Adaptive Noise Removal from Biomedical Signals using Warped Polynomials

Dr. ir. W. Philips

ELIS Technical Report
January, 1995

Adaptive Noise Removal from Biomedical Signals using Warped Polynomials

Dr. ir. W. Philips

Department of Electronics and Information Sytems
Univeristy of Gent
St.-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Tel.: 32-9-264.33.85 Fax: 32-9-264.35.94
E-mail: philips@elis.rug.ac.be

Abstract

This paper presents the time-warped polynomial filter (TWPF), a new adaptive filter for removing stationary noise from non-stationary biomedical signals. The filter fits warped polynomials to large segments of such signals. This can be interpreted as low-pass filtering with a time-varying cut-off frequency. In optimal operation, the filter's cut-off frequency equals the local signal bandwidth. However, the paper also presents an iterative filter adaptation algorithm, which does not rely on the (complicated) computation of the local bandwidth.

The TWPF has some important advantages over existing adaptive noise removal techniques: it reacts immediately to changes in the signal's properties, independently of the desired noise reduction; it does not require a reference signal and can be applied to non-periodical signals. In case of quasi-periodical signals, applying the TWPF to the individual signal periods leads to an optimal noise reduction. However, the TWPF can also be applied to intervals of fixed size, at the expense of a slightly lower noise reduction. This is the way non quasi-periodical signals are filtered.

The paper presents experimental results which demonstrate the usefulness of the adaptive filter in several biomedical applications: noise removal from ECG, respiratory and blood pressure signals and base line restoration of electro-encephalograms (EEGs).

1 Introduction

Biomedical signals are often contaminated by noise from sources such as power line interference and disturbances due to movement of the recording electrodes. In addition, biomedical signals often interfere with one another, e.g., signals due to muscle contractions often contaminate EEGs and ECGs. Often, the signal-to-noise ratio (SNR)[1] of biomedical signals is low. Therefore, filters for removing noise and interference are important in the medical practice.

Noise removal is complicated by the fact that the characteristics (e.g., local spectrum) of almost all biomedical signals vary in time. For instance, electrocardiograms (ECGs) tend to vary quasi-periodically, with each period corresponding to one heart beat; sometimes, even the shape of the ECG beats varies (slowly or abruptly) in time. Other signals such as electroencephalograms (EEGs) do not vary quasi-periodically, but have a local spectrum that changes in time.

In order to be successful, any noise removal procedure for biomedical signals must be adaptive, i.e., it must track the changing signal characteristics. Most adaptive noise removal techniques are based on the least mean squares (LMS) principle or on the recursive least squares (RLS) principle [9]. They gradually reduce the mean squared error between the *primary input signal* (i.e., the signal to be filtered) and some *reference signal*. Noise and interference removal techniques can be broadly classified into two categories, depending on whether or not a reference signal is actually recorded:

- **Techniques with a recorded reference signal.** Most techniques proposed for adaptive noise or interference removal are of this type; they use a reference input to estimate the noise in the primary signal and subtract the noise estimate from the primary signal.

  In the majority of techniques, the reference signal consists of noise which is correlated with the noise in the primary input [13, 14, 17, 18, 19]. In a few techniques, the reference signal consists of a recording at a different position of the same biomedical signal [14]. In that case, the noise in the primary and reference inputs must be uncorrelated.

- **Techniques using an artificial reference signal.** In this "single input" case the reference signal is derived from the recorded signal. For instance, "predictor" filters use a delayed version of the input signal as a reference. In filters for noise removal from ECGs the reference input may be an impulse sequence generated by a QRS-detector [14, 10]. Alternatively, it can be a QRS-synchronous repetition of an estimated ECG beat, obtained by averaging many beats. This was the case in [20], where the interfering ECG signal was removed from a magneto-encephalogram (MEG) and also in [12], where the ECG was the useful signal.

A disadvantage of the first set of techniques is that they require extra hardware (electrodes, wires, amplifiers, etc.) to record the reference signal; sometimes even, a suitable reference signal cannot be obtained at all. In this paper, we only consider "single input" filters.

Most noise removal techniques for biomedical signals rely on the quasi-periodical nature of either the useful signal or the noise and exploit this in different ways: for instance, in [13] a filter with a periodically varying impulse response is adapted once every ECG period; on the other hand, the procedure in [20] uses a quasi-periodical artificial reference signal. Such techniques are obviously not applicable when neither the signal nor the noise are quasi-periodical. Also, they distort the signal when spurious deviations from quasi-periodicity occur (e.g., ventricular extrasystolic beats)[2] and they rely on a correct and accurate QRS-detection. Finally, all of the techniques fail when QRS-complexes are wrongly detected and some of them even require that the QRS-positions are determined with high accuracy [11]. This can be a serious disadvantage

[1] In this paper we will consider interference an additional source of noise.

[2] In [20] and [14] this property was positively exploited to remove separate normal and abnormal heart beats.

in some applications, especially at moderate to low SNRs when even state of the art QRS-detectors [22, 23] fail from time to time.

A final disadvantage, which is typical of all LMS- and RLS-type filters (not only those exploiting quasi-periodicity) is that their adaptation speed depends on the desired output SNR. In fact, the higher the required level of noise reduction the slower the filter adapts.

This paper presents a new type of adaptive filter for removing stationary noise from highly non-stationary biomedical signals. In contrast to the LMS- and RLS-type adaptive filters, the adaptation speed of the new adaptive filter is independent of the desired level of noise reduction; in fact, the new filter reacts *immediately* to changes in the signal's morphology. Furthermore, it doesn't use a reference signal (not even an artificial one) and it is not restricted to quasi-periodical signals.

The new adaptive filter is called the *time-warped polynomial filter* (TWPF), because it approximates large (definitely non-stationary), possibly overlapping signal segments by warped polynomials [1]; the latter are functions $s_N(f;t)$ of the form $s_N(f;t) = p_N\Big(f(t)\Big)$, where $p_N(x)$ is a polynomial of degree N in x and $f(t)$ a strictly increasing function. Both the *warping function* $f(t)$ and the polynomial coefficients are adapted such as to minimize the mean squared error between $s_N(f;t)$ and the input signal $s(t)$. Because the *warping function* $f(t)$ is adapted to the signal being filtered, the method is both adaptive and non-linear.

The filter operates according to the principle that even highly non-stationary biomedical signals can be accurately represented by a time-warped polynomial of low degree[3] whereas stationary noise cannot. Hence, the filter output $s_N(f;t)$ is a smoothed version of the input signal $s(t)$. The degree of noise reduction increases with increasing N and high noise reductions can be achieved without distorting the signal.

On quasi-periodic signals, the adaptive TWPF performs optimally when the signal segments each correspond to exactly one period. Of course, this can only be achieved in conjunction with period-detection (e.g., QRS-detection). This might seem a significant disadvantage. However, in contrast to traditional filters, the effects of errors in period-detection on the filter's performance are always benign; even more importantly, the filter's performance remains more than adequate when the intervals are not period-synchronized at all, or when the signal is not quasi-periodical.

Section 2 presents the theory of the TWPF in detail; in particular it relates the optimal warping function to the local signal bandwidth and it discusses the noise reduction properties of the filter. Next, section 3 presents the algorithmic details of the adaptive filter. The main component of the filter is an algorithm for adapting the warping function to a particular signal. This algorithm is a significantly modified version of an earlier one [2], which was developed for compression purposes: the new algorithm is both faster, robuster and generally better suited for filtering. The final section presents some results obtained with the new adaptive filter on ECG, EEG, repiratory and blood pressure signals. The examples illustrate both fixed-interval filtering and period-synchronized filtering. They also demonstrate the superior performance of the adaptive TWPF over non-adaptive low-pass filters.

2 Approximation by warped polynomials

As mentioned in the introduction, the TWPF approximate each of the (possibly overlapping) segments of the input signal by a warped polynomial. In this section we summarize and the theory of time-warped polynomial approximation presented in [1, 2] and extend it. We first introduce some definitions and notations; next, we investigate the noise reduction properties of the non-adaptive TWPF, which has a fixed (non signal-adapted) warping function. Finally, we discuss the adaptive TWPF and we show how to optimally adapt its warping function to

[3] In the case of ECGs, this was demonstrated in [2].

the signal. We also compute the theoretical improvement in noise reduction of the optimally adapted TWPF over the optimal non-adaptive ideal low-pass filter.

