Warped Polynomials and their Applications in Signal and Image Processing

Dr. ir. W. Philips

Department of Electronics and Information Systems (ELIS)
Medical Image and Signal Processing group (MEDISIP)
Institute Biomedical Technology (IBITECH)

A summary of the author's research on warped polynomials

ELIS Technical Report WP 96-02
January, 1996

UNIVERSITEIT GENT
Elektronica en Informatiesystemen

MEDISIP

IBITECH
Institute Biomedical Technology
GENT

This work was awarded the "Scientific Award Siemens 1996"

Preface

This document summarizes the author's research on adaptive image and signal representations using warped polynomial approximations and their applications in (mainly) coding and filtering. Some of the research results described in this document are explained in much greater detail in the author's Ph.D. thesis. However, this document additionally presents many recent new results concerning, e.g., adaptive filtering, and image and contour coding. It also discusses a few possible extensions and new applications of the theory.

Acknowledgments

The author wishes 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.

He is also grateful to his colleagues and friends (including researchers from other universities and research institutions), for the many fruitful discussions concerning his research. In particular, he would like to thank prof. dr. ir. J. Van Campenhout, prof. dr. ir. M. Vanwormhoudt and prof. dr. I. Lemahieu for stimulating him in his research, and ir. K. Denecker for critically reviewing the first version of this manuscript.

Contents

List of Figures

1 Introduction

This document presents an overview of the author's research on warped polynomials and their applications in signal and image processing. Basically, this research deals with the representation of signals and images in terms of signal- and image-dependent base functions. Such *adaptive* representations have important applications in signal and image coding and filtering, where they may lead to higher compression ratios and better signal/noise separation, respectively.

Adaptive representations have received little attention in literature, due to a great number of potential problems. For instance, it is not easy to define base functions which can be adapted to a signal or image and which are at the same time easy to compute and more efficient[1] than fixed base functions. Even if such base functions are available, adapting them to a specific signal or image is in general a complicated optimization problem, which can only be solved within a reasonable time by exploiting prior information about the optimal solution.

This document shows that in spite of these difficulties, adaptive signal and image representation can be made both feasible and practical by using orthonormal warped polynomial base functions. The following sections summarize the elaborate theory concerning warped polynomials (and their extensions). This theory explains what warped polynomials are and how they are used in adaptive signal representation. It also allows a unified treatment and an easy comparison of many existing transforms. This document also presents several new practical coding and filtering techniques for signals, images and contours which are based on warped polynomial approximation. In each case, the techniques are shown to outperform non-adaptive techniques, both theoretically and experimentally.

This document is only a summary of the author's research on warped polynomial approximation. Readers interested in the technical details are referred to the Ph.D. thesis [3] (in Dutch) and the articles [2,4,6,10,15,25]. They are also encouraged to consult the ELIS WWW-server[2] for the most recent information.

In the remainder of this section, we first describe some of the possible practical applications of the developed coding and filtering techniques. Next, we discuss adaptive coding in general in somewhat greater detail. Finally, we present a more detailed overview of the contents of this document.

1.1 The importance of data compression and adaptive filtering

Initially, digitized signals are represented as series of sample values and digitized images as two-dimensional arrays of pixels. Data compression is the (exact or approximate) coding of signals and images in an alternative form, which requires fewer bits. Compression allows signals and images to be transmitted faster and more economically over communication channels or, alternatively, to transmit more signals simultaneously and in real time over those channels.

Some compression techniques are *lossless*, which means that the signal can be exactly recovered from the coded data. *Lossy* coding techniques, on the other hand, only allow the signal to be approximately recovered from the coded data but they are generally much more efficient than lossless techniques. In most applications (e.g., consumer electronics),

[1]The following sections will make clear what “efficient” means in this context.

[2]http://www.elis.rug.ac.be/ELISgroups/mbv/philips/

small errors can be tolerated and therefore lossy techniques are widely used and accepted. Lossy data compression is essential in, e.g., the digital transmission and storage of the following signals:

- **Biophysical signals.** In a not very distant future, monitoring equipment in ambulances will transmit biophysical signals such as electrocardiograms (ECGs), electroencephalograms (EEGs) and blood pressure signals to a nearby hospital over low bit rate radio channels. This will allow the physician in the hospital to provide ambulance personnel with better-founded instructions. Compression is required to transmit the required number of signals (e.g., three ECG leads and one blood pressure signal) simultaneously, in real-time and without loss of quality.

  Remote monitoring of patients in home care is another application in which real-time transmission of biophysical signals is essential. In this case, the signals are transmitted over ordinary telephone lines, which, just like radio channels, offer only a limited bit rate. Home care will become more and more important because it enables important reductions in health care costs.

  Telephone transmission of biophysical signals is also required in tele-diagnosis, which is becoming more and more important. For instance, in Spain, workers who are exposed to harsh environments (e.g., building workers exposed to very hot weather) are required (by insurance companies) to undergo regular health checks. Some companies now use portable devices to record ECG and blood pressure signals on-site and transmit them to a remote diagnosis center by telephone. This approach is much more efficient than sending a worker to a doctor or hospital, which takes at least a few hours. Airline and shipping companies are also very interested in tele-diagnosis, because it may save lives, e.g., on transatlantic flights.

  Finally, biophysical signals sometimes need to be stored very compactly. For instance, in long-term recorders for ECGs or other biophysical signals on portable RAM-recorders, compression is needed to achieve the desired recording time (e.g., 24 hours) with a reasonably sized RAM. Compression would also be beneficial in implantable automatic defibrillators; for diagnostic purposes, these devices store ECG-events that trigger defibrillation.

- **Music.** The digital compact cassette (DCC) and the mini-disc system (MDS) are devices that would be difficult and expensive to realize without some form of compression. In fact, they are based on transform coding (sub-band coding and DCT-coding, respectively).

  Digital Radio (already available on satellite and currently being tested terrestrially in London) also uses compression because uncompressed digital signals would require a much larger bandwidth than analog FM-signals and because bandwidth is costly, especially in terrestrial and cable systems.

- **Medical images.** Picture archiving is becoming an important problem in hospital information systems and compression is needed to reduce its cost. In the medical environment, compression also allows large images to be transmitted faster over local area networks and telephone lines. This is especially important in tele-diagnosis applications,[3] which are of particular importance to remote areas and third-world countries.

- **Satellite images.** Huge amounts of image data are generated each day by earth, climate and weather observation satellites. Again, efficient archiving and transmission is essential for economic reasons.

- **Digital video.** Compression techniques are an integral part of digital television, which would require an excessive amount of bandwidth if no compression were used. The recently adopted MPEG television standard combines DCT-coding with predictive techniques. MPEG is also employed in multi-media applications, such as CDI and in video phones.

Most of these applications use transform coding. Also, the DCT is often the preferred transform. The compression techniques described in this document are extensions of the DCT and are therefore useful in all of these applications.

The techniques in this document can not only be used for compression but also for filtering. In many applications, signals are contaminated by noise or by other undesirable signals. Sometimes the desired and undesired signal (or image) components can be adequately separated using classical time-invariant filters; however, often this is not possible, e.g., because the spectrum of the desired signal and the spectrum of the undesired signal (e.g., noise) overlap to much. In this case, adaptive filters, which adapt to the time-varying statistics of the signal are often used. However, in some applications their performance is mediocre or even bad. Examples of such applications are:

- **Noise removal from ECGs and other biomedical signals.** ECG signals are low-voltage signals and are often very noisy. Because the signals are highly non-stationary, adaptive filters must be used. However, classical adaptive filters usually cannot adapt quickly enough to the rapidly changing signal characteristics without compromising their noise reduction ability. Special purpose filters have been developed to solve this problem, but these solutions have other disadvantages (such as requiring an extra signal or noise reference signal or relying heavily on periodicities in the signal).

- **Baseline correction of EEGs.** The typical amplitude of EEG signals is many times smaller than the typical amplitude of muscle signals (electromyograms). Therefore, EEGs are usually superimposed on a large-amplitude baseline, which is caused by, e.g., eye movements. Classical adaptive filters are not able to separate both components, again because they react too slowly.

- **Speckle reduction in ultrasound images.** Due to Rayleigh scattering, ultrasound images always contain speckles (spots). In order to reduce the amplitude of the speckles, while retaining the edges in the images, adaptive filters must be used.

In the first two of these applications, the author has already shown that adaptive filters based on warped polynomial approximation do not suffer from the disadvantages of classical adaptive filters and are able to adequately separate the desired and undesired signal

[3] For instance, doctors in rural areas can seek the advice of specialists in dedicated hospitals, by transmitting them the relevant images.

components (see section 3). Furthermore, the results in section 3.4 suggest that adaptive warped polynomial image approximations may achieve similar results in the third application.

1.2 Adaptive transform coding of signals and images

Almost all of the most efficient lossy signal[4] compression techniques available up to date are based on transform coding: they approximate a signal by a weighted sum of (usually orthonormal) base functions. The coefficients of the base functions in this weighted sum are quantized and coded and then stored or transmitted instead of the original sample values.

In practice, the coefficients are computed by applying a discrete orthogonal transform to the vector of the samples of the original signal, hence the name "transform coding". The base vectors of the discrete transform are the orthonormal sample vectors of the (approximately) orthonormal base functions. Properties of transform coding methods are often more easily interpreted in terms of continuous signals and base functions. Therefore, this document will stick to this continuous terminology. However, all the developed compression methods take the sampled nature of signals and images into account.

The most popular transform for coding signals and images is undoubtably the Discrete Cosine Transform (DCT). From a stochastical point of view, the DCT is close to optimal (in the sense that the bit rate for coding the quantized coefficients is the smallest possible for a given distortion) for coding signal segments or image blocks generated by a stationary stochastic process with only local correlation .[5] Even though most practical signals satisfy the local correlation assumption, they can hardly ever be considered stationary. Nevertheless, an ensemble of a large number of *segments* from a large number of signals usually is. This is because the random alignment of the signal w.r.t. the intervals washes away any non-stationarities. For this reason, the DCT usually is the best transform *on average* for coding a large number of signal segments or image blocks. However, the DCT is often *not* the best transform for coding a particular segment, because the local properties of the DCT base functions (e.g., local frequency and local amplitude) are time-independent (or position-independent in the case of images), while the corresponding local properties of a signal (e.g., local bandwidth and local amplitude) are not.

The previous discussion shows that there is no point in looking for a transform with better coding properties than the DCT, except in two cases:

1. when the transform, i.e., its base functions are allowed to differ from interval to interval. Indeed, even though no transform can outperform the DCT *on average* on all intervals, it is possible to find transforms which outperform the DCT on a specific interval. Clearly, such a transform is both signal- and interval-dependent.

