiently search for the minimum \( N_P \) by successive evaluations. Using binary search, only about 17 evaluations of (28) are required, as compared to \( N_P \) steps of the recursive algorithm of Table I. For example, if we can assume that the impulse response is at most \( N_P = 256 \), only eight evaluations of (28) are required, as compared to up to 256 steps using the recursive algorithm of Table I.

Fig. 3 shows the 90%, 95%, and 99% effective lengths of the all-pole filter with a double real pole as a function of the
pole radius. The curves are similar to those of a single pole (compare with Fig. 1), but the EL now naturally rises faster with the pole radius.

D. Second-Order All-Pole Filter: Distinct Real Poles

The case of distinct real poles also requires separate treatment. The transfer function is

\[ H(z) = \frac{1}{1 - \left(1 - a_1 z^{-1}\right)\left(1 - a_2 z^{-1}\right)} \]

where we again assume real-valued poles such that \(|a_1| < 1\), \(|a_2| < 1\) and \(a_1 \neq a_2\). The impulse response now becomes

\[ h(n) = \begin{cases} 0, & n < 0 \\ \frac{a_1^n a_2^m - a_2^n a_1^m}{a_1 - a_2}, & n \geq 0 \end{cases} \]

and the accumulated energy is obtained as

\[ E_A(N_P) = \frac{1}{(a_1 - a_2)^2} \sum_{n=0}^{N_P} \left( a_1^{n+1} - a_2^{n+1} \right)^2 \]

\[ = \frac{1}{(a_1 - a_2)^2} \left[ a_2^2 (1 - a_2^{N_P+1}) \left( 1 - a_1^2 \right) + a_1^2 (1 - a_1^{N_P+1}) \left( 1 - a_2^2 \right) - 2a_1 a_2 \left( 1 - \left( a_1 a_2 \right)^{N_P+1} \right) \right] \].

The total energy is

\[ E = \frac{1}{(a_1 - a_2)^2} \left[ a_1^2 \left( \frac{1}{1 - a_1^2} + \frac{a_1^2}{1 - a_2^2} \right) - 2a_1 a_2 \left( \frac{1}{1 - a_1 a_2} \right) \right].\]

Similarly to the double-pole case, the expressions (32) and (33) do not yield \(N_P\) in closed form, but a search algorithm can be employed.

Fig. 4 shows the 90% EL as a function of the one pole in range \((-1, 1)\) when the other pole is fixed at \(a_2 = 0.5, 0.8, 0.9,\) and 0.95. Note that now the sign of the pole is important: when \(a_2 = -a_1\) or nearly so, the EL attains a minimum. This can even cause the length to be suppressed to zero as here for \(a_2 = 0.5\) when \(a_1\) is in the neighborhood of \(-0.5\).

IV. On the Effect of Zeros

All the above results consider all-pole filters only. In this section we show how the zeros affect the EL of recursive filters’ impulse response. A general first-order filter will be studied in detail after which general conclusions are drawn for higher-order filters.

A. General First-Order Recursive Filter

Consider a general first-order recursive filter with the transfer function

\[ H(z) = \frac{1 - b z^{-1}}{1 - a z^{-1}} \]

where \(a, b,\) and \(c\) are real-valued and \(|a| < 1\). The impulse response is now

\[ h(n) = \begin{cases} 0, & n < 0 \\ c(a-b)a^{n-1}, & n \geq 0 \end{cases} \]

The accumulated energy \(E_A(N_P)\) is (for \(N_P > 0\))

\[ E_A(N_P) = c^2 \left[ 1 + \sum_{n=1}^{N_P} (a-b)^2 a^{2(n-1)} \right] = c^2 + c^2 (a-b)^2 \frac{1- a^{2N_P}}{1- a^2} \]

from which the total energy is also obtained as a limit \((N_P \to \infty)\) as

\[ E = c^2 \frac{1 - 2ab + b^2}{1 - a^2}.\]
The EL can now be solved as

$$N_P = \left\lceil \frac{\log(1 - P/100) + \log[L(a,b)]}{\log(\alpha^2)} - 1 \right\rceil$$  \hspace{1cm} (38)