Without loss of generality, we normalize the segments such that they correspond to the time-interval $[-1,1]$. We denote the input signal $s(t)$, $t \in [-1,1]$ and we assume that it was sampled at the $L+1$ time-instances $t_k = -1 + 2k/L$, $0 \le k \le L$. With $g \cdot h = (2/L)\sum_{k=0}^{L} g(t_k)h(t_k)$ we denote the *inner product* of the functions $g(t)$ and $h(t)$ on $[-1,1]$ and with $\|g\| = \sqrt{g \cdot g}$ the *norm* of $g(t)$. A function $g(t)$ is called *normalized* if $\|g\| = 1$; when $g \cdot h = 0$, the functions $g(t)$ and $h(t)$ are *orthogonal* on $[1-,1]$. Two functions that are normalized and mutually orthogonal are called *orthonormal.*

A strictly increasing function $f(t)$ with $f(1) = 1$ and $f(-1) = -1$ is called a *warping function*; such a function maps the interval $[-1,1]$ onto itself. If $p_N(x)$ is a polynomial of degree N and $f(t)$ a warping function, the function $q_N(f;t) = p_N\big(f(t)\big)$ is called a *warped polynomial* of degree N. In the following we stick to the notation $q_N(f;t)$, which explicitly shows the dependency on the warping function.

Warped polynomial approximations of signals are most easily computed in terms of the orthonormal warped polynomials $\phi_n(f;t)$, $n = 0, \ldots L$ associated with the warping function $f(t)$. These functions are defined by the recurrence

$$\begin{aligned} N_0\phi_0(f;t) &= \phi_0'(f;t) = 1 \\ N_n\phi_n(f;t) &= \phi_n'(f;t) = \Big(f(t) - \alpha_n\Big)\phi_{n-1}(f;t) - \beta_n\phi_{n-2}(f;t) \quad 1 \le n \le L \end{aligned} \tag{1}$$

where $\alpha_n = (f\phi_{n-1}) \cdot \phi_{n-1}$, $\beta_1 = 0$, $\beta_n = (f\phi_{n-1}) \cdot \phi_{n-2}$ and $N_n = \|\phi_n'\|$. It is easily seen that $\phi_n(f;t)$ is a warped polynomial of degree n.

When $f(t)$ is fixed, the warped polynomial of degree N that minimizes the mean squared approximation error $\|s - s_N\|$ is given by

$$s_N(f;t) = \sum_{n=0}^{N} A_n(f)\phi_n(f;t), \tag{2}$$

where $A_n(f) = s \cdot \phi_n$. The *coefficients* $A_n(f)$ depend on the base functions $\phi_n(f;t)$ and consequently on the warping function $f(t)$; of course, the coefficients also depend on the signal $s(t)$, but this (obvious) dependency is not reflected in the notations.

When $N \ll L$, the approximation $s_N(f;t)$ can be interpreted as the output of an ideal low-pass filter with a time-varying cut-off frequency $\nu_N(f;t) \triangleq N b\,(f;t)\,/(2\pi)$, where $b(f;t) \triangleq \left|\frac{d}{dt}\arccos f(t)\right|$ is called the *frequency function.* This result, which was derived in [2], is based on an approximation of $\phi_n(f;t)$ which holds only when $\nu_n(f;t) \le L/4$, where $L/4$ is the Nyquist-frequency (i.e., half of the sampling frequency) of the sampled segment. It can be shown that for $0 \le N \le L$,

$$\nu_N(f;t) \triangleq \min\Big(\frac{N}{2\pi}b(f;t), \frac{L}{4}\Big) \tag{3}$$

is a good approximation of the true cut-off frequency. As $\nu_N(f;t)$ never exceeds the Nyquist frequency, it is called the anti-aliased version of $\nu_N'(f;t) \triangleq N b\,(f;t)\,/(2\pi)$.

Assume that the signal $s(t)$ consists of a "desired" low-pass signal $d(t)$ and some unwanted white noise $n(t)$; let $B(t)$ and B_g be the local and global bandwidth [24] of $d(t)$, respectively. The following relationship between B_g and $B(t)$ will prove to be very important:

$$B_g \approx \max_{-1 \le t \le 1} B(t). \tag{4}$$

We now denote the warped approximations of $d(t)$ and $n(t)$ with $d_N(f;t)$ and $n_N(f;t)$. In noise removal applications the filter must not distort the useful signal $d(t)$, i.e., $d(t) \approx d_N(t)$

with good accuracy. According to eq. (3), and taking into account that $B(t)$ never exceeds the Nyquist frequency $L/4$, the useful signal is not distorted if and only if

$$\nu_N'(f;t) = N b\,(f;t)\,/(2\pi) \geq B(t),\ \forall t \in [-1,1]. \tag{5}$$

On the other hand, a good filter should reduce the noise energy $\|n(t)\|^2$ as much as possible. It is easily shown that for white gaussian noise $\|n_N\| \approx \sqrt{(N+1)/(L+1)}\,\|n\|$; therefore, the net noise reduction depends on the approximation order N only, and not on the warping function $f(t)$.

When eq. (5) is satisfied, i.e., when $d_N(t) \approx d(t)$, the TWPF increases the SNR by a factor G, where

$$G \triangleq \left(\frac{\|d_N\|}{\|n_N\|} \Big/ \frac{\|d\|}{\|n\|}\right)^2 \approx \left(\frac{\|n\|}{\|n_N\|}\right)^2 \approx \frac{L+1}{N+1}. \tag{6}$$

An adaptive filter should maximize G, i.e., minimize N, but without distorting the signal. According to eq. (5), the lowest possible N for which there is no signal distortion is

$$N^-(f) \triangleq 2\pi \max_{-1 \leq t \leq 1} B(t)/b\,(f;t)\,. \tag{7}$$

Clearly, $N^-(f)$ depends on the warping function and on the local signal bandwidth $B(t)$.

To find the warping function $f^*(t)$ that minimizes $N^-(f)$ we first note that $\int_{-1}^{1}\left|\frac{d\theta}{dt}\right| dt = \pi$ because $f(\pm 1) = \pm 1$. Consequently, the time-mean $\overline{\nu_N'(f;t)}$ of $\nu_N'(f;t)$ depends only on N:

$$\overline{\nu_N'(f;t)} \triangleq \int_{-1}^{1} \nu_N'(f;t)\,dt/2 = \frac{N}{2\pi}\int_{-1}^{1}\left|\frac{d\theta}{dt}\right| dt = N/4. \tag{8}$$

This implies that minimizing N is always equivalent to minimizing $\overline{\nu_N'(f;t)}$. Under the restriction (5), the minimal N results when $\nu_N(f;t) = \nu_N'(f;t) = B(t)$, $\forall t \in [-1,1]$; consequently, $b\,(f^*,t)$ is proportional to $B(t)$ and the corresponding minimal order $N^-(f^*)$, which we now abbreviate as N^* is given by $N^* \triangleq N^-(f^*) = 4\overline{B(t)}$.

The optimal TWPF is adaptive because its warping function $f^*(t)$ depends on the local signal bandwidth $B(t)$. According to eq. (6), the optimal filter increases the SNR by a factor

$$G^* \approx (L+1)/(N^*+1) \approx L/\left(4\overline{B(t)}\right). \tag{9}$$

In comparison, the optimal cut-off frequency of a time-invariant ideal low-pass filter equals the global signal bandwidth B_g. Such a filter reduces the noise energy in a segment by a factor $G_{\mathrm{lp}} \approx L/(4B_g)$, which is the ratio of the highest noise frequency (i.e., the Nyquist frequency) and the filter's bandwidth. Combined with eq. (4), this implies that

$$G^*/G_{\mathrm{lp}} \approx B_g/\overline{B(t)} = \left(\max_{-1 \leq t \leq 1} B(t)\right)/\overline{B(t)}, \tag{10}$$

which is never smaller than 1. Therefore, the optimal TWPF always performs better than the best ideal low-pass filter. As the local bandwidth of many biomedical signals deviates significantly from a constant value, the ratio (10) can be quite substantial.