2. when the signal segmentation, i.e., the division of the signal into intervals is made signal-dependent in such a way that the ensemble of signal segments has non-stationary statistics. This is usually only feasible in the case of quasi-periodical signals, where the intervals may be aligned with signal periods (or parts thereof). Note that in this case, the same transform may be used for coding an ensemble of

[4]The following discussion equally applies to images, which may be considered signals of two variables.

[5]This means that there is little or no correlation between signal samples or image pixels that are far apart (in time or position).

signals (e.g., all ECGs in a medical data base). However it may be more profitable to use a different transform for each signal, or even for each segment, because the non-stationary segment statistics may vary from signal to signal and from segment to segment.

Obviously, the success of these approaches depends on the possibility of efficiently describing the transform used to code each segment or each signal (if the same transform is used on all segments). The most straightforward approach for describing the transform consists of directly coding the base functions for each segment, but this is not feasible because coding a single base function requires about the same number of bits as coding a signal segment.

Clearly, what is needed is a "simple" transform, i.e., one which depends on a relatively small number of parameters so that only the parameters need to be stored for each segment. At the same time, the transform should have better coding properties than the DCT in order to achieve a significantly lower total bit rate (i.e., the sum of the coefficient bit rate and the parameter bit rate) than simple DCT coding. In other words, the transform should be flexible enough to adapt to each signal segment so that the number of coefficient bits for coding this segment (with a fixed distortion) is significantly lower than for the DCT.

Of course, flexibility and simplicity are conflicting demands, which need to be balanced to achieve the minimal total bit rate. Indeed, the coefficient bit rate will decrease with *increasing flexibility*, while the parameter bit rate will decrease with increasing simplicity, i.e., with *decreasing flexibility*. It is easily seen that neither the least flexible transform (i.e., a fixed transform such as the DCT) nor the most flexible transform (which spends almost all of the bits on coding its base functions and almost none on coding the coefficients) achieves the lowest bit rate.[6]

In addition to the above requirements, it is desirable that the base functions of the transform can be computed relatively quickly, given values for the transform parameters. Even more importantly, it must be possible to robustly and quickly determine optimal or at least close to optimal values of these parameters for a given signal segment. In general this highly dimensional (a number of 10 or 20 transform parameters is not unusual) optimization problem can only be solved satisfactorily and quickly by exploiting prior knowledge about the optimal solution. Clearly such knowledge can only be derived from theoretical insights on how the coding properties of the transform depend on the transform parameters and on some easily interpretable signal properties.

In the following, we will refer to the use of signal- and/or interval-dependent transforms as *adaptive transform coding*. It should be noted that in the literature, the term "adaptive transform coding" usually refers to signal-dependent coefficient quantization or coding and not to the use of a signal-dependent base. This is because signal-dependent bases have been studied only rarely and often with little success. Indeed, it is not easy to define transforms which satisfy the above requirements (simple, flexible, having good coding properties, quickly computable, easy to optimize and theoretically well-founded) and therefore only few papers have been published on the subject. Nevertheless, the author has achieved this goal by introducing the warped polynomial transforms [1], and by developing an elaborate theory on their coding properties [2] and an efficient algorithm for adapting them to a

[6] In fact, the most flexible transform uses a normalized version of the signal in a segment as its only base function and therefore needs to code only a single coefficient. However coding the base function is essentially equivalent to coding the original signal so that this most flexible transform is completely useless.

specific signal or image [4, 10]. The following sections describe these results in greater detail.

1.3 Overview

The following sections present an overview of the theory of warped polynomial signal approximation and its practical applications. Signal approximations are usually expressed as a linear combination of orthonormal base functions and so section 2 first introduces the orthonormal warped polynomials, which are the base functions used in warped polynomial approximations. It shows that these base functions are approximately amplitude and frequency modulated cosines and it relates their local frequency and amplitude to the warping function and weight function from which they are derived.

Next, section 2 demonstrates that warped polynomial approximation of a signal is equivalent to passing this signal through a low-pass filter with time-varying cut-off frequency. It then goes on to show that the number of base functions needed to represent a given signal depends on the relationship between the local signal bandwidth and the local frequency of the base functions and that it becomes minimal when both are proportional.

Section 2 also presents a practical algorithm for computing the optimal local frequency and the corresponding optimal base functions. Finally, it theoretically compares warped polynomial approximation to wavelet approximation and it discusses the connection between warped polynomial approximation and non-uniform sampling.

Section 3 presents some of the applications of warped polynomial signal and image approximation that have already proved successful. It starts with an ECG approximation method based on (non-warped) polynomial approximation, mainly to illustrate how the theory of section 2 can be used to explain the coding properties of existing transforms and to suggest improvements. Next, it discusses an adaptive warped polynomial ECG coding method, in which the warping function is adapted to the ECG. Section 3 also shows how this method may be used to adaptively filter not only ECGs, but also other signals (such as pneumotachs, electro-encephalograms and blood pressure signals).

Section 4 extends the theory of section 2 concerning "ordinary" warped polynomials to another kind of signal-adaptable orthonormal base functions, called "periodical" warped polynomials; as their name implies, the base functions of section 4 are periodical functions, and as such they are more suited for representing periodical signals than "ordinary" warped polynomials. Section 4 summarizes the main points of the theory of approximation by periodical warped polynomials and illustrates how it may be used in contour coding.

Finally, section 5 discusses a few possible extensions of the theory of warped polynomial approximation and a few possible new applications, and section 6 summarizes the main results of the author's research on warped polynomials.

2 The theory of orthogonal time-warped transforms

The base functions of orthonormal warped transforms are orthonormal *time-warped* or *spatially warped* polynomials. In this section, we first define these functions and we briefly describe their most important properties. Next we discuss some computational aspects, mainly to show that orthonormal warped polynomials can be computed quickly and accurately. Then, we discuss signal expansions in terms of warped polynomials: we describe their properties and we show how the warped polynomials can be adapted to a particular signal so as to obtain the most economical signal approximation. Finally, we discuss the connections of warped polynomial approximation to the theory of non-uniform sampling and sub-Nyquist uniform sampling (i.e., error-free uniform sampling at a rate lower than the Nyquist rate) and we compare the efficiency of warped polynomial approximations and wavelet approximations.

2.1 Definition and properties

The base functions of orthonormal time-warped transforms are orthonormal time-warped polynomials (OTWPs). These functions depend on two parameter functions: a *warping function* $\tau(t)$ and a *weight function* $w(t)$. The former is a strictly increasing function which maps the interval $[-1, 1]$ onto itself and the latter is a function strictly positive on $[-1, 1]$. More precisely, the OTWPs $\phi_n(t)$, $n = 0, 1, \ldots$ are the functions obtained by orthonormalizing the powers $\tau^n(t)$, $n = 0, 1, \ldots$ w.r.t. either the continuous inner product

$$\langle f \, , \, g \rangle = \int_{-1}^{1} f(t) g(t) \, w(t) dt \tag{1}$$

or w.r.t. the discrete inner product

$$f \cdot g = \frac{2}{L} \sum_{k=0}^{L} f(t_k) g(t_k) w(t_k), \text{ where } t_k = -1 + 2k/L. \tag{2}$$

The functions $\phi_n(t)$ which are orthonormal w.r.t. the discrete inner product are called discrete OTWPs and they are the ones used in practical applications. However, the continuously orthonormal OTWPs are easier to use in theoretical derivations and, as we will see later, in practical situations they very well approximate their discrete counterparts.

Many well-known and widely used transforms such as the discrete cosine transform (DCT), the discrete sine transform (DST) and polynomial transforms such as the discrete Legendre transform (DLT) are special cases of time-warped transforms. This is important, because it means that the theory of warped polynomial signal approximation (see section 2) directly applies to those transforms.

A first important result, derived in chapter 3 of [3] is that the continuously orthonormal OTWPs are approximately amplitude and frequency modulated cosines:

$$\phi_n(t) \approx A(t) \cos\Big(n\theta(t) + \Gamma(\theta(t))\Big), \tag{3}$$

where $\cos\Big(\theta(t)\Big) = \tau(t)$ and $A(t) = \sqrt{\frac{2}{\pi} \left| \frac{d\theta}{dt} \right| / w(t)}$. The phase function $\Gamma\Big(\theta(t)\Big)$ depends in a complicated way on $w(t)$ and $\tau(t)$, but it is generally a slowly varying function of θ. For large enough n, the local angular frequency $\omega_n(t) = 2\pi\nu_n(t)$ of $\phi_n(t)$ is approximately $n \left| \frac{d\theta}{dt} \right|$. The function $\left| \frac{d\theta}{dt} \right|$ is called the *frequency function*. Note that all base functions corresponding with the same $\tau(t)$ and $w(t)$ have the same local amplitude and that their local frequencies $\nu_n(t)$ are multiples of a fundamental local frequency $\nu_0(t)$.

The approximation 3 is the basis of most of the theoretical results concerning warped signal and image approximation. It relates the local frequency and the local amplitude of the base functions to the warping function and the weight function. Furthermore it provides a recipe for generating orthonormal base functions with a desired local amplitude

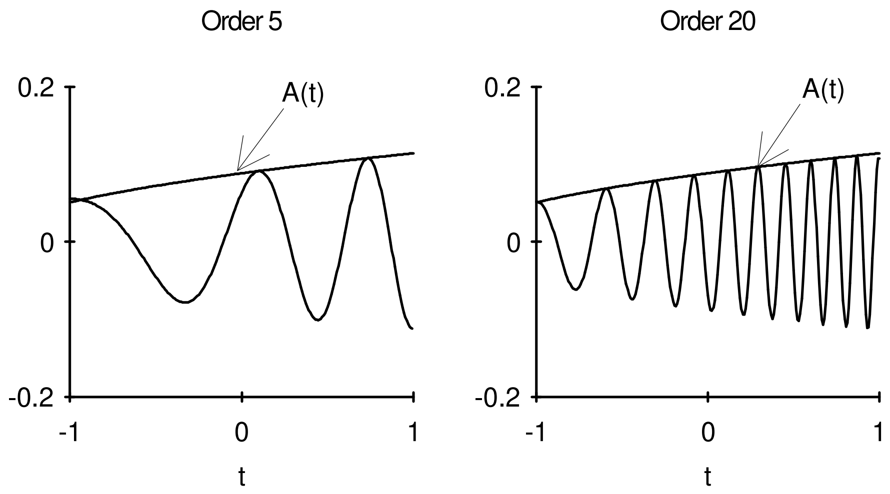

Figure 1: Orthonormal time-warped polynomials with linearly increasing frequency (and constant weight function) and their local amplitude $A(t)$. Near $t = -1$, the local frequency is five times higher than near $t = -1$.

and fundamental local frequency. Indeed, given $A(t)$ and $\nu_0(t)$ it is very easy to compute the corresponding values of $w(t)$ and $\tau(t)$, and then the corresponding base functions.