where

$$L(a,b) = \frac{1 - 2ab + b^2}{1 - b/\alpha^2}. \hspace{1cm} (39)$$

It is seen that (38) is the same as (8) except for an additive new term \(\log[L(a,b)]\). Since \(\log(\alpha^2) < 0\), this term increases the length of the impulse response when (39) is smaller than unity, which happens when

$$|\beta - a| > |\alpha|. \hspace{1cm} (40)$$

In the limit, the additional term goes asymptotically toward the minimum value \(\log[L(a,b)] \approx \log(\alpha^2)\) when \(|\beta| \rightarrow \infty\), which means that the impulse response (38) is lengthened by one sample at most. In this case the numerator approximates a unit delay, i.e., \(1 - b/\alpha^2 \approx b\).

On the other hand, the impulse response is shorter than (or equal to) that without the zero when \(|\beta - a| < |\alpha|\) (or in the neighborhood of that region—quantization of \(N_P\) causes some variation). For zeros close enough to the pole, the impulse response is suppressed down to zero. This is expected, because when \(b = a\), the zero exactly cancels the pole and the impulse response reduces to a unit impulse \(h(\eta) = \delta(\eta)\).

Fig. 5 shows the 90% EL of a first-order filter with a fixed pole \(\alpha = 0.9\) and the zero varying from \(-1\) to \(1\), clearly demonstrating that the length reduces to zero when \(b\) is close to the value \(a = 0.9\) (pole-zero cancellation), and that the EL is increased by one sample at most.

The special case of the EL of a first-order all-pass filter is elaborated in the Appendix.

**B. N Zeros**

The conclusions for the first-order filter can readily be generalized for higher-order filters. Consider a general recursive transfer function

$$H(z) = \frac{B(z)}{A(z)} \hspace{1cm} (41)$$

with the numerator \(B(z)\) of order \(M_B\). Assuming a fixed denominator, the longest possible impulse response corresponds to a delay of \(M_B\) units (one per each zero) and it is attained when the highest-order coefficient \(b_{M_B}\) of \(B(z) = b_0 + b_1 z^{-1} + \cdots + b_{M_B} z^{-M_B}\) is large enough compared to the others. The EL of the filter (41) has thus an upper bound

$$N_P\{H(z)\} \leq N_P\left\{ \frac{1}{A(z)} \right\} + M_B. \hspace{1cm} (42)$$

The smallest possible EL for the high-order filter is zero which naturally occurs due to (approximate) cancellation of all of the poles by corresponding zeros. This result will be used in the next section to obtain a general bound for high-order filters.

**V. HIGH-ORDER FILTERS**

Analytical treatment of higher-order filters soon becomes cumbersome, as can be seen already by comparing the expression for a single real pole (7a) to that of a double real pole (28). Instead of trying to derive complicated formulas of questionable utility, only approximate upper bounds will be derived here.

Let us focus on the case of effective length for a relatively large \(P\) (90%-99.99%) so that most of the energy has arrived by the time index \(N_P\) and we can neglect the tail of the impulse response. We define the length-\(N_P\) truncated impulse response as

$$h_{TR}(n) = \begin{cases} h(n), & \text{for } n = 0, 1, \ldots, N_P, \\ 0, & \text{otherwise}. \end{cases} \hspace{1cm} (43)$$

As the truncated impulse response is genuinely finite-length, we can obtain a simple approximative limit for the length of the convolution of two impulse responses \(h_1(n)\) and \(h_2(n)\) with effective lengths \(N_{P_1}\) and \(N_{P_2}\) as

$$N_P\{h_1(n) * h_2(n)\} \approx N_P\{h_{1TR}(n) * h_{2TR}(n)\} \leq N_{P_1} + N_{P_2}. \hspace{1cm} (44)$$

This follows because the length of the convolution of two sequences of lengths \(N_{P_1} + 1\) and \(N_{P_2} + 1\) is equal to \(N_P = (N_{P_1} + 1) + (N_{P_2} + 1) - 1 = N_{P_1} + N_{P_2} + 1\). or \(N_P = N_{P_1} + N_{P_2}\). (Remember that the effective length is one shorter than the number of coefficients!) Applying this result for several convolutions, we can express a general formula suitable for a filter consisting of \(K\) subsections

$$N_P\{h_1(n) * h_2(n) * \cdots * h_K(n)\} \approx N_P\{h_{1TR}(n) * h_{2TR}(n) * \cdots * h_{KTR}(n)\} \leq N_{P_1} + N_{P_2} + \cdots + N_{P_K}. \hspace{1cm} (45)$$