We conclude this section with some important remarks. First, we emphasize that B_g and $B(t)$ are the global and local bandwidth of the *noiseless* signal $d(t)$. In practice, B_g and $B(t)$ must be measured on the *noisy* signal $s(t)$, because this is the only signal available. The optimal low-pass and adaptive TWPF filters designed using the measured B_g and $B(t)$ cannot perform as well as those designed using the exact B_g and $B(t)$: in particular, the estimated values of B_g and $B(t)$ may be lower than the true values, with signal distortion as a result. To avoid this, the value of $B(t)$ has to be overestimated, which decreases the noise reduction of both filters.

As shown in section 3, it is both possible and practical to estimate the optimal warping function without computing the local bandwidth. For this reason, we will not discuss the consequences of errors in $B(t)$ in detail. The filter adaptation algorithm in section 3 only requires determining a close upper bound N of N^*. Fortunately, N^* depends on the *mean* of $B(t)$, which can be estimated quite robustly without computing $B(t)$.

Finally, we note that in the previous discussions, we have explicitly required that both the TWPF and the low-pass filters do not distort the signal. In both cases, higher degrees of noise reduction can be achieved when some signal distortion is tolerated. Unfortunately, the signal distortion introduced by a TWPF or low-pass-filter filter not only depends on the filters (local) cut-off frequency and the signals (local) bandwidth, but also on the details of the local spectrum of the signal. However, in practice the acceptable signal distortion is often quite low. In that case, the results for the distortionless case hold with good accuracy.

3 The adaptive filter

Section 2 dealt with the theoretical properties of the TWPF. It showed that the approximation order N should be kept as small as possible, in order to obtain the highest possible noise reduction. On the other hand, N is bounded below by the requirement of no signal distortion. The optimal N results when the frequency function $b\,(f;t)$ is proportional to the local signal bandwidth.

In this section, we first present an algorithm for determining the optimal warping function $f(t)$ for a given signal segment. The algorithm is a much improved version of an earlier adaptation algorithm [2], which was developed for compression purposes. It does not require estimating the local bandwidth $B(t)$ and this eliminates some important practical and computational problems. In this section, we also present two different approaches for dividing the signal into segments, for use with the filter. We discuss the advantages and disadvantages of both methods and give some hints on selecting appropriate values for the different filter parameters.

According to section 2, the optimal frequency function can be computed from the local signal bandwidth $B(t)$, which could be estimated using, e.g., the windowed Fourier-transform [25]. This is not really practical when signal statistics vary rapidly, because of difficulties in determining an optimal window width. However, as demonstrated in [2], it is quite feasible to estimate of $b\,(f^*;t)$ directly, without computing $B(t)$ first. Table 1 outlines an improved version of the iterative algorithm that was used for computing $b\,(f^*;t)$ in [2]. The algorithm iteratively updates the frequency function $b(t)$, and computes the corresponding warping function $f(b;t)$ using the relation

$$f(b;t) = \cos\left(\int_t^1 b(u)\,du\right). \tag{11}$$

Note that $f(b;t)$ is a valid warping function if and only if $b(t) > 0$ and $\overline{b(t)} = \pi/2$. The algorithm assumes that an estimate N of N^* is known and computes the frequency function $b(t)$ that minimizes $\|s - s_N\,(f(b))\|$. When the signal is corrupted by white gaussian noise, the resulting $s_N\,(f(b);t)$ is a maximum-likelihood estimate of the noiseless signal $d(t)$. Later, we will return to the problem of selecting a suitable value for N. First, we describe the adaptation algorithm and its improvements over the one in [2].

The algorithm starts with an initial, constant estimate of $b(t)$, which is iteratively refined. The refinement procedure (step 2 in table 1) increases $b(t)$ in those time-intervals where the smoothed squared approximation error is larger than average and decreases $b(t)$ elsewhere, at least if this leads to a smaller total mean squared approximation error. According to the filter-interpretation, increasing $b(t)$ at a certain time-instant lowers any local signal distortion near t, The algorithm for filtering a single segment

Input: the input signal segment $s(t)$.

Output: the filtered segment $s_N(f;t)$.

Parameters: the intermediate filter order N, the final filter order M $(M \leq N)$, the excess filter order a $(a > 0)$, the window width W of the running average filter and the number of iterations n_{it}.

Notations:

- For any given signal $s(t)$, $e_N(s,b;t) \stackrel{\triangle}{=} s(t) - s_N\left(f(b);t\right)$, where $f(b;t)$ is the warping function corresponding to $b(t)$.
- $\mathcal{A}\left(b,N;t\right)$ is the "anti-aliased" version of the frequency function $b(t)$; it is computed from $b(t)$ by the algorithm in table 2 and satisfies $2N\mathcal{A}\left(b,N;t\right) \leq L\pi$.

The algorithm:

1. **Initialization:** Copy the signal $s(t)$ to $s'(t)$; let $b(t) = \pi/2$.
2. **Improve the warping function**, by performing the performing the following steps n_{it} times:
   (a) Compute the squared error signal $e_N^2(s',b;t)$ and smooth it, using a running-average filter with window width W. Let $\underline{e'_N(s',b;t)}$ denote the smoothed version of $e_N^2(s',b;t)$. Compute $b_u(t) = e'_N(s',b;t) - \overline{e'_N(s',b;t)}$.
   (b) Minimize $\|e_N\left(s',\mathcal{A}\left(b+\alpha b_u,N\right);t\right)\|$ w.r.t. α. Using the optimal α, compute $b'(t) = \mathcal{A}\left(b+\alpha b_u,N;t\right)$ and copy $b'(t)$ into $b(t)$.
   (c) With $f'(t)$ the warping function corresponding to $\mathcal{A}\left(b,N+a,t\right)$, compute $s_{N+a}\left(f';t\right)$ and copy $s_{N+a}\left(f';t\right)$ into $s'(t)$.
3. **Filter the segment** $s(t)$, i.e., compute the final approximation $s_M(f;t)$, with $f(t)$ the warping function corresponding to $\mathcal{A}\left(M,b;t\right)$.

Table 1: The algorithm for adaptively filtering a signal segment.

while decreasing $b(t)$ has the opposite effect. Therefore, in the absence of noise, the algorithm reduces the total signal distortion and eliminates it completely if $N \geq N^*$. Stationary noise, if present in the input signal, tends to drive the algorithm towards a constant $b(t)$ and away from its optimal value: this is because the amplitude of the noise component $d(t) - d_N\Big(f(b); t\Big)$ of the approximation error is large where $b(t)$ is small and vice versa. With decreasing SNR, this effect becomes more and more important However, the results of section 4 demonstrate that its influence remains benign in practice, even at moderately low SNR.

The algorithm for anti-aliasing a frequency function

Input: a frequency function $b(t)$.

Output: the anti-aliased version of $b(t)$, which satisfies eq. (12).

Parameters: the approximation order N, and a small positive number ϵ, which is a guard against numerical errors.

Notations:

- $l = L\pi/(2N)$, the value to which $b(t)$ should be limited.
- $S = \{k : 0 \leq k \leq L, b(t_k) < l\}$ and $\#S$, the number of elements of S, i.e., the number of samples $b(t_k)$ smaller than the limit l.
- $S' = \{k : 0 \leq k \leq L, b(t_k) \geq l\}$.

The algorithm:

1. **Initialization:**
   (a) Compute the total "excess bandwidth" $b_e = \sum_{k \in S'} \Big(b(t_k) - l\Big)$.
   (b) Limit $b(t)$: for all $k \in S'$, replace $b(t_k)$ with l.
2. **Redistribute the excess bandwidth:** While $b_e > \epsilon$
   (a) Compute $\#S$ and let $\Delta = b_e/\#S$.
   (b) For each k in S, i.e., for each sample with $b(t_k) < l$
      i. Let $\Delta' = \min(\Delta, l - b(t_k))$
      ii. Replace $b(t_k)$ with $b(t_k) + \Delta'$.
      iii. Decrease b_e by Δ'

Table 2: The algorithm for anti-aliasing a frequency function.