In principle, any desired local frequency and amplitude can be achieved, with one minor restriction: the fundamental local frequency $\nu_0(t)$ must satisfy the following equation for the corresponding $\tau(t)$ to be a valid warping function:

$$(1/2)\int_{-1}^{1} \nu_0(t) = 1/4. \tag{4}$$

However, this condition, which states that the time mean of $\nu_0(t)$ must equal $1/4$ is not very restrictive.

Note that in practice, $w(t)$ often cannot be chosen freely (e.g., $w(t)$ is often a constant). In this case only $\nu_0(t)$ can be specified, but this is not a severe restriction as $A(t)$ has a far smaller influence on the coding properties of the base functions than $\nu_0(t)$ [3].

Fig. 1 and fig. 2 illustrate how base functions with a specific local frequency (and with $w(t) = 1$) can be generated. In these figures, the desired local frequencies are linearly increasing and piecewise constant, respectively. The actual base functions indeed have the desired local frequency, which confirms the accuracy of the approximation (3).

A second important conclusion reached in [3] is that the discrete OTWPs differ very little from the corresponding continuously orthonormal OTWPs, as long as the order $n < cL$, where the number c is given by

$$c = \frac{\pi}{2} / \max_{t\in[-1,1]} \left|\frac{d\theta}{dt}\right| \tag{5}$$

and depends on the warping function only. In high-compression coding, signals are represented with a small number of base functions $\phi_n(t)$, $0 \leq n \leq N$, where $N \ll L$. There, the condition $n < cL$ is normally satisfied and there is often no need to make the distinction between the two kinds of OTWPs.

Sometimes, it is necessary to also use base functions with $n \geq cL$ (e.g., in warped image coding, see section 3.4). At the moment no good approximation is known for the discretely

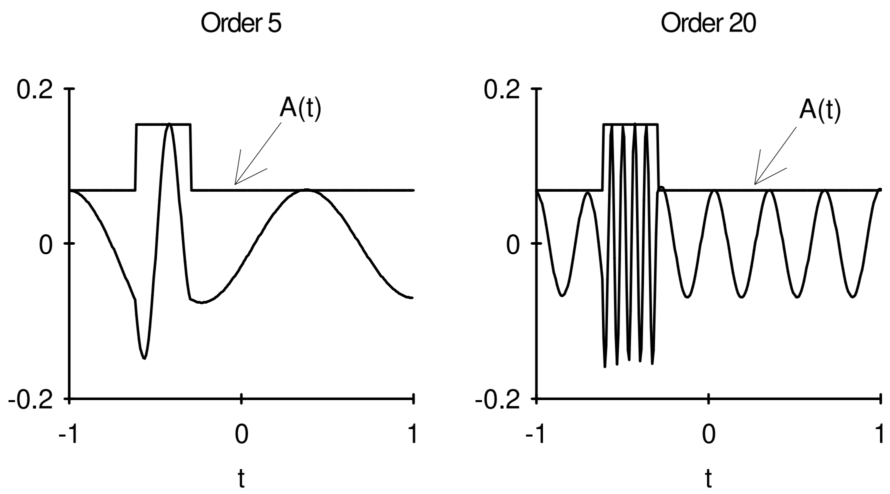

Figure 2: Orthonormal time-warped polynomials with piecewise constant local frequency (and constant weight function) and their local amplitude $A(t)$. In the central region, the local frequency is five times higher than in the extremal regions.

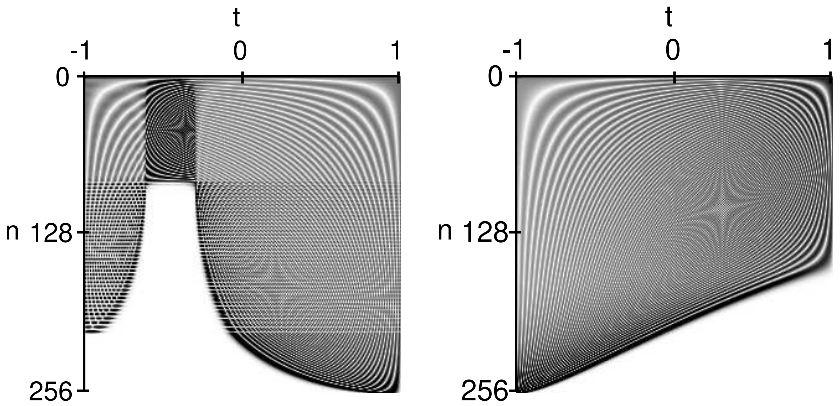

Figure 3: Absolute values of the base functions of fig. 2 and fig. 1 for all orders n. White corresponds to a low absolute value.

orthonormal OTWPs when $n \geq cL$. However, the author has demonstrated in [3] that when $n \geq cL$, aliasing occurs in those regions where $\nu_n(t)$ (i.e., the local frequency of the corresponding continuously orthonormal base functions) exceeds the Nyquist frequency $(L+1)/4$ (i.e., half the sampling rate). Due to this aliasing, the amplitude of the sampled base functions becomes very small (but not zero) in those regions.

The aliasing at high orders is illustrated in fig. 3, which displays the absolute values of the base functions of fig. 1 and fig. 2 for all orders. The vanishing of the local amplitude is clearly visible in fig. 3.

2.2 Computational aspects

The discrete OTWPs satisfy a three-term recurrence. Using this recurrence, they can be generated very quickly. This is especially important in adaptive coding, where the

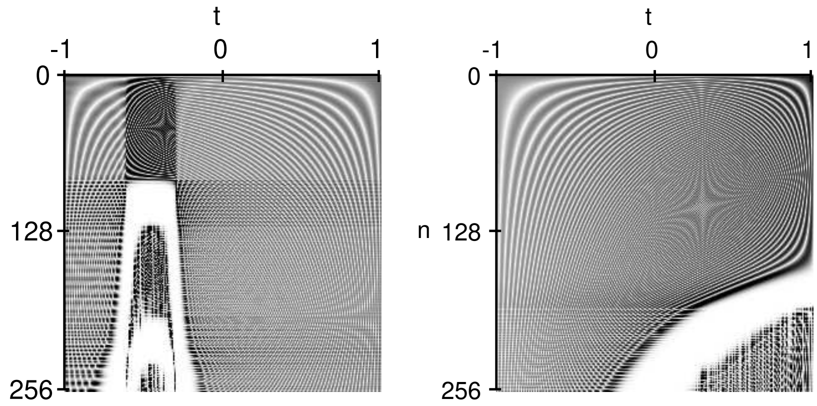

Figure 4: Absolute values of the base functions of fig. 2 and fig. 1 for all orders n, when the base functions are computed by the three-term recurrence. White corresponds to a low absolute value.

base cannot be pre-computed. An important question, which has not yet been resolved theoretically concerns the numerical stability of the recurrence. However, experimental results in [3] demonstrate that the recurrence is stable at least as long as $n \leq cL$. When n is increased above this limit, this remains true at first, but soon after stability is lost. Fig. 4 illustrates this behavior in the case of the base functions with linearly increasing and piecewise constant local frequency. Note that base functions of order n larger than cL can still be generated by the Gram-Schmidt orthogonalization procedure, if they are needed for some reason. This procedure is more stable, but it requires many more computations.

A numerical test, which detects the onset of numerical instability more precisely has been developed. However, this test is rarely needed because in practice n often remains smaller than cL and therefore the recurrence can be used safely.

Another important question, which has only recently been addressed, concerns the computational precision required by the recursion for generating the orthonormal base functions. The answer to this question is important in view of the implementation of the compression and filter algorithms on digital signal processors and application specific integrated circuits. Preliminary results have shown that this recursion can be implemented using both 32 bit fixed-point and 32 bit floating-point arithmetic. A 16 bit fixed-point implementation did not prove sufficiently accurate yet, but this implementation could be improved. Some more research is needed to quantify the errors resulting from a reduction in computational precision.

2.3 Warped polynomial signal expansions and their properties

Both warped polynomial signal coding and warped polynomial filtering are based on signal approximation by warped polynomials. Usually, the signals to be filtered or coded are divided into intervals which are approximated separately. We now concentrate on approximating a single such segment by a warped polynomial of degree N. For the moment we assume that the warping function $\tau(t)$ and the weight function $w(t)$ are given and that $w(t) = 1$ (the most popular choice). Furthermore —without loss of generality— we normalize the segment such that it corresponds to the time interval $[-1, 1]$.

Usually, the quality of a signal approximation $s_N(t)$ is measured by the (weighted) mean squared error (MSE) between $s_N(t)$ and $s(t)$. This MSE is defined as $\|s_N - s\|$, where the notation $\|f\|$ stands for the *discrete norm* $\|f\| \triangleq (f \cdot f)^{1/2}$ of the function $f(t)$. The optimal warped polynomial approximation $s_N(t)$ of a signal segment $s(t)$, $t \in [-1, 1]$ according to this least mean squares (LMS) criterion can be very easily expressed in terms of the orthonormal warped polynomials $\phi_n(t)$ associated with $\tau(t)$ and $w(t)$:[7]

$$s_N(t) = \sum_{n=0}^{N} A_n \phi_n(t), \tag{6}$$

where $A_n = s \cdot \phi_n$.

Except when $N = L$, i.e., when the number of coefficients equals the number of samples, $s_N(t)$ will not be exactly equal to $s(t)$. This is called the *truncation error.* Furthermore, in coding applications, the coefficients are quantized, i.e., represented with a limited accuracy and then coded, e.g., using variable-length words. This causes an additional *quantization error.* A detailed discussion on how these errors and the coefficient bit rate depend on the warping and weight functions can be found in [3, ch. 4]. The results show that the truncation error depends very strongly on $\tau(t)$, especially when N is much lower than L, but that it depends only mildly on $w(t)$. Furthermore, the results indicate that the total mean squared quantization error is basically independent of $\tau(t)$ and $w(t)$, but that $w(t)$ and $\tau(t)$ do influence the distribution of the quantization error in the approximation interval.

In conclusion, warped polynomial approximations corresponding to different warping functions may differ strongly in truncation error, but not in quantization error. For this reason, we will not discuss quantization effects in the remainder of this document. Instead, we will focus on the truncation error, in order to compare different warped polynomial transforms.

In most applications, the signal $s(t)$ to be approximated consists of a desired signal component $d(t)$ and an undesired noise component $n(t)$, i.e., $s(t) = d(t) + n(t)$. In that case, $s_N(t) = d_N(t) + n_N(t)$ where $d_N(t)$ and $n_N(t)$ are the approximations of $d(t)$ and $n(t)$, respectively. Usually, it is required that $d(t)$ is very well approximated, i.e., that the *signal distortion* $\|d - d_N\|$ is small. On the other hand, the *output noise energy* $\|n_N\|$ should be kept as small as possible. For any orthogonal transform, this output noise energy is proportional to N when the *input noise* $n(t)$ is white (which is often a reasonable assumption). In fact, assuming no distortion of the desired signal component $d(t)$, the output signal-to-noise ratio (SNR), defined as $\|d_N\|/\|n_N\|$, is $10\log(L/N)\,\mathrm{dB}$ lower than the input SNR $\|d\|/\|n\|$.