Let us then consider a general recursive transfer function where poles are divided into at most second-order real-coefficient sections as follows:

$$H(z) = \frac{B(z)}{A(z)} = \frac{B(z)}{\prod_{k=1}^{K+1} A_k(z)} \hspace{1cm} (46)$$
where the numerator $B(z)$ is of order $M_B$, and $K_A$ denotes the number of sections in the denominator. Combining (45) and (46) with the result (42) from Section IV.B, we obtain an approximative upper bound for the EL as

$$N_P\{H(z)\} = N_P\left\{ \frac{B(z)}{\prod_{k=1}^{K_A} A_k(z)} \right\} \leq M_B + \sum_{k=1}^{K_A} N_P\left\{ \frac{1}{A_k(z)} \right\},$$

(47)

This is a general-purpose result which can be applied to any kind of stable filters when the factorization to first or second-order real-coefficient sections is available. Note that the obtained estimate for EL is an approximate upper bound and it may be pessimistic, particularly for filters with poles and zeros close to each other resulting in partial pole-zero cancellation.

Let us illustrate the use of the bound for some practical high-order filters. Fig. 6(a) and (b) show the EL of classical Butterworth and Chebyshev II lowpass filters of order 6. The EL of the total impulse response is given as a function of the 3 dB cutoff frequency for the Butterworth filter and as a function of the passband cutoff frequency for the Chebyshev filter. The passband and stopband ripples were specified as 0.5 dB and 40 dB for the Chebyshev filter and the stopband cutoff frequency was minimized.

The EL was estimated by splitting the filters into three second-order sections and using (47) and the formulas for second-order filters given in Section III-B. Also the exact effective length was computed for reference using the algorithm of Table I. Fig. 6(a) and (b) show these two curves and also their differences to enable easier comparison.

The EL’s of Butterworth filters are estimated remarkably well using the lengths of second-order subsections. As Fig. 6(a) shows, the bound (47) gives only two to five samples longer estimates than the accurate one for cutoff frequencies up to 0.6 times the Nyquist frequency. Only for very wideband filters is the estimate more pessimistic, which is due to the fact that poles of the filter come disturbingly close to the zeros which are all at $z = -1$ (i.e., at the Nyquist or half the sampling frequency) and pole-zero cancellation distorts the estimate.

It is also worth noting that the exact EL is not a monotonic function of the cutoff frequency. This is caused by the fact that the length of a second-order section with complex-conjugate poles is not monotonic either but oscillates between the upper and lower bounds, depending on the pole angle as demonstrated in Fig. 2(a) and (b).

The EL estimation for Chebyshev II filters [Fig. 6(b)] yields results similar to those for Butterworth filters. For most practical filters, except for very narrowband or very wideband filters, the estimate is accurate within three to six samples. This may be because the Chebyshev II filter has maximally flat magnitude response in the passband like the Butterworth filter, which effectively keeps the poles at a safe distance from the unit circle. Somewhat surprisingly, the fact that the zeros are distributed in the stopband instead of being fixed at $z = -1$ seems not to have much effect except for very wideband filters.

It should be emphasized that, when the EL can determined off-line and an accurate result is required, the exact algorithm of Table I should be preferred. Only when a high-order filter is tuned on-line in a cascade or parallel configuration of first and second-order subsections, it may be attractive to use low-order formulas to enable fast and simple on-line EL estimation.

### VI. Examples

Let us then consider some real-life examples where the estimation of the length of the impulse response is crucial.