There are some essential differences between the improved algorithm in table 1 and the algorithm which was originally presented in [2]. These changes reduce computation time and improve the filter's performance. First of all, in the improved algorithm, only frequency functions $b(t)$ satisfying the anti-aliasing condition

$$N b(t)/(2\pi) \leq L/4 \tag{12}$$

are considered. The algorithm in table 2 appropriately modifies any frequency function to make it comply with the condition (12); this involves reducing $b(t)$ to $l \triangleq L\pi/(2N)$ wherever it exceeds l, and uniformly increasing $b(t)$ elsewhere to assure that it remains a valid frequency function. The old algorithm does not perform anti-aliasing and often produces a frequency function $b(t)$ that violates the condition (12); according to section 2, such a $b(t)$ cannot be optimal. Furthermore, experimental results to be published elsewhere, indicate that the recursion (1) tends to become unstable as soon as the condition (12) is violated anywhere. This is usually not a problem when N is relatively low (e.g., $N < 30$), but it can cause serious trouble with larger values of N (e.g., $N > 50$). The anti-aliasing step eliminates this (possible) problem for all N.

The second improvement of the adaptation algorithm is the replacement in step 2a of the polynomial approximation filter in the old algorithm by a simpler and many times faster running average filter. The original algorithm was meant for compression applications, where it is essential that the warping function $f(t)$ can be coded efficiently. For this reason, $f(t)$ was approximated by a low-degree polynomial in the old algorithm, which can be interpreted as a complicated way of filtering $b(t)$. In filter applications, coding the warping function is superfluous, hence the simplification.

Another improvement is that the number of iterations n_{it} in step 2 of the new algorithm is now fixed; in the old algorithm, the iteration was terminated when $\|s' - s_n\|$ no longer decreased significantly. This change is motivated by the observation that after one or two iterations, the adaptation algorithm typically produced a very close to optimal approximation, but that the convergence test was often not powerful enough to detect this, which sometimes resulted in a large number of superfluous iterations. With the new algorithm, this can no longer occur.

The final change in the algorithm, is the inclusion of the extra step 2c, which results in a more robust determination of the frequency function when the input SNR is low. This is because the signal $s_{N+a}(f';t)$ in step 2c is a filtered version of $s(t)$, which contains most of the signal when the approximation order a is sufficiently large, but which is less noisy than $s(t)$; this is the case even with a relatively small a and even at the end of the first iteration because the adaptation algorithm converges very quickly.

In step 3 of the algorithm, the filter order N is increased to M. The idea is to allow N to be somewhat lower than N^* during the adaptation of the warping function, in order to reduce the influence of noise on the warping function. This causes some mis-adaptation, which is compensated by choosing M sufficiently larger than N^*. In this way signal distortion is avoided in the final approximation.

The adaptive TWPF filter operates on signal segments. In the remainder of this section, we discuss two different segmentation strategies and we show how to select a suitable value for N. First, we show that in principle the chosen segmentation strategy does not affect the improvement in SNR, *but only as long as the order N in each interval is chosen optimally (i.e., $N = N^*$).*

Assume now that a large interval I, consisting of $L + 1$ samples is divided into $K + 1$ (non-overlapping) sub-intervals $I_k = [a_k - L_k/2, a_k + L_k/2]$, which are individually remapped to the standard interval $[-1, 1]$ and then filtered by the method described above. From now on, symbols with a subscript k refer to quantities associated with the interval I_k, while symbols without such a subscript refer to I. For instance, $L_k + 1$ is the number of samples in I_k and $L + 1 = \sum_{k=0}^{K} (L_k + 1)$ is the number of samples in I.

It is easily shown that filtering all intervals increases the global SNR (i.e., the SNR measured on I) by a factor $G = (L + 1)/(N_k + 1)$, where $N \triangleq -1 + \sum_{k=0}^{K} (N_k + 1)$ is the *global approximation order.* We now assume an optimal order selection in each interval, i.e., we assume that

$$N_k = N_k^* \triangleq L_k / \left(4\overline{B_k(t)}\right), \tag{13}$$

where in accordance with the theory of section 2, $B_k(t)$ is the local bandwidth of the segment I_k, *after this is remapped to the time-interval* $[-1,1]$. If $B'(t)$ denotes the local bandwidth of I, treated as the time-interval $[0,L]$, then $B_k(t) = (L_k/2)\, B'(a_k + tL_k/2)$ and

$$N^* \triangleq -1 + \sum_{k=0}^{L} (N_k^* + 1) \approx \sum_{k=0}^{K} 2 \int_{a_k - L_k/2}^{a_k + L_k/2} B'(t)\, dt = \int_0^L B'(t)\, dt. \tag{14}$$

This shows that $G^* \approx L/N^*$ is independent of the interval division.

An important practical problem is the selection of the numbers N_k. To avoid signal distortion, N_k should not be smaller than N_k^*, which can be estimated in different ways: firstly N_k^* can be estimated by computing $B_k(t)$. Unfortunately this requires many computations and requires the careful selection of some parameters (such as the window-width of the short-term Fourier-transform); secondly, N_k^* can be determined by computing the output signal for a few values of N, while monitoring the mean squared approximation error, which decreases linearly when $N > N^*$ and, for typical signals, much faster than linearly when $N \leq N^*$.

A third and much simpler solution, which is employed in section 4 is to experimentally determine an upper bound Δ (independent of k) for $\overline{B_k(t)}$ on a few typical signals of the type to be filtered. (e.g., on a few ECGs). Assuming the law of large numbers applies, the bound Δ can be tighter when the intervals are larger, because then the relative variations of N_k are smaller. For instance, assume that the interval lengths $L_k = L/(K+1)$ are equal; also assume that the expected value of $\int_a^b B'(t)\, dt$ is $\mu(b-a)$ and its variance is $\sigma^2(b-a)$, where μ and σ are constants. In that case, N_k might be set to the expected value of N_k^* plus m times its variance, where m is suitably large number. This leads to $\Delta \approx 2\mu L/K + 2m\sigma\sqrt{L/K}$. Hence $N \approx (K+1)\Delta \approx 2\mu L + 2m\sigma\sqrt{LK}$, which increases with K. This shows that large intervals are to be preferred, because they lead to a smaller N and consequently to a larger noise reduction.

The previous discussion implicitly assumes that $B'(t)$ is stationary, i.e., that the statistics of $B'(t)$ do not change in time. However, even if this is not the case, the adaptive TWPF performs better than the optimal non-adaptive low-pass filter, as long as an upper bound Δ for N_k^* much smaller than L/K can be found. As illustrated in the following section, this is the case for many practical signals.

An important class of biomedical signals (e.g., ECGs, blood pressure, etc.) are quasi-periodic. For such signals, it is best to synchronize the filter intervals to the individual signal periods. This way, the statistical fluctuations of N_k^* are smallest. Note that the different intervals need not have exactly the same length, and that small alignment errors between the filter intervals and the signal periods can be tolerated. In most segmentation methods large alignment errors occur very infrequently; furthermore, large errors can be detected by monitoring the length of and the mean squared approximation error in the different segments. This allows the filter algorithm to switch temporarily to a larger or a smaller N_k, depending on whether the approximation error and the interval length are above or below average. Finally, the results in section 4 demonstrate that quasi-periodical signals can be adequately filtered, even without synchronizing the filter segments.

Finally, we need to address a minor problem that is not apparent from the theory above: because different signal segments are filtered independently, small discontinuities arise at the interval boundaries. The magnitude of the discontinuities is reduced by slightly overlapping neighboring intervals I_k and only using the central part I'_k of the approximated intervals in the output signal. In this case L_k denotes the size of I_k, while L'_k is the size of I'_k. When the signal is quasi-periodical, the discontinuities can also be concealed by aligning the interval boundaries with high-frequency signal features (e.g., the QRS-complexes of ECG signals) [2].

We end this section with an estimation of the computational complexity of the filter algorithm. To this end, we note that the minimization w.r.t. α in step 2b is performed iteratively. Let n_α be the maximum number of iterations for this purpose. Using the results in [2], it can be verified that filtering an interval I_k requires of the order of $12 n_{\mathrm{it}} n_\alpha N_k L_k$ floating point

TWPF parameters

	Curve	L_k+1	L'_k+1	M_k+1	N_k+1	W	a	Mean cut-off (Hz)
Fig. 1	C	256	200	40	30	5	20	78
Fig. 2	B,C	1000	1000	40	40	20	20	2
Fig. 3	B	880	800	40	40	10	20	0.68
Fig. 4	B	128	100	30	30	5	20	29.3
Fig. 4	C	QRS	QRS	30	30	5	20	26.6

FIR-filter parameters

	Curve	β	γ	Order	Cut-off (Hz)
Fig. 1	D	0.08	0.05	139	80
Fig. 2	B,D	0.04	0.04	61	3.8

DCT filter parameters

	Curve	L_k+1	L'_k+1	M_k+1	Cut-off (Hz)
Fig. 3	C	880	800	40	0.68
Fig. 3	D	880	800	60	1.02

Table 3: Parameters of the filters used in the different figures.

operations (flop). Typically, N_k is roughly proportional to L_k. In the case of $K+1$ equally sized intervals, the total computational complexity per output sample is roughly proportional to the interval size L/K. In practice a compromise must be made between the conflicting requirements of a high noise reduction (large intervals) and low computational complexity (small intervals).