The previous discussion shows that it is desirable to keep N as small as possible to reduce the output noise. Furthermore, in coding applications N should also be kept low to minimize the number of coefficients and hence the bit rate. Consequently, a useful criterion for comparing different orthonormal warped transforms is the minimal order N with which an "almost error-free" approximation (i.e., an approximation for which $\|d - d_N\|$ is negligible) can be achieved.

[7] The fact that the optimal coefficients A_n can be computed so easily is one reason why expansions in terms of orthonormal base functions are preferred over non-orthonormal expansions.

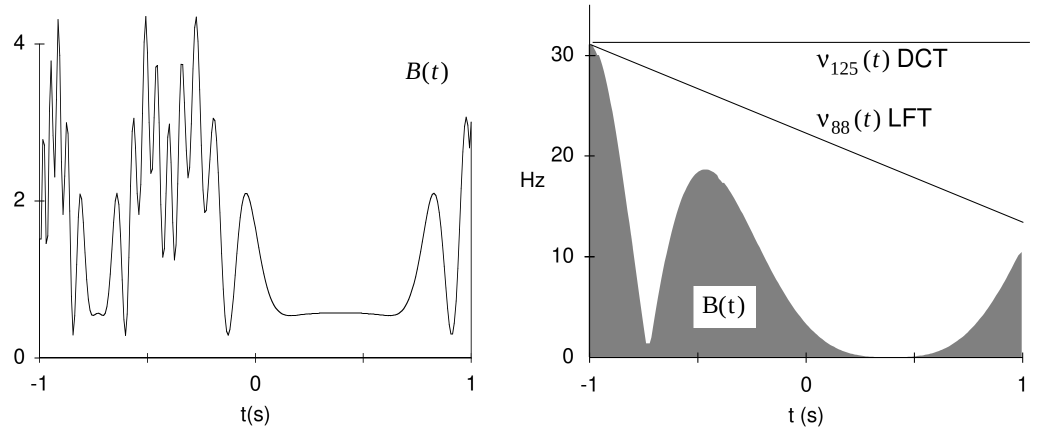

Figure 5: Almost error-free approximation of an artificial signal with minimal approximation order for two different transforms (the DCT and the LFT). Left: the actual signal. Right: its local bandwidth $B(t)$ (shaded area) and the local cut-off frequencies $\nu_N(t)$ of the transforms for $N = N_-$.

In [2,3], the author has shown that the approximation $s_N(t)$ can be interpreted as the output of an ideal non-stationary low-pass filter, to which $s(t)$ is applied as the input. The low-pass filter has a time-varying cut-off frequency $\nu_N(t) \approx N \left|\frac{d\theta}{dt}\right| /(2\pi)$ which equals the local frequency of the highest-degree base function in eq. (6). Almost error-free approximation occurs whenever this cut-off frequency exceeds the local bandwidth $B(t)$ of the desired signal $d(t)$ throughout the approximation interval, i.e., whenever

$$N \left|\frac{d\theta}{dt}\right| \geq 2\pi B(t), \qquad \forall t \in [-1,1]; \tag{7}$$

Clearly, it is possible to satisfy this condition for any transform, by choosing N large enough. Furthermore, the lowest possible N for which eq. (7) is satisfied is approximately equal to

$$N_- \triangleq \max_{t\in[-1,1]} 2\pi B(t)/\left|\frac{d\theta}{dt}\right|. \tag{8}$$

This very important result is explained in more detail in [2,3,18]. It is the basis for the comparison of different time-warped transforms, as illustrated by the following example: fig. 5 depicts the situation $N = N_-$ for an artificial signal and for two different transforms: the DCT for which $\nu_0(t)$ is constant and the linear frequency transform (LFT) for which $\nu_0(t)$ is linearly decreasing. Note that in the figure, N indeed equals N_-, because for both transforms N cannot be reduced without violating eq. (7). Also note that in figures like fig. 5, the area A enclosed by the curve $\nu_N(t)$ and the time axis has an important interpretation: it equals $N/2$ because eq. (4) implies that

$$\int_{-1}^{1} \nu_N(t) = N/2. \tag{9}$$

Therefore fig. 5 shows graphically that N_- is lower for the LFT than for the DCT, i.e., that the LFT is better suited to code (or filter) this signal than the DCT. Actually, in this example N_- equals 88 for the LFT and 125 for the DCT.

The previous example is somewhat artificial, but it illustrates how the theory can be used to explain why certain transforms perform better (have a lower value of N_-) than others on certain signals, even when the local signal bandwidth is only known qualitatively. In particular, theoretical insights like this have lead the author to the conclusion that the discrete Legendre transform (DLT) is better suited for coding (and filtering) so-called R-R-segments (see fig. 9 and fig. 11) of electrocardiograms (ECGs) than the DCT which was proposed earlier for this purpose [3, 4].

When N is lower than N_-, i.e., lower than the minimal approximation order for almost error-free approximation, different kinds of errors are introduced in the approximated signal $d_N(t)$. The filter interpretation allows these errors to be easily studied and categorized. For instance, it is easily seen that narrow peaks in the signal are widened and lowered in amplitude if they occur in regions where $B(t)$ exceeds $\nu_N(t)$. Furthermore, the widening of such peaks causes ringing or Gibbs effect, i.e., oscillations with large extent and with a local frequency of about $\nu_N(t)$. Low-pass and ringing effects are examined in greater detail in [2, 3]. One interesting conclusion is that the amplitude of the Gibbs effect in a given region depends not only on the warping function but also on the weight function, which means that Gibbs effect can be reduced by introducing a suitable weight function (see [2, 3] for more details).

2.4 Optimal warped polynomial signal expansions

2.4.1 Theory

In section 2.3 we already mentioned that the optimal time-warped transform for approximating a signal segment is the one with the lowest N_-, i.e., the one which achieves almost error-free approximation with the lowest possible approximation order.

Using eq. (8) and eq. (4), it is easily seen that N_- becomes minimal when the fundamental local frequency $\nu_0(t)$ of the base functions is proportional to the local signal bandwidth $B(t)$. This can also be deduced using the "area" interpretation of N (see previous section). Indeed, the area A underneath any local cut-off frequency $\nu_N(t)$ can be reduced without violating the condition (7) for error-free approximation, by lowering $\nu_N(t)$ until $\nu_N(t) = B(t)$ for all $t \in [-1, 1]$. When this situation is achieved, N_- equals[8]

$$N^* \triangleq 2 \int_{-1}^{1} B(t)\, dt \tag{10}$$

according to eq. (9) and $\nu_0(t)$ equals

$$\nu_0^*(t) = B(t)/N^*. \tag{11}$$

It is easily seen that no transform can have a value of N_- lower than N^*, without violating eq. (7). This leads us to conclude that the frequency function of the optimal transform is given by eq. (11) in which N^* is given by eq. (10); furthermore, only the local bandwidth $B(t)$ must be known to compute the optimal frequency function.

The difference in N_- between a fixed (non-signal-adaptive) transform such as the DCT and the optimally warped transform (OWT), for which $\nu_0(t)$ is proportional to $B(t)$ can

[8]The attentive reader will notice that N^*, as defined by eq. (10), is not an integer as it should be. In fact, the correct value of N^* is the smallest integer larger than the right-hand side of eq. (10). In order not to overcomplicate the notations, we will make abstraction of this.

be quite considerable. For example, for the signal in fig. 5, N^* equals 22. This means that for the OWT $N_- = N^* = 22$, which is much lower than the corresponding value for the DCT (which is 125, see above). Therefore, the OWT needs 5.7 times fewer coefficients than the DCT to store the signal. Obviously, this is important in coding applications; the lower value of N_- is also important in filter applications, because the signal-to-noise ratio of the OWT approximation is $10\log(125/22)\,\mathrm{dB} = 7.5\,\mathrm{dB}$ better than that of the DCT approximation (see section 2.3).

2.4.2 A practical algorithm

The previous theory on optimal warped approximations provides valuable insights and implies that the optimal transform for coding or filtering a signal may be determined by computing $B(t)$. However, this computation poses a number of practical difficulties, which are not easily overcome: the local bandwidth $B(t)$ of the desired signal $d(t)$ is usually determined by thresholding the local spectrum of the (noisy) signal $s(t)$, which is in turn computed using a windowed Fourier transform. The computed value of $B(t)$ therefore depends on the input SNR, on the selected threshold and on the selected window function. Unfortunately, there are no clear guidelines for optimally selecting these parameters because it is simply not clear what "optimal" means in this context.

To avoid these difficulties, the author has developed a quickly converging iterative algorithm [4,10] for computing the optimal (or at least a close to optimal) warping function for a given signal. This algorithm does not compute $B(t)$ but rather attempts to minimize the mean squared error (MSE) $||s_N - s||$ between $s(t)$ and $s_N(t)$ by iteratively updating the frequency function $\nu_0(t)$, while keeping N fixed. In other words the algorithm attempts to minimize the functional $\mathcal{F}_N(\nu_0)$, which associates to each warping function $\nu_0(t)$ the corresponding value of $||s_N - s||$.

It can be shown that $\nu_0^*(t)$, defined by eq. (11) is the only frequency function that globally minimizes $\mathcal{F}_N(\nu_0)$ when N equals N^*. When $N > N^*$, there may be more than one frequency function that globally minimizes $\mathcal{F}_N(\nu_0)$; in this case, all of the optimal frequency functions satisfy eq. (7). In other words, the algorithm always produces a signal approximation without signal distortion when $N > N^*$ and it produces the optimal approximation when $N = N^*$.

To compute the optimal approximation, the value of N^* must be known. In practice N^* can sometimes be estimated by exploiting theoretical knowledge about the signal. For instance, ECG signals consist of different waves, the typical bandwidths and durations of which are well-known. In conjunction with eq. (10) this knowledge can be used to estimate N^*. Alternatively, N^* can be estimated by running the algorithm for different values of N and examining the corresponding minimal value of $||s_N - s||$; this value increases rapidly when N drops below N^* and slowly decreases for $N \geq N^*$.