#### A. Implementation of Linear-Phase IIR Filters

As discussed in the Introduction, linear-phase IIR filters can be implemented by cascading a minimum-phase IIR filter $H(z)$ and its maximum-phase counterpart $H(z^{-1})$. For this the effective length of $H(z)$ must be determined.
In [19], Kormylo and Jain designed a third-order elliptic lowpass filter for the processing of a noisy ECG signal. The filter specifications were: passband ripple $A_p = 0.05$ dB, passband cutoff frequency $\omega_p = 0.175\pi$ (or 35 Hz for 400 Hz sampling frequency), and stopband attenuation $A_p = 16$ dB. For the cascaded linear-phase system the ripple values are of course doubled, i.e., the composite stopband attenuation is 32 dB.

For block implementation, an estimate for the length of the impulse response of the elliptic filter is required. In [19] it was suggested (apparently heuristically) that the length of four times the time constant $\tau$ of the pole with the largest radius should be used, which yields the length estimate of 24.25 sample intervals (using Smith’s approximation, i.e., time constant $\tau = 1/(1 - r_{\text{max}})$—in [19], no figures were given). The desired 32 dB stopband attenuation suggests that at most or \% of the impulse response energy can be lost in the truncation, which corresponds to \% samples, which is not far from the 4 $\tau$ estimate.

In [23], Powell and Chau employed a seventh-order elliptic lowpass filter with the passband ripple $A_p = 0.005$ dB, passband cutoff frequency $\omega_p = 0.057\pi$ and stopband attenuation $A_p = 35$ dB. Requiring that a bound for the maximum amplitude of transient errors be 70 dB below the signal level, it was derived in [23] that the block length of 200 samples is necessary. By requiring the residual energy of the impulse response to be below 70 dB, i.e. $P = 100\% \times (1 - 10^{-7}) = 99.99999\%$, results in the exact EL of $N_p = 160$ samples. Hence, assuming that the energy-based criterion is suitable for the application, 20\% savings in the processing delay can be achieved by using the proposed EL of the impulse response.

B. Transients in Tunable Fractional Delay All-Pass Filters

Let us then consider a tunable fractional (noninteger) delay implemented with an all-pass filter. Tunable fractional delays find applications, e.g., in speech processing and modeling of musical instruments (see [7]). All-pass filters are especially well suited for approximating a fractional delay in a feedback loop as the gain of the filter is guaranteed to be exactly unity at all frequencies and stability problems due to a larger-than-unity gain can be forgotten. As discussed in [7], the coefficients of an all-pass filter that approximates a desired noninteger (group) delay in a maximally flat sense are obtained in closed form and can thus be easily changed on-line. The change of coefficients however causes transients which can be very disturbing in audio applications.

The amplitude of the transients depends on the magnitude of the change in coefficients, but the length of the transient depends mainly on the impulse response of the filter after the change. Several approaches for eliminating these transients have been proposed, as discussed in [15]. The simplest approach is, perhaps, to execute the coefficient changes in enough steps; the best is to eliminate the transients completely using the novel technique proposed in [12]–[15]. The transient elimination technique is based on active cancellation of the transient for a chosen span of time. For this we need an estimate of the effective length of the filter’s impulse response.

The total delay of an $N$th-order all-pass filter is $N + d$ where $d$ is the fractional delay. Here we focus on fractional delay approximation in the range $d = -0.5 \ldots 1.0$, for reasons to be explained soon. For the design equations and delay approximation characteristics of these filters, see [7]. Fig. 7(a) and (b) show the exact effective length of the impulse response of a first and second-order maximally flat fractional delay all-pass filter as a function of the approximated delay, corresponding to $P = 99\%$ (solid line) and $P = 99.99\%$ (dashed line).
first samples of the transient will guarantee about 20 dB (40 dB) attenuation of the transient power. For a more detailed discussion of transient elimination, see [15].

It is also worth noting that the choice of the range for delay approximation may have a significant effect on transients. Often it is enough to approximate fractional delay in the range of one unit delay, which can usually be chosen as desired. If the range is chosen as [−0.2, 0.8], the transient length can be cut down by as much as three samples, which will facilitate the implementation of the transient eliminator. This choice will also affect the average accuracy of the delay approximation which has similarly asymmetric characteristics (see [14]).