4 Results and discussion

This section presents some results obtained with the adaptive TWPF on different kinds of signals: ECGs, EEGs, blood pressure and respiratory signals. The considered applications are noise removal and base line correction. Both fixed-interval filtering and period-synchronized adaptive filtering (for quasi-periodical signals) are illustrated. Also, the TWPF is compared to non-adaptive FIR low-pass filters and to a low-pass filter based on the discrete cosine transform (DCT). The FIR filters were designed using the procedure in [26]; they have maximally flat stop and pass bands and depend on two parameters: the normalized cut-off frequency β and the normalized transition bandwidth γ. The DCT low-pass filter operates on the same overlapping intervals I_k as the TWPF and approximates the signal in those intervals by the best (in the mean squares sense) linear combination of the M_k+1 lowest-frequency DCT base functions. As in the TWPF, only the central parts I'_k of the approximated intervals I_k are used in the output signal to reduce possible discontinuities. It is easily seen that the DCT filter is an ideal low-pass filter with a cut-off frequency of approximately $M_k\nu_s/(2L_k)$, where ν_s is the sampling frequency.

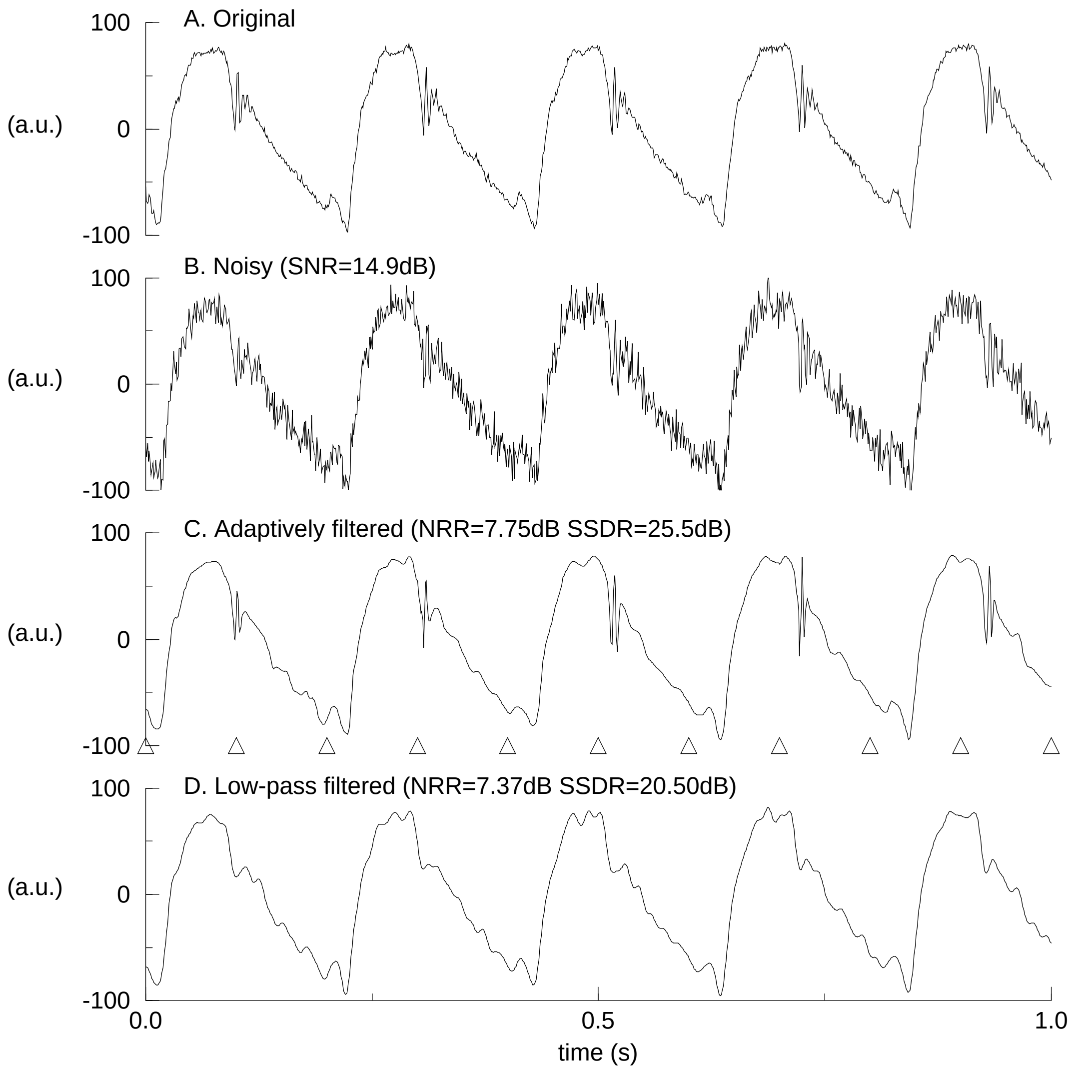

Figure 1: Heart pressure signal of a dog (sampled at 1kHz): A. the original signal; B. signal with added white noise; C. adaptively filtered signal; D. low-pass filtered signal. The filter parameters are listed in table 3.

The relevant parameters of the filters used in the different examples are listed in table 3. This table also indicates the filter order of the FIR filters, the 6-dB cut-off frequency of the low-pass filters and the mean local cut-off frequency of the TWPF. Finally, the boundaries of the non-overlapping parts I'_k of the TWPF intervals are marked by triangles in the figures.

The performance of any filter depends on the level of noise reduction it provides and on the signal distortion it causes. We quantify this performance using the *noise reduction ratio* $\mathrm{NRR} = 20\log(\|n\|/\|n_o\|)$ and the *signal-to-signal-distortion ratio* $\mathrm{SSDR} = 20\log(\|d\|/\|d_o\|)$, where $d(t)$ and $n(t)$ denote the signal and noise component of the filter input and $d_o(t)$ and $n_o(t)$ the corresponding output components.

The first example aims to estimate the NRR and the SSDR of the adaptive TWPF in a practical application: noise removal from a canine blood pressure signal. Measuring the NRR and the SSDR requires the noiseless signal $d(t)$ to be known. In the example, $d(t)$ is a signal taken from Signal Processing Information Base (SPIB)[4], which is assumed to be noise-free (fig. 1A). After adding computer-generated white noise (fig 1B), the signal is filtered with a fixed-interval adaptive TWPF.[5] The value of M_k (40) is chosen such that the signal quality is adequate in all intervals. A value of M_k lower than 40 would be acceptable in most intervals, but as explained in section 3, detecting when this is the case would add to the complexity of the filter.

For comparison, fig. 1 also shows the result of low-pass filtering the noisy signal. The cut-off frequency of the low-pass filter was adjusted to achieve approximately the same theoretical noise reduction as the adaptive TWPF: indeed, the theory of section 2 predicts that for the adaptive TWPF $G = L_k/N_k = 256/40 = 6.4$, while for the low-pass filter $G = 500/80 = 6.25$ (the ratio of the Nyquist frequency and the filter's cut-off frequency); these values correspond to a theoretical NRR of 8.06dB and 7.95dB, respectively. The actual (measured) NRR of the TWPF (7.75dB) is quite close to its theoretical value. As the low-pass filter is only an approximation of an ideal filter, its measured NRR (7.37dB) is somewhat lower than the theory predicts.

Even though both filters reduce the noise to approximately the same level, they distort the signal to quite different degrees: the SSDR of the TWPF is 5dB better than that of the low-pass filter. Fig. 1 clearly shows that the low-pass filter severely distorts the high-bandwidth regions of the signal; this results in the disappearance of the high-frequency oscillations. The TWPF performs much better in this respect.