In practice, it is not necessary to accurately determine N^*. Indeed, a suboptimal approximation with $N > N^*$ may be quite acceptable in many applications, provided that N is significantly smaller than the value of N_- for the DCT. For many practical signals (e.g. ECGs), it is not difficult to find a value of N which satisfies this requirement because of the large difference between N^* and the value of N_- for the DCT. In such a situation, it is sufficient to determine a reasonably tight upper bound for N, which is of course much easier than accurately estimating N^*

Even when $N < N^*$, the algorithm may be useful. This situation arises, e.g., in signal

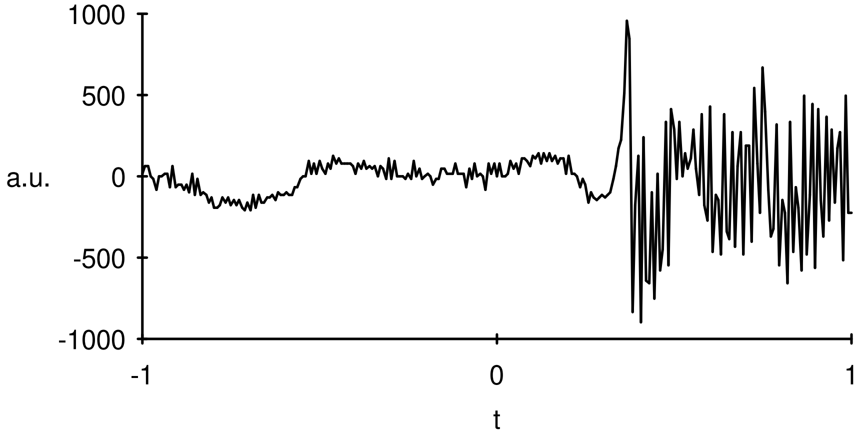

Figure 6: A segment of a speech signal.

coding, where N may be limited to achieve a very low bit rate at the expense of some signal distortion. In this case, the algorithm attempts to minimize the total signal distortion, which is a desirable goal.

Note that the algorithm can also deal with parameterized frequency functions or warping functions. For instance, in coding applications, it may be desirable to use frequency functions which depend on a few parameters only (see section 3.2 for more details). In this case, the algorithm converges to those frequency function parameters which minimize the mean squared approximation error.

2.4.3 Example

The following small example illustrates both the operation of the iterative algorithm, and its potential: fig. 6 shows a segment of a speech signal, which is to be approximated using $N + 1 = 128$ transform coefficients. At the start of the first iteration of the algorithm, the frequency function is constant, and the corresponding approximation is the DCT128 approximation of fig. 7 (which is clearly inadequate, by the way).

After convergence, the frequency function has adapted to the local bandwidth of the signal: the frequency is now much higher on the right (see fig. 7) and consequently, the right-hand portion of the signal is now reconstructed almost perfectly (see the error signal in fig. 7). On the left, the approximation error has not changed much. This because the signal on the left is rather noise-like and (probably) cannot be approximated by any transform using only 128 coefficients.

Overall, the quality of the approximation has been much improved by the algorithm; in fact, the RMS approximation error has been reduced by a factor of 7.4 (see [3, p. 150]). Note that fig. 7 does not inform us on whether the final approximation is of sufficient auditory quality. However, the only purpose of the example is to illustrate the use of the iterative algorithm. Nevertheless, the results are sufficiently interesting to warrant further investigation of speech applications.

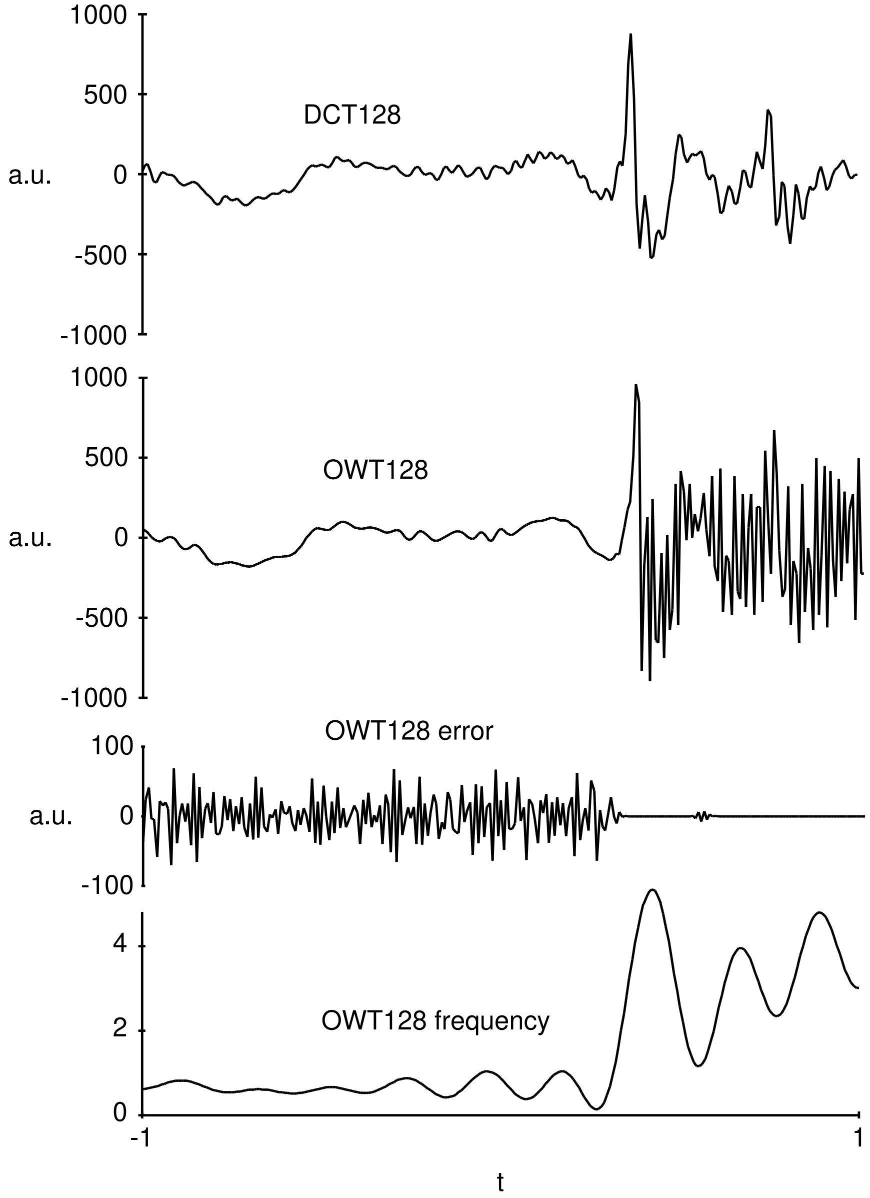

Figure 7: Top: the warped polynomial approximations of the signal in fig. 6 before (DCT128) and after (OWT128) adapting the frequency function. Middle: the error signal of the final approximation. Bottom: the final frequency function.

2.5 Comparison to wavelet coding

The adaptive compression method of this section adapts to the local signal bandwidth by changing the fundamental frequency of its base functions. However, the time-varying nature of the local bandwidth may also be exploited by using base functions which are well localized in time and frequency. A typical example of such base functions are wavelets. In this case, adaptation to the time-varying signal spectrum is not achieved by adapt-

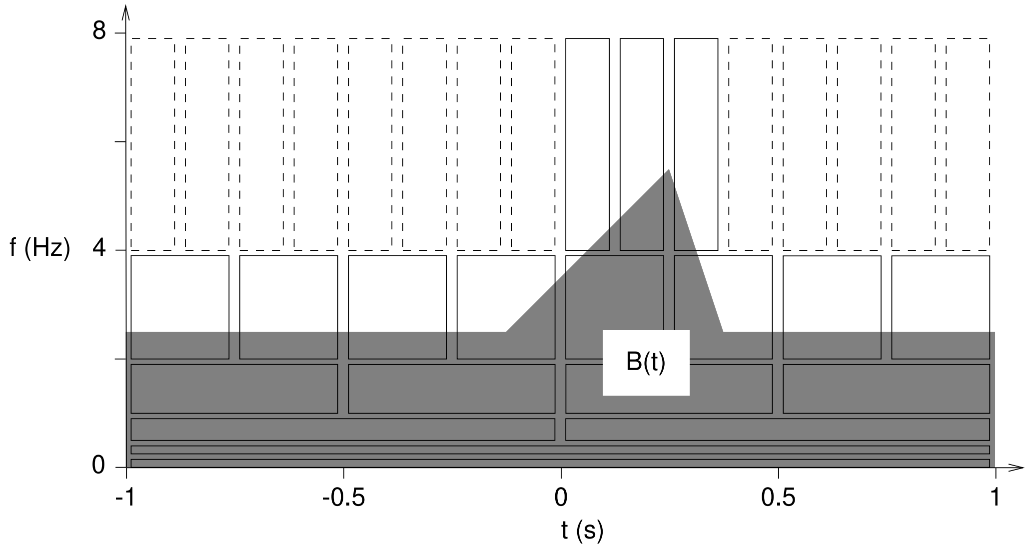

Figure 8: Wavelet coding of a signal with local bandwidth $B(t)$ (indicated by the gray area). Each rectangle represents a single wavelet in time-frequency space. Only the 18 solid rectangles need to be used to code the signal almost error-freely. To code the same signal with optimally adapted orthonormal warped polynomials, only 13 base functions would be needed.

ing the base functions to the signal, but rather by selectively using wavelets in a signal approximation (see below).

At this moment, no detailed comparison of warped polynomial approximation and wavelet approximation is available. However, in this section, we present some theoretical arguments which suggest that the former may be more efficient in the sense that they require fewer coefficients to represent a give signal. In order to explain this, we will now graphically derive the wavelet equivalent of eq. (10), which relates the minimum number of coefficients for almost error-free warped polynomial approximation to the local signal bandwidth.

In order to do this, we first repeat a few important properties of wavelet decompositions (albeit with a great deal of simplification):

1. In order to represent a signal of $L + 1 = 2^k$ samples with a full wavelet base, k groups of wavelets G_j, $j = 0 \dots k-1$ are needed. For $j > 0$, each group G_j consists of 2^j wavelets, which are all translated versions of one another and which are band-pass functions. Group G_0 on the other hand consists of 2 wavelets, one of which is band-pass and the other low-pass.

2. All band-pass wavelets are approximately scaled (and translated) versions of one another.[9] For this reason all of these wavelets have approximately the same duration-bandwidth product.

[9] Actually this is a gross simplification, especially for the wavelets in groups with high indices.

3. The duration of each wavelet in group G_j is approximately 2^{-j}, and its bandwidth is approximately 2^{j-1} when the approximation interval is $[-1, 1]$.

These properties are much more easily visualized in a time-frequency plane. In this plane, each wavelet is represented by the rectangle in the time-frequency plane in which the local spectrum of the wavelet is non-negligible. Fig. 8 shows an example of a wavelet decomposition in the time-frequency plane for $L + 1 = 32$. Graphically, the properties above may be interpreted as follows:

1. There are k different kinds of rectangles, each kind having a different aspect ratio. There are 2^j rectangles of the j-th kind, except for $j = 0$.