VII. CONCLUSION

A new approach for determining the EL of the impulse response of a recursive filter based on the accumulated energy of the impulse response was proposed. The energy-based measure is argued to be better suited for many signal processing problems than former techniques that focus on the amplitude of the impulse response or the time constant of the system. Alongside a simple recursive algorithm to determine the EL for any stable recursive filter, closed-form formulas were derived for low-order all-pole filters which yield valuable insight in the effect of individual poles on the EL. The effect of zeros was studied and an approximate upper bound was derived for estimating the EL for higher-order filters using formulas for low-order filters.

The results find applications in several fundamental and advanced signal processing problems. The EL can be used to characterize the number of samples contaminated by the attack transient in the case when the coefficients of a recursive filter are changed on-line or the input signal changes drastically. The results may also be employed in context of a novel transient elimination technique or in constructing linear-phase IIR filter structures that employ time-reversal and truncation of impulse responses of recursive subfilters.

APPENDIX

A. Effective Length of First-Order All-Pass Filter

A special but useful case of a pole-zero filter is the first-order all-pass filter. Its transfer function is

\[ H(z) = \frac{-a + z^{-1}}{1 - az^{-1}} \] \hspace{1cm} (A1)

where the coefficient \( a \) is real-valued and \( |a| < 1 \). The corresponding impulse response is

\[ h(n) = \begin{cases} 0, & n < 0 \\ -a, & n = 0 \\ (1 - a^2)^n, & n \geq 1 \end{cases} \] \hspace{1cm} (A2)

The total energy (4) of the impulse response of any all-pass filter always equals to unity because \(|H(e^{j\omega})| \equiv 1\) by definition. The accumulated energy \( E_A(N_P) \) of the impulse response can be elaborated as

\[ E_A(N_P) = a^2 + (1 - a^2)^2 \sum_{n=1}^{N_P} (a^2)^{n-1} \]

\[ = a^2 + (1 - a^2)(1 - a^{2N_P}) \] \hspace{1cm} (A3)

which yields the EL of the impulse response in closed form as

\[ N_P = \frac{\log(1 - P/100) - \log(1 - a^2)}{\log(a^2)}. \] \hspace{1cm} (A4)

ACKNOWLEDGMENT

This work was carried out in part in School of Electronic and Manufacturing Systems Engineering (now the Department of Electronic Systems), University of Westminster, London, U.K. The authors would like to thank their colleagues there, Dr. J. Mackenzie and Dr. T. Wicks, for useful comments on the manuscript. Also the insightful and meticulous comments by the anonymous reviewers helped to remove some errors and improve the readability of the paper, and are gratefully acknowledged.

REFERENCES


Timo I. Laakso (SM’95) was born in Vantaa, Finland, in 1961. He received the Master of Science in Technology degree, the Licentiate of Science in Technology degree, and the Doctor of Science in Technology degree in electrical engineering from the Department of Electrical Engineering, Helsinki University of Technology (HUT), Espoo, Finland, in 1987, 1990, and 1991, respectively.

From 1992 to 1994 he was with the Nokia Research Center, Helsinki, working on the research of third-generation mobile communication systems. From 1994 to 1996, he was Lecturer at the University of Westminster, London, U.K. He is currently Professor of Telecommunications, HUT. He has published about 100 journal and conference publications on design and implementation of digital filters and mobile communications systems. His interests are in the design and implementation of digital signal processing algorithms for telecommunications.

Vesa Välimäki (S’90–M’92) was born in Kuorevesi, Finland, in 1968. He received the Master of Science in Technology degree, the Licentiate of Science in Technology degree, and the Doctor of Science in Technology degree in electrical engineering from Helsinki University of Technology (HUT), Espoo, Finland, in 1992, 1994, and 1995, respectively.

He was with the HUT Laboratory of Acoustics and Audio Signal Processing from 1990 until 1995. In 1996, he spent six months as a Postdoctoral Research Fellow in the Signal Processing Group, School of Electronic and Manufacturing Systems Engineering (now the Department of Electronic Systems), University of Westminster, London, U.K. He then returned to HUT, where he holds the position of Senior Assistant in Acoustics and Audio Signal Processing. Currently, he is on leave as a Postdoctoral Researcher at the Academy of Finland. His research interests include digital filter design, digital sound synthesis, and active noise control.

Dr. Välimäki is a member of the Audio Engineering Society, the Acoustical Society of Finland, and the International Computer Music Association.