Fig. 2 shows an EEG signal with a large base line wander, caused by eye movement. The amplitude of the base line is about eight times larger than the amplitude of the useful signal. Even though the bandwidth of the base line is much smaller than that of the EEG, it is difficult to separate both signals using non-adaptive filters; this is because the spectra of both signals overlap. This is confirmed by fig. 2, which shows the base line estimated using a low-pass FIR filter (fig. 2B, dotted line) and the reconstructed EEG (fig. 2D) obtained by subtracting the base line estimate. As the useful frequency range of the EEG extends from about 1 Hz to about 35 Hz, the cut-off frequency ν_c of the low-pass filter must be chosen as low as possible, to distort the EEG as little as possible. On the other hand, ν_c must be high enough to faithfully extract the base line. In the example, $\nu_c = 3.8$ Hz is the best compromise between these conflicting requirements.

Unfortunately, the non-adaptive procedure above removes the diagnostically important δ-waves (1 to 4 Hz) from the EEG. At the same time, it allows the highest local frequencies of the base line to leak into the reconstructed EEG; this results in artifacts which might easily be mistaken for actual EEG waves; for example, the area enclosed by the rectangle in fig. 2D contains such an artifact. Better results can be obtained with an adaptive TWPF: fig. 2B shows that the base-line estimated by the TWPF is much flatter over-all, except there where the base line fluctuates rapidly. As a result, most of the δ-waves are retained in the reconstructed EEG (fig. 2C). This is confirmed by fig. 2E, which shows that the local cut-off frequency of the TWPF is typically about 1 Hz. On the other hand, the TWPF cut-off frequency is much higher than that of the FIR-filter when base line fluctuations occur. As a result, the high local frequencies of the base line don't cause artifacts.

The third example shows the filtering a noisy pneumotach.[6] It shows mainly normal breathing, interrupted by a few seconds of no breathing. Fig. 3 illustrates that it is difficult to remove the noise using low-pass filters, without distorting the signal: fig. 3C and fig. 3D display the result of applying DCT low-pass filters with different cut-off frequencies to the noisy signal. Neither of these filters retain all high-frequency signal components. The filter with the highest

[4]Source: gopher://bellona.cs.rice.edu/11/SPIB.

[5]The TWPF is adapted *after* adding the noise.

[6]Source: gopher://sunserv0.cae.wisc.edu:70/00/ece/biomed/signals/respiratory/CHAN7.TXT; department of biomedical engineering, university of Wisconsin-Madison.

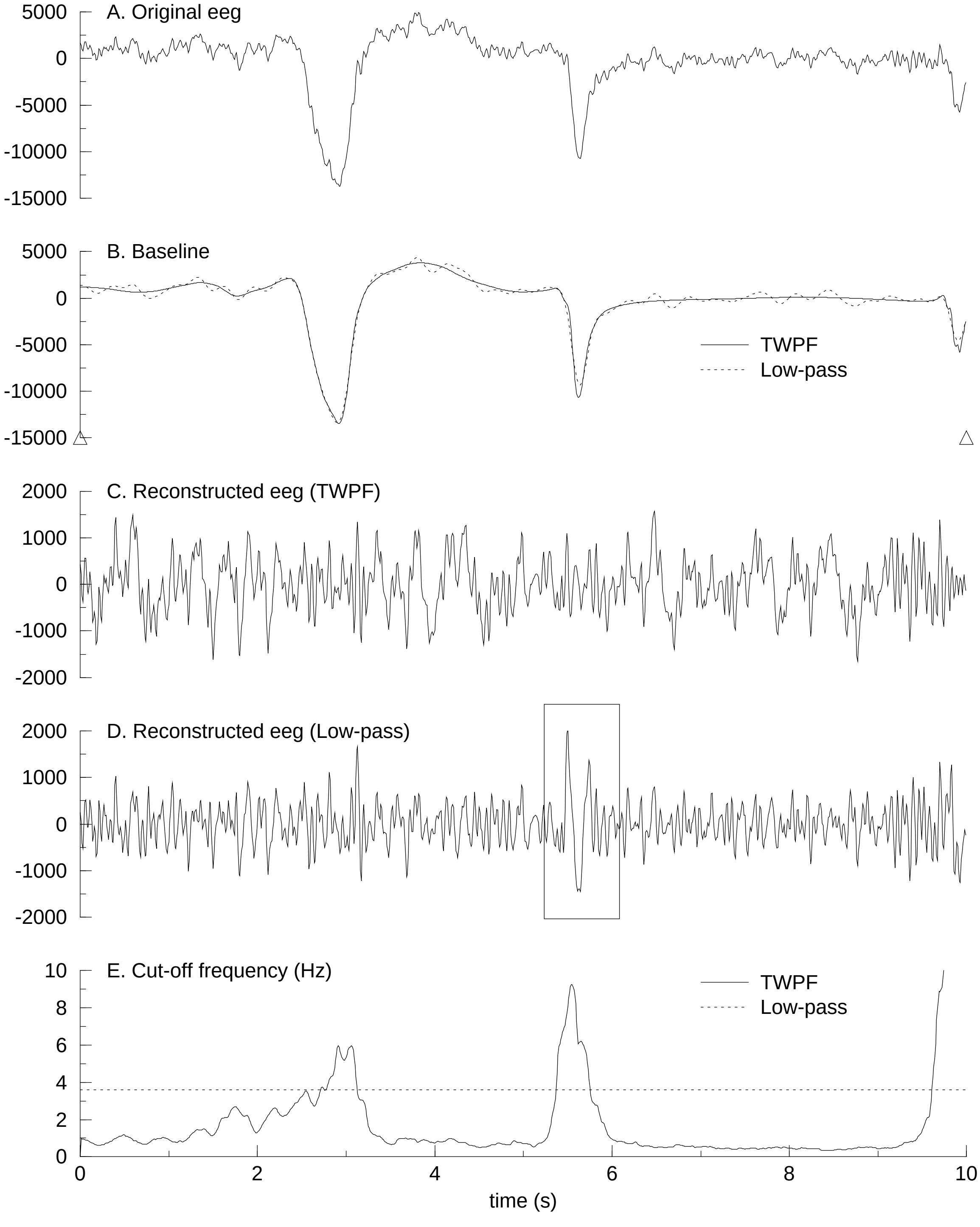

Figure 2: EEG signal with base-line wander (sampled at 100Hz): A. the original signal; B. the estimated baselines using adaptive and non-adaptive filtering; C. the EEG after adaptive base line removal; D. the EEG after non-adaptive base line removal; E. the cut-off frequencies of the adaptive and non-adaptive base line estimators. The filter parameters are listed in table 3.