2. All rectangles have the same area, because this area equals the duration-bandwidth product which is the same for all wavelets. The area is $1/2$, when the approximation interval is $[-1, 1]$

3. The width of a rectangle of the j-th kind equals 2^{-j}, and its height 2^{j-1}.

Now suppose that a signal with local bandwidth $B(t)$ is represented by the wavelet expansion

$$s(t) = \sum_{j=0}^{k-1} \sum_{i=0}^{c_j} A_{i,j} \Phi_{i,j}(t), \tag{12}$$

where $c_j = 2^j - 1$ for $j > 0$ and $c_0 = 2$ and where $\Phi_{i,j}(t)$ denotes the i-th wavelet of group G_j. Intuitively, it is clear that wavelets $\Phi_{i,j}(t)$ whose local frequency spectrum has no components in the area $0 \leq f \leq B(t)$, $-1 \leq t \leq 1$ will not appear in such an expansion, i.e., that the corresponding coefficient $A_{i,j}$ will be negligible.[10] Also, assuming that the local spectrum of $s(t)$ is low-pass, all other coefficients will be non-zero in general.

Consequently, the minimum number N_w of wavelets needed to approximate $s(t)$ almost error-freely equals the minimum number of wavelet-rectangles needed to completely cover the area $0 \leq f \leq B(t)$, $-1 \leq t \leq 1$ in the time-frequency plane. In the example of fig. 8, this area is shown in gray and the wavelet-rectangles required to cover it are shown in solid line.

Note that in practice, it is necessary to specify which wavelets actually appear in the signal expansion eq. (12) (unless, all coefficients $A_{i,j}$ are coded, even those which are zero), or in graphical terms, which rectangles are solid in fig. 8. Actually, this data has the same purpose as the transform parameter data in adaptive warped polynomial coding: it provides information about the local signal bandwidth $B(t)$.

Using the previous results, we are now able to compare the efficiency of optimal warped polynomial and optimal wavelet expansions. To this end, we first note that the area A of time-frequency space covered by the wavelets with non-zero coefficients is usually significantly larger than the area A' underneath $B(t)$; in any case, A cannot be smaller than A'. Now, according to eq. (10), A' equals $N^*/2$ where N^* is the minimal approximation order required for almost error-free warped polynomial approximation of $s(t)$. On the other hand, A' equals $N_w/2$, because the area of each wavelet-rectangle is $1/2$. Therefore we may conclude that N_w is usually significantly larger than N^*. In other words, a

[10] Indeed, if this were not the case, then the local bandwidth of the right-hand side of eq. (12) could not be $B(t)$, and eq. (12) would not hold.

wavelet approximation is usually less efficient than the corresponding warped polynomial approximation. In fact a detailed examination of fig. 8 shows that in the example N_w equals 18, while N^* equals 13.

The previous discussion also gives a simple interpretation for the relative inefficiency of wavelets: covering the gray area in fig. 8 with "rectangular" wavelets takes up more area in time-frequency space than covering the gray area with more malleable orthonormal warped polynomials. This is another beautiful example of the power of a signal approximation theory based on local bandwidth (as opposed to one based on stochastics). Of course, the above results remain to be validated experimentally, and should therefore be used with great caution. Nevertheless, preliminary results [35] seem to confirm the theoretical results (see section 3.2.3).

2.6 Connection to non-uniform sampling and sub-Nyquist sampling

2.6.1 Non-uniform sampling

Shannon's sampling theorem states that it is possible to reconstruct a band limited signal from its uniformly taken samples provided the sampling frequency is larger than $2B$, where B is the signal's bandwidth. Therefore, about $2B\Delta$ samples are needed to represent a band limited signal with duration Δ and bandwidth B.

Not surprisingly, there also exists a "local" sampling (see [3] for more details): if the signal is locally band limited to $B(t)$, then it can be (approximately) reconstructed from its non-uniformly taken sample values, provided the local sampling frequency $\nu(t)$ satisfies $\nu(t) \geq 2B(t)$ for all t. In this case, the minimum number of samples required to represent a signal segment I equals $N_s \triangleq 2\int_{t\in I} B(t)\,dt$. Interestingly, when $I = [-1, 1]$ this number exactly equals N^*, the minimum number of coefficients for almost error-free warped polynomial approximation. Furthermore, the optimal sampling frequency $\nu(t)$ is proportional to the fundamental frequency $\nu_0(t)$ of the optimal warped transform.

Of course, the close relationship between non-uniform sampling and warped polynomial approximation is no coincidence [3, ch. 2] and was noted before by other researchers (even though no practical methods for warped polynomial approximation existed). Up till now, the theory of *signal-adaptive* non-uniform sampling has not produced many practical applications, because of the difficulties involved in determining the local signal bandwidth and in non-uniformly resampling the signal (which requires some sophisticated interpolation). However, using the theory of warped polynomial approximation, this problem is now solved: indeed, $\nu_0(t)$ and hence $\nu(t)$ can be easily computed using the algorithm of section 2.4.2. The theory of warped polynomial approximation also solves the interpolation problem because the warped polynomial approximation $s_N(t)$ may be computed easily for values of t which do not coincide with the original uniform samples [3].

2.6.2 Sub-Nyquist uniform sampling

The previous discussion demonstrates that non-uniform sampling is an area of great potential for the theory of warped polynomial approximation. However, this theory also leads to some new interesting results in connection with uniform sampling: it shows that (at least theoretically) signals may be recovered from their uniformly spaced sample values, even when the sampling frequency is lower than the Nyquist rate $2B$, provided that

- the sampling frequency is higher than twice the mean local bandwidth.
- the local bandwidth $B(t)$ is known.

It can be shown that the minimal sampling rate of this theorem is usually much lower (and never higher) than the Nyquist rate [3]. Therefore, the theorem implies that sampling at a sub-Nyquist rate without loss of information is possible, *provided the local signal bandwidth is known.*

In practice, the theorem has very few applications. This is because any practical algorithm for reconstructing the signal from its sub-Nyquist sample values is numerically unstable (this can be shown using the theory of warped polynomial approximation). However, sub-Nyquist sampling can be made to work in practice, by appropriately distorting the sub-Nyquist samples, while they are taken. Actually, the kind of "approximate" sub-Nyquist sampling is the basis of the warped image coding method of section 3.4.

We end this section by showing that is not difficult to understand why the sub-Nyquist sampling theorem holds. Indeed, a warped polynomial approximation $s_N(t)$ is basically a polynomial function of the warping function: $s_N(t) = p_N(\tau(t))$. Now, like any other polynomial of degree N, $p_N(x)$ is completely determined by *any set* of $N+1$ sample values $p_N(x_k)$, $0 \leq k \leq N$, the only restriction being that all x_k should be different. In particular the sample instants x_k may be chosen as $x_k = \tau(t'_k)$, where t'_k, $0 \leq k \leq N$ are $N+1$ uniformly spaced sample instants in the interval $[-1, 1]$. In that case, the samples $p_N(x_k)$ are uniformly taken samples of $s_N(t)$.

Now suppose that $\tau(t)$ is the warping function of the optimal warped transform for the signal $s(t)$ and suppose that $N = N^*$. In this case $s(t) \approx s_N(t)$, to within very good approximation and the above results imply that $s(t)$ may be reconstructed in the interval $[-1, 1]$ from $N^* + 1$ uniformly taken samples. As the length of this interval is 2, the corresponding sampling frequency is about $N^*/2$, which according to eq. (10) is indeed twice the mean local bandwidth.

3 Applications of orthogonal warped transforms

In this section we describe some of the practical applications of warped polynomial signal and image approximations. We focus mainly on the results and we do not describe the different compression and filter methods in great detail. All of the results in this section have been published in open literature, and the interesting reader is referred to the relevant publications for more details.

3.1 ECG coding with discrete Legendre polynomials

In the past, the coding properties of transforms have been evaluated from a stochastical point of view: the signal was described by a stochastical model, and the efficiency of the transform was evaluated by computing the expected entropy of the transform coefficients for signals generated by this model. However, usually this approach does not provide much insight. Indeed, only in very special cases, e.g., when the signal's statistics are stationary it is possible to compute the efficiency of a few transforms (e.g., the DCT and the discrete Fourier transform) analytically.

Consequently, it is not surprising that when researchers experimented with different transforms, they could determine which transform performed best on a given signal, but

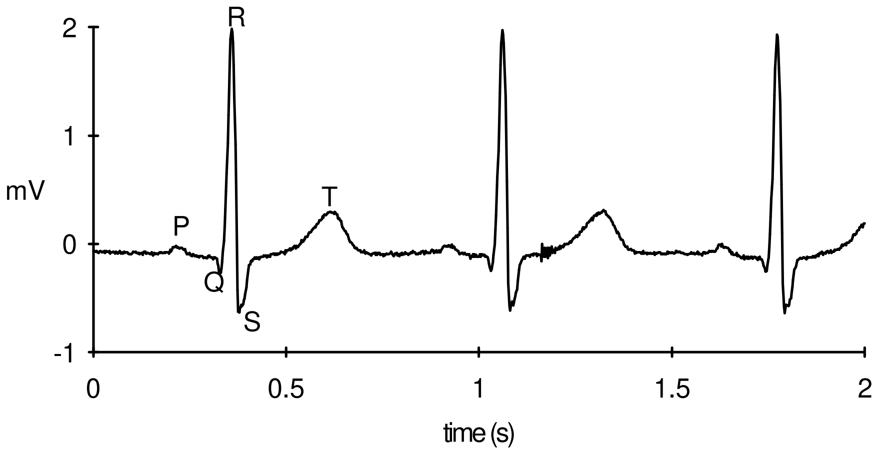

Figure 9: An original ECG signal.

they could not explain why, except in vague terms.[11] More importantly, this lack of understanding prevented them from suggesting better transforms for a given signal. In contrast, the theory of section 2.3 explains the coding properties of a (warped) transform in terms of a very simple signal property (local bandwidth) and a very simple property of the transform (its frequency function).

We now demonstrate how this theoretical insight may be used to improve coding schemes, using ECG coding as an example. Fig. 9 shows an original ECG. In order to code such a signal using an orthogonal transform, it has to be divided in intervals. The most logical choice is to use intervals which correspond to one ECG period, because for this choice the coefficients will change little from interval to interval, which can be exploited to code them more efficiently. However, on such intervals, traditional transforms such as the DCT and the DFT require many coefficients to accurately represent the signal. For instance, fig. 10 demonstrates that it is not possible to adequately approximate the ECG of fig. 9 using only 31 DCT coefficients. Earlier findings by other researchers indicate that this is also true for the DLT. However, as we will see below, their bad results are due to the fact that (for aesthetic reasons) they aligned the approximation intervals such that the most striking signal features (the QRS-complexes, see fig. 9) are located centrally.

The theory of section 2 easily explains the failure of the earlier DLT methods; it shows that a good approximation with few coefficients is possible whenever the local frequency $\nu_0(t)$ of the transform matches the local signal bandwidth $B(t)$, in the sense that $\nu_0(t)$ is large or small whenever $B(t)$ is and vice versa (ideally, they should be proportional). Also, the theory shows that the local frequency of the DLT transform (for which $\tau(t) = t$), is large near the edges of the approximation interval and low in the center of such an interval. On the other hand, the local bandwidth of an ECG signal is large near the QRS-complexes and low elsewhere. Therefore this bandwidth matches the local frequency only on an R-R interval (because only then the high-bandwidth regions are located near the interval boundaries). This not only explains the failure of earlier DLT methods but also suggests a remedy: segment the ECG in R-R intervals.

[11] For instance, one researcher examined the use of polynomial approximation for ECG coding, he found that it didn't work very well and then commented that "high-degree polynomials are not suited to code ECGs because they have a tendency to wiggle."

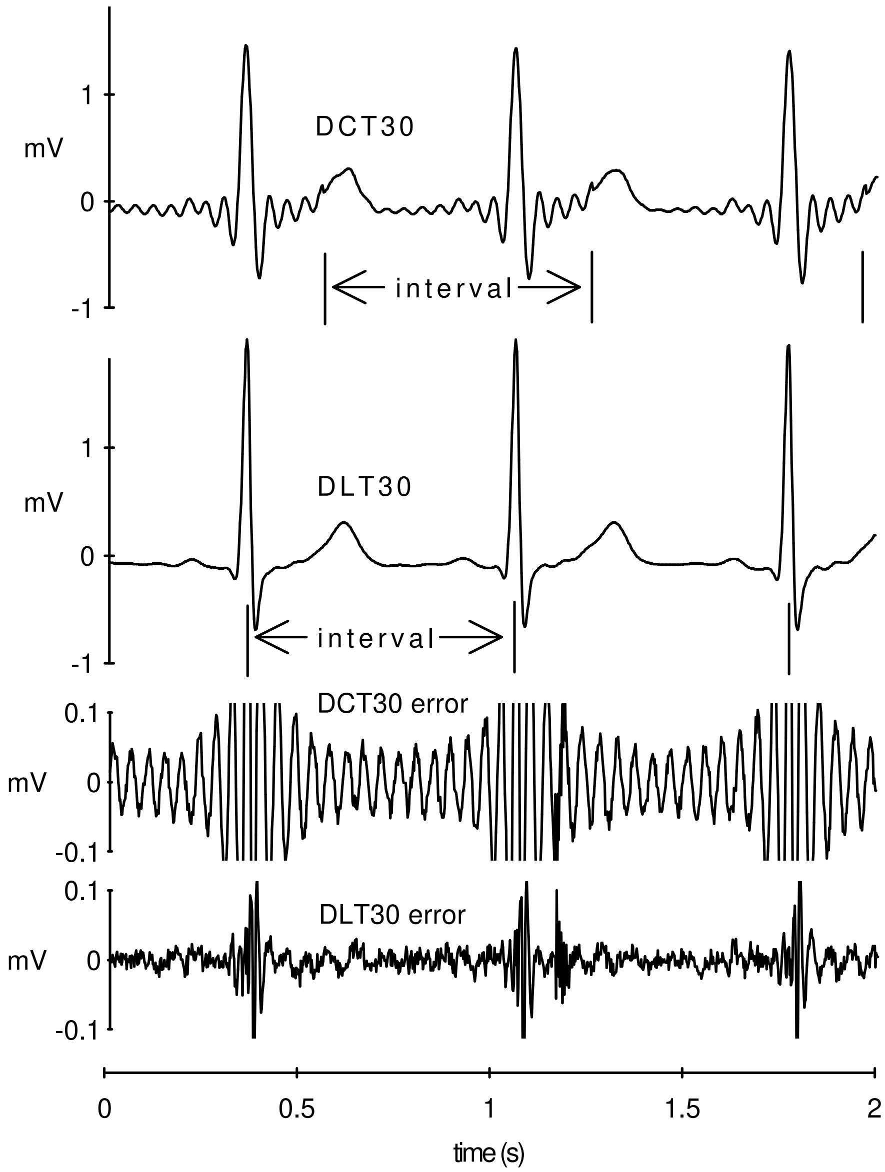

Figure 10: Top: DCT and DLT reconstructions of order 30 (quantized coefficients) of the ECG signal in fig. 9. Bottom: the corresponding error signals.

Fig. 10 shows that the proposed solution works: the DLT approximation is quite faithful (though not perfect). In fact, entropy-coding the quantized coefficients requires about 240 bit/s, which corresponds to a compression ratio of 21 [3, p. 126]. In contrast, the DCT approximation in fig. 10 requires a significantly higher bit rate (310 bit/s, compression ratio 16) even though it is of much worse quality.[12]

The reader may wonder whether an R-R interval choice may similarly improve the performance of the DCT. The theory predicts that this is not the case, because the frequency function of the DCT is constant. Experimental results [3, p. 127] confirm this.

[12]In fact, this can also be explained by the theory; see [3, ch. 4].

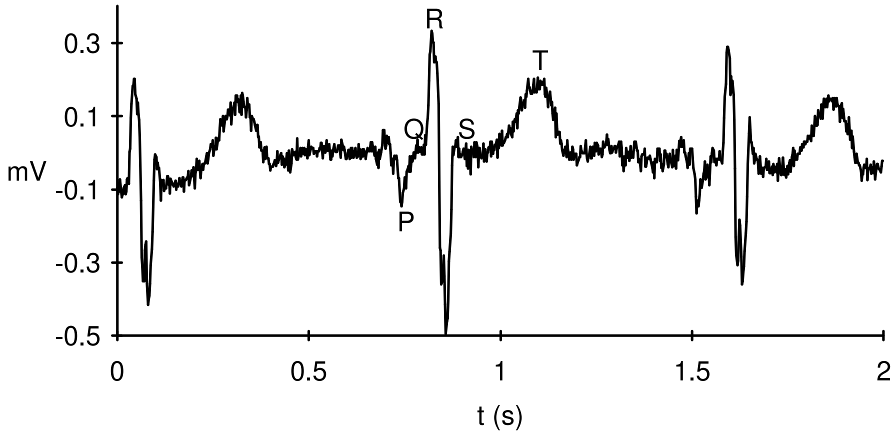

Figure 11: A noisy original ECG signal.

Fig. 11 shows a second ECG which is much noisier than the ECG of fig. 9. The new ECG is also one of the most difficult to code (among many examined ECGs), because it has a complicated waveform. Consequently, it is not surprising that an adequate approximation of this signal requires at least 50 coefficients, even when using the DLT. Fig. 12 shows the DCT and DLT approximations of order 50 of the new ECG. Again, the DLT approximation is of much better better quality. However, even the DLT approximation is not perfect: it is somewhat noisy, and some of the R-S waves have been slightly distorted.

More results in [15, ch. 5] show that the DLT requires only 30 to 50 quantized coefficients per R-R interval, to accurately represent most ECGs. Typically, ECGs can be coded at 300 bit/s using the DLT; corresponding DCT reconstructions are distorted by the Gibbs effect, even when using 50 to 100 coefficients (which corresponds to a rate of 400 to 500 bit/s).

3.2 Adaptive coding of biomedical signals

3.2.1 A practical adaptive coding method

The theory of section 2.4 clearly demonstrates that signals may be very efficiently represented using the optimal warped polynomial transform, in the sense that only few coefficients are needed to represent (i.e., code) the signal. Of course, the optimal transform for coding a signal interval depends on the signal in that interval. This makes it necessary to store a description of the transform used in each interval along with the coefficient data. Clearly, the bit rate required for this purpose must be kept small, in order not to nullify the reduction in coefficient bit rate over a fixed transform. In adaptive warped polynomial signal coding, this requirement is met by using frequency functions which depend on a small number of parameters.

Typically, either the warping function or the frequency function is restricted to be a polynomial of low degree, which is expanded as a linear combination of orthonormal polynomials. The coefficients in this expansion are quantized and coded. In essence, this means that the warping function (or the frequency function) is coded just like the signal itself, but using a fixed transform.

Approximating the frequency function by a low-degree polynomial is usually justified

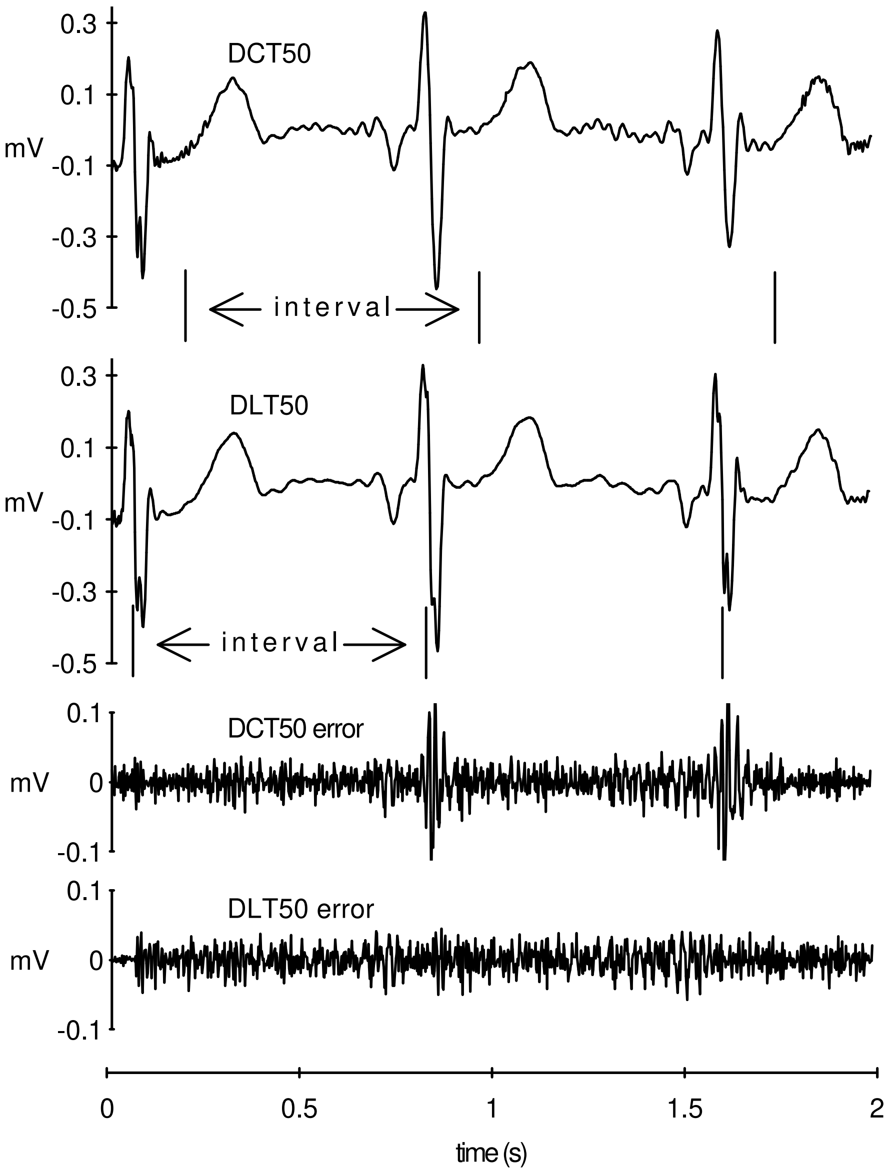

Figure 12: DCT and DLT approximations of order 50 of the ECG in fig. 11 and the corresponding error signals.