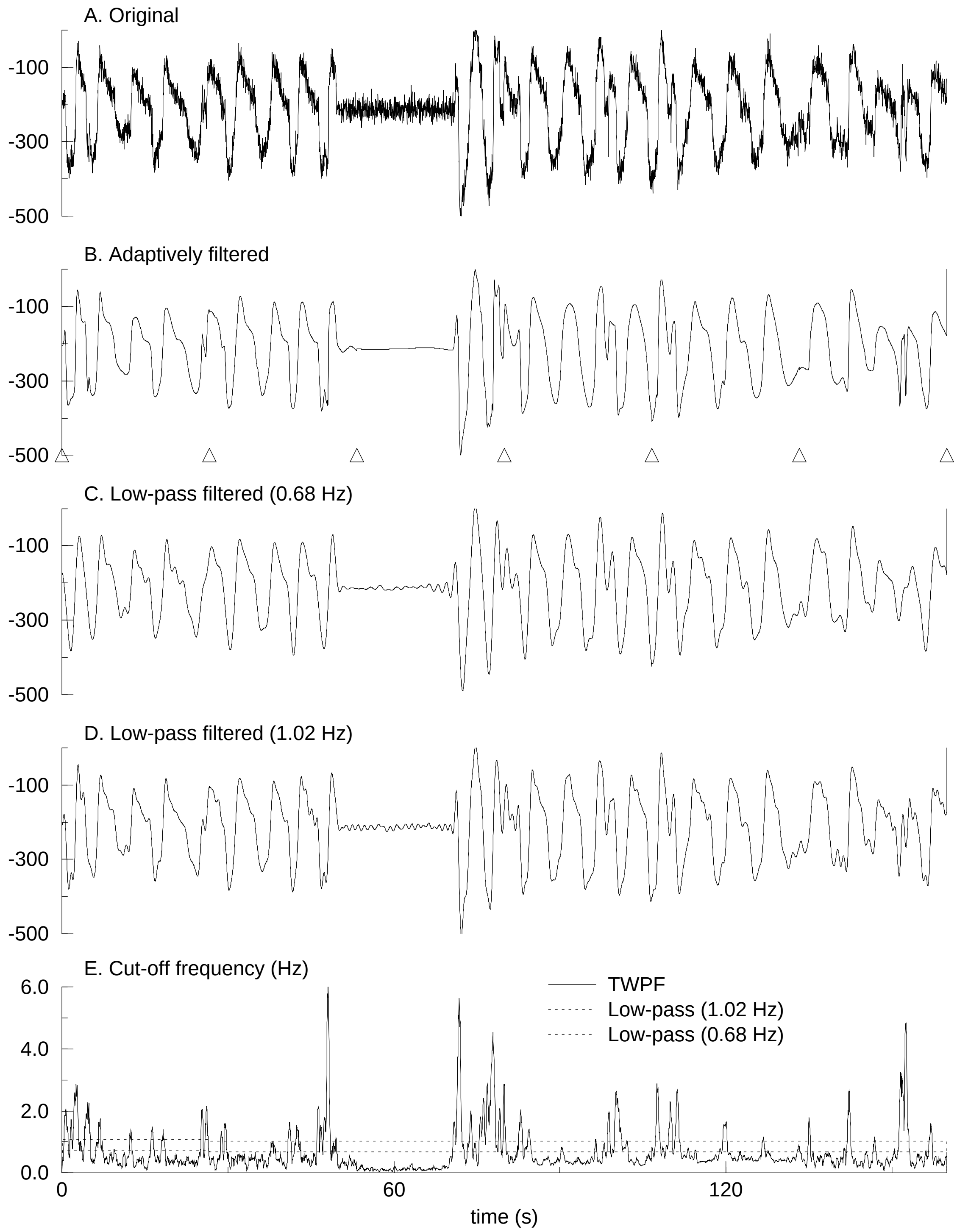

Figure 3: Pneumotach signal (sampled at 30Hz): A. the original noisy signal; B. adaptively filtered signal; C,D. low-pass filtered signals; E. the cut-off frequencies of the adaptive and non-adaptive filters. The filter parameters are listed in table 3.

cut-off is somewhat better in this respect, but it leaves more noise in the signal (this is best seen in the no breathing period). Theoretically, the adaptive TWPF of fig. 3D achieves the same noise reduction as the DCT filter with the lowest cut-off frequency. In addition, by comparing fig. 3B to fig. 3C and to fig. 3D, it is easily seen that the signal reconstructed using the TWPF is of much better quality than the signals reconstructed using the DCT.

In the previous examples, the TWPF is always compared to *non-adaptive* filters. The reader might feel the need for a comparison with traditional adaptive filters (LMS or RLS) that have been used on biomedical signals. As explained in the introduction, two types of such filters have been described in literature. The first type uses a reference signal, which is not available here. Therefore, there is little point in considering it here. The second type of adaptive filter can be applied to quasi-periodical signals only. Therefore it cannot be used in the last two examples. Because it requires a long training period, it is not easily applied in the blood pressure example either. Furthermore, it requires a very accurate segmentation [11] and does not react to sudden changes in the signal's morphology. Clearly, traditional adaptive filters are very useful, *but in quite different applications*. Therefore, a comparison with such filters seems inappropriate. Other types of adaptive filters such as adaptive predictors might be envisaged for noise removal, but they would basically converge to the optimal non-adaptive filter [13] and they would only do so after a long training period. For all of these reasons, traditional adaptive filters are not considered in this article.

To conclude this section, we present one more example (fig. 4) which compares fixed-interval adaptive filtering to QRS-synchronized filtering on a a noisy ECG signal; the signal is an extract of ECG 2207, lead 2 of the AHA arrhytmia database [27]. As the local bandwidth of of the QRS-complexes is much higher than that of the other waves, $N_k^* = \int_{a_k-L_k/2}^{a_k+L_k/2} B'(t)\,dt$ will be almost independent of a_k and L_k, as long as one and only one QRS complex is contained in I_k. In QRS-synchronized filtering this is always the case; hence N_k^* will vary little with k and a fixed $N_k = N$ is appropriate. In fixed-interval filtering, the interval width is chosen small enough, such that one interval never contains more than one QRS-complex. Therefore, N_k^* is never higher than in QRS-synchronized filtering, and the same fixed $N_k = N$ can be used. Of course, when an interval I_k does not contain a QRS-complex, N_k is much larger than N_k^* and the noise removal is not optimal in that interval.

The difference in noise reduction between both types of TWPF can be estimated using eq. (6) together with the fact that in fig. 4B L_k is typically about 140. From this it follows that the NRR is about 6.7dB for the QRS-synchronized filter and about 6.3dB for the fixed-interval filter. Clearly, the difference is small. This is confirmed by the graphs in fig. 4. The cut-off-frequencies of both adaptive filters are shown in fig. 4D and fig. 4E, respectively. In the vicinity of the QRS-complexes, the cut-off frequencies equal the Nyquist frequency (125 Hz) and the filters are effectively turned off. This also demonstrates the operation of the anti-aliasing procedure of table 2.

The examples above illustrate the potential of the adaptive TWPF in several applications. Currently, extensions of the filter technique for removal of noise from medical images are being investigated.

5 Conclusion

The adaptive TWPF is a new method for adaptively removing noise from biomedical signals. It approximates segments of such signals by warped polynomials of a degree which is low compared to the number of samples in the segment. In contrast to other methods, the TWPF requires neither a reference signal, nor a quasi-periodical input. Also, the TWPF reacts immediately to changing signal properties (local bandwidth), independently of the desired noise reduction.

From a theoretical point of view, the TWPF is an ideal low-pass filter with a time-varying cut-off frequency, which is adapted to the local signal bandwidth. When the filter is optimally

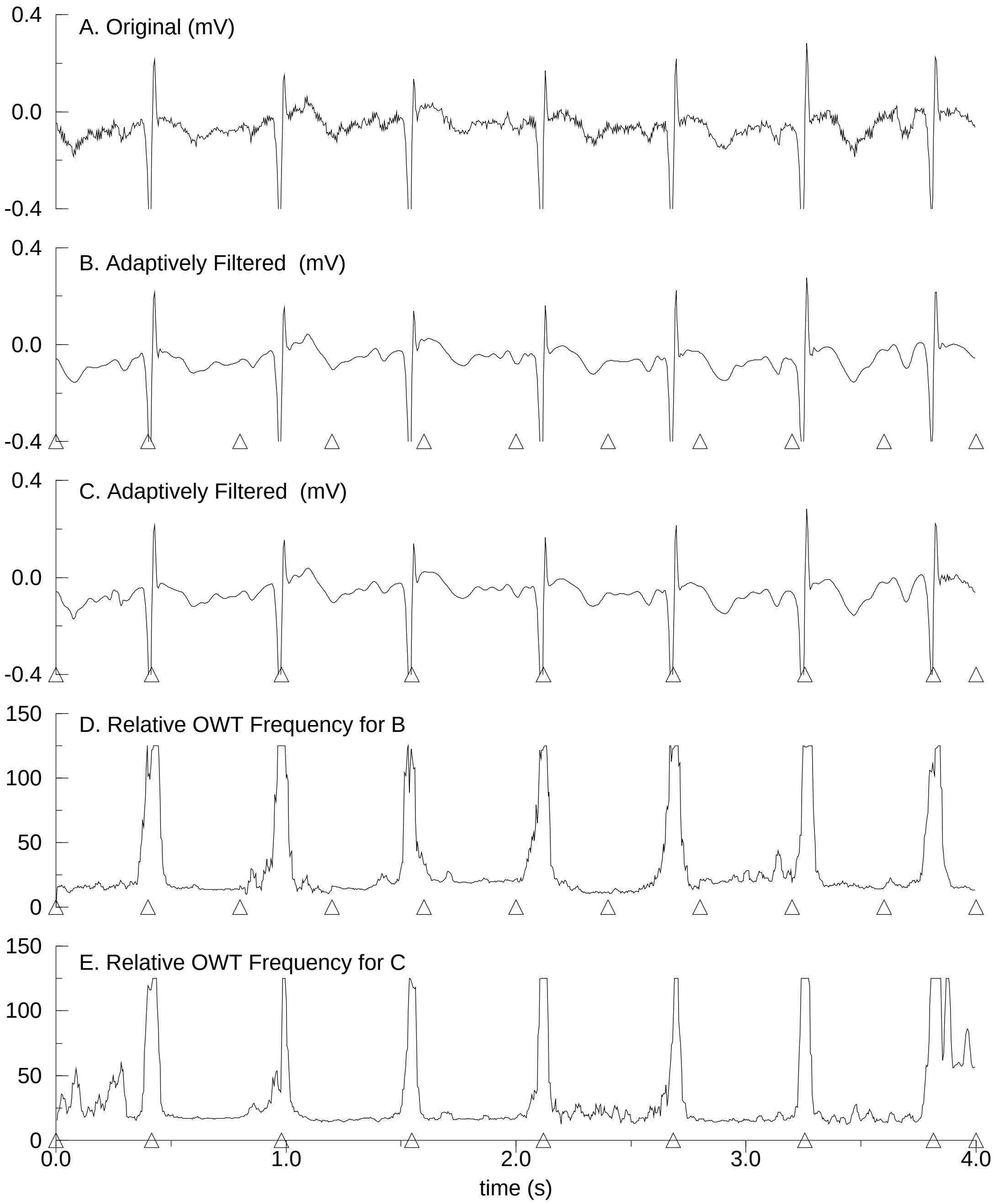

Figure 4: ECG signal (sampled at 250 Hz): A. the original signal; B. fixed-interval adaptive filtering; C. QRS-synchronized filtering; D,E. the local cut-off frequencies of the filters in B and C. The filter parameters are listed in table 3.

adapted, it reduces the noise energy in a given interval by a factor which is inversely proportional to the *mean* of the local signal bandwidth in that interval. In contrast, the corresponding factor for a non-adaptive filter is generally much smaller, because it is inversely proportional to the *maximum* of the local bandwidth.

In practice, the filter adapts iteratively, without computing the local bandwidth. The adaptation algorithm is a much improved version of an earlier algorithm which was specifically developed for signal compression. The new algorithm is faster and robuster; it can be applied to (possibly overlapping) intervals of fixed or arbitrary length or, in the case of quasi-periodical signals, to signal periods. Fixed-interval filtering is simpler because the signal segmentation is trivial; on the other hand period-synchronized filtering theoretically achieves a closer to optimal noise removal.

The paper demonstrates the advantages of the TWPF over other techniques in removing noise from ECG, pneumotach and blood pressure signals. It also shows that the TWPF is able to correct the base line of EEG signals. The examples confirm that the TWPF is able to remove stationary noise or noise-like signals from several types of non-stationary signals. Finally, the results show that in practice the difference in noise reduction between fixed-interval filtering and period-synchronized filtering can be quite small.

Acknowledgment

The author would like to thank the Belgian National Fund for Scientific Research (NFWO) for financially supporting his work through a mandate of "postdoctoral research fellow" and through the NFWO projects 39.0051.93 and 31.5831.95.

References

[1] W. Philips, "Coding properties of time-warped polynomial transforms," *Signal Processing*, vol. 37, pp. 229–242, May 1994.

[2] W. Philips, "ECG data compression with time-warped polynomials," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 1095–1101, Nov. 1993.

[3] D. Cochran and J. Clark, "On the sampling and reconstruction of time-warped bandlimited signals," in *Proceedings of ICASSP-90*, vol. 3, pp. 1539–1541, Apr. 1990.

[4] G. Szegö, *Orthogonal Polynomials*, vol. 23 of *American Mathematical Society Colloquium Publications*. Fourth ed., 1975.

[5] W. Philips, "Time-warped polynomials for signal representation and coding," in *Signal Processing VI: Theories and Applications*, vol. 2, (Amsterdam), pp. 925–928, Elsevier, Aug. 1992. Proceedings of EUSIPCO-92, Brussels.

[6] W. Philips, *Vervormde Polynomen voor de Compressie van Eén- en Tweedimensionale Signalen*. PhD thesis, ELIS, RUG, Gent, Belgium, 1992–1993. In Dutch.

[7] A. Habibi and P. Wintz, "Image coding by linear transformation and block quantization techniques," *IEEE Trans. Comm. Technology*, vol. 19, pp. 50–62, Feb. 1971.

[8] M. Tasto and P. Wintz, "Image coding by adaptive block quantization," *IEEE Trans. Comm. Technology*, vol. 19, pp. 957–972, Dec. 1971.

[9] S. Haykin, *Adaptive Filter Theory*. London: Prentice Hall, second ed., 1991.

[10] P. Laguna, R. Jané, O. Meste, P. Pool, P. Caminal, R. Hervé, and N. Thakor, "Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: Comparison with signal averaging techniques," *IEEE Trans. Biomed. Eng.*, pp. 877–885, Oct. 1992.

[11] F. Lippens, P. Savard, and M. Shenasa, "Performance degradation of the coherent averaging technique by molding effect," *IEEE Trans. Communications*, pp. 856–858, Aug. 1989.

[12] A. Colomer and V. Terré, "Electrocardiogram enhancement by adaptive filtering," in *Signal Processing VII: Theories and Applications*, vol. 1, (Amsterdam), pp. 58–61, Elsevier, Sept. 1994. Proceedings of EUSIPCO-94, Edinburgh, Scotland.

[13] E. Ferrara and B. Widrow, "The time-sequenced adaptive filter," *IEEE Trans. Circ. Syst.*, pp. 519–523, June 1981.

[14] N. Thakor and Y.-S. Zhu, "Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhytmia detection," *IEEE Trans. Biomed. Eng.*, pp. 785–794, Aug. 1991.

[15] S. Roberts, T. Chaumeron, and J. Magdolen, "Adaptive noise removal from complex signals using the wavelet transform," in *Signal Processing VII: Theories and Applications*, vol. 1, (Amsterdam), pp. 66–69, Elsevier, Sept. 1994. Proceedings of EUSIPCO-94, Edinburgh, Scotland.

[16] A. Hambley, R. Moruzzi, and C. Feldman, "The use of intrinsic components in an ECG filter," *IEEE Trans. Biomed. Eng.*, vol. 21, pp. 469–473, Nov. 1974.

[17] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin, "Adaptive noise cancelling: Principles and applications," *IEEE Proceedings*, vol. 63, pp. 939–952, Dec. 1975.

[18] E. Ferrara and B. Widrow, "Fetal electrocardiogram enhancement by time-sequenced adaptive filtering," *IEEE Trans. Biomed. Eng.*, pp. 458–460, June 1982.

[19] M. Yelderman, B. Widrow, J. Cioffi, E. Hesler, and J. Leddy, "ECG enhancement by adaptive cancellation of electrosurgical interference," *IEEE Trans. Biomed. Eng.*, pp. 392–398, July 1983.

[20] P. Strobach, K. Abraham-Fuchs, and W. Härer, "Event-synchronous cancellation of the heart interference in biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 343–350, Apr. 1994.

[21] M. Ferdjallah and R. Barr, "Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 529–536, June 1994.

[22] L. Sörnmo, O. Pahlm, and M. Nygards, "Adaptive QRS detection: A study of performance," *IEEE Trans. Biomed. Eng.*, vol. 32, pp. 392–401, June 1985.

[23] J. Pan and W. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, pp. 230–236, Mar. 1985.

[24] D. Slepian, "On bandwidth," *IEEE Proceedings*, vol. 64, pp. 292–300, Mar. 1976.

[25] F. Hlawatsch and G. Boudraux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, pp. 21–67, Apr. 1992.

[26] J. McClellan, "FIR filter design and synthesis," in *Programs for Digital Signal Processing*, ch. 5, pp. 5.3–1–5.3–6, IEEE Press, 1979. Edited by the Digital Signal Processing Committee.

[27] American Heart Association, ECRI, 5200 Butler Pike, Plymouth Meeting, PA 19462, U.S., *AHA Arrhythmia Database*.