in practice. This because for many signals, the local bandwidth and therefore the optimal frequency function can be well approximated by such such a slowly varying polynomial function. Furthermore, even if this is not the case, it is often quite easy to find a close to optimal frequency function which is slowly varying.[13]

[13]Indeed, it is always possible to find a slowly varying polynomial $p(t)$ which is an upper bound of $B(t)$ for all t. Now, if this upper bound is close in the sense that $2N' \triangleq \int_{-1}^{1} p(t)\,dt$ is not much larger than $2N^* \triangleq \int_{-1}^{1} B(t)\,dt$, then the frequency function $\nu_0'(t) \triangleq p(t)/N'$ is close to optimal; this is because the corresponding N_- equals N' and is therefore not much larger than N^*.

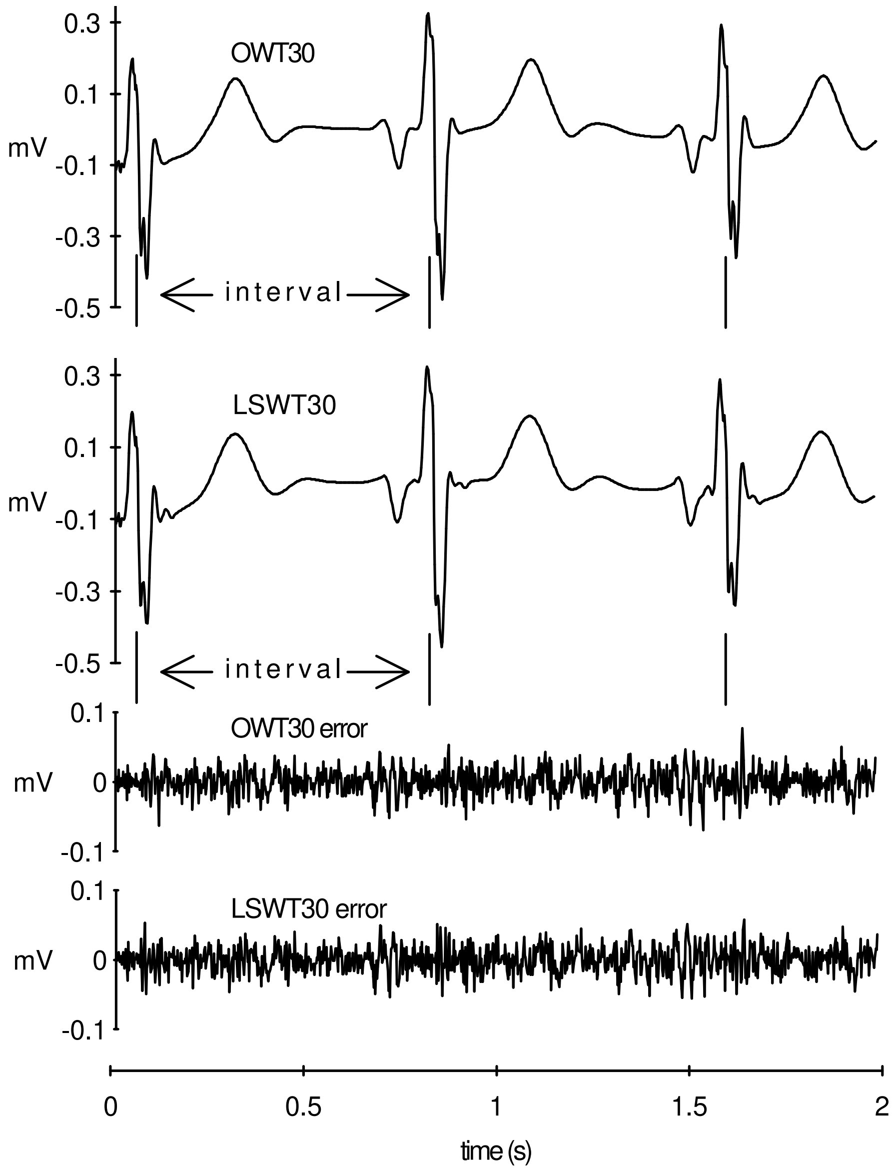

Figure 13: OWT and LSWT approximations of order 30 of the ECG in fig. 11 and the corresponding error signals.

Before proceeding with some examples, we wish to point out a peculiar fact: when the warping function is a polynomial the actual signal approximation $s_N(t)$ is also a polynomial in t (because it is a polynomial function of the warping function). Therefore, the coding method may be viewed as a polynomial approximation method, albeit a very special one. Indeed, the polynomial degree of $s_N(t)$ is much higher than N. For instance, when the warping function is of degree M, $s_N(t)$ is a polynomial of degree MN in t. Also, the polynomial coefficients of $s_N(t)$ are not independent; indeed, they are uniquely determined by only $M + N + 2$ numbers, i.e., the $N + 1$ coefficients A_n in eq. (6) together with the $M + 1$ polynomial coefficients of the warping function.

3.2.2 Some experimental results

The adaptive coding method with a polynomial warping function has already been successfully applied in ECG compression [4]. At a given compression ratio, it provides an even better signal quality than the DLT-technique mentioned earlier. It allows ECG signals to be coded with very high accuracy using only about 250 to 300 bit/s (depending on the heart rate). More importantly, the adaptive method is very insensitive to errors in QRS-detection (e.g., see [3, p. 137]). The DLT method, on the other hand, only performs well on R-R-segments and is therefore very sensitive to segmentation errors. The robustness of the adaptive method is very important, because it guarantees an adequate signal quality in all circumstances.

To demonstrate the advantages of the the optimally warped transform, we return to the ECG of fig. 11, which was difficult to code, even with the DLT. The OWT approximation of order 30 of this ECG is shown in fig. 13. From an RMS point of view the OWT30 approximation is of the same quality as the DLT50 approximation in fig. 12 (see [3, p. 138]), even though its approximation order is much lower (30 vs. 50). However, the actual visual quality (as judged by a cardiologist) of the OWT30 approximation is much better. Indeed, the OWT30 reconstructs the R-S waves somewhat more accurately than the DLT50; more importantly, the OWT30 approximation is much less noisy than the DLT50 approximation. Of course, this last result is due to the lower approximation order, in accordance with the theory of section 2.

The OWT30 approximation is also the winner from a coding point of view: it requires a rate of only 250 bit/s (compression ratio 20) for coding the coefficients as compared to 360 bit/s (compression ratio 14) for the DLT [3, p. 138]. Of course, some extra bit rate is needed to code the frequency functions, which are polynomials of degree 19. However, the parameter bit rate can be kept very low by exploiting the fact that the parameters change little from interval to interval [4]. Therefore, the total bit rate is not much higher than the coefficient bit rate when sufficiently long ECGs are coded (e.g., 8 seconds or more).

Finally, fig. 13 presents evidence that it is not even necessary to use warping functions with relatively many (e.g., 10 to 20) parameters. The LSWT30 approximation in this figure has a piecewise constant frequency function which is completely determined by only four parameters. Despite the much lower number of parameters, the LSWT30 approximation is of the same quality as the OWT30 approximation and it requires about the same coefficient bit rate [3, p. 138]. For ECG coding, this result is not so important, because the parameter bit rate is low anyway (see the above discussion). However, in other applications, a low parameter bit rate may be more critical.

3.2.3 Remarks

- In [35], the OWT was compared to a wavelet-based compression technique, which represents signals by the location and amplitude of the extrema of their wavelet transforms. The (somewhat preliminary) results suggest that the time-warped method allows a better signal quality at low bit rates and is robuster in the presence of noise. However, more research is needed to validate these conclusions.

- The adaptive coding method is not restricted to ECG coding: the results in [10] indicate that it may also be used for coding blood pressure signals and electromyograms.

3.3 Adaptive filtering with time-warped polynomials

An important property of approximation by low-degree time-warped polynomials is that white noise in signals is removed, or its level at least significantly reduced. In fact, a time-warped signal approximation can be interpreted as the output signal of a time-varying filter with local cut-off frequency $\nu_N(t)$ (see section 2.3).

When the warping function of the time-warped polynomial is optimally adapted to the signal, $\nu_N(t)$ equals the local bandwidth $B(t)$ of the noiseless signal and all signal components are reconstructed perfectly. Furthermore, noise components with local frequency above $\nu_N(t)$ are completely removed.

In fact, the signal-to-noise ratio of an optimally adapted warped polynomial approximation is inversely proportional to N^* (see section 2.3), i.e., to the mean local bandwidth $\overline{B}$ of the (noiseless) input signal (see eq. (10)). In contrast, the cut-off frequency of a non-adaptive ideal low-pass filter must be higher than the global signal bandwidth B, if the signal is not to be distorted. As a result, the output signal-to-noise ratio of such a filter is inversely proportional to the *maximum local bandwidth* of the input signal.

More precisely, it has bee shown [10] that the signal-to-noise ratio of the optimal warped polynomial approximation is 10 log $\left(B/\overline{B}\right)$ dB better than that of the best non-adaptively low-pass filtered signal. As the maximum local bandwidth of a signal is often much larger (and never smaller) than the mean local bandwidth, the adaptive filter therefore usually performs much better as is confirmed by the experimental results below (see [10] for more details).

3.3.1 Noise removal from biomedical signals

Sections 3.1 and 3.2.2 already demonstrated that warped polynomial (both adaptive and non-adaptive) filtering efficiently and accurately removes noise from ECG signals. The results in [10] show that other biomedical signals, e.g., pneumotachs (which are actually electromyograms of the lungs) and blood pressure signals, can be filtered just as efficiently and accurately.

In this section, we will restrict ourselves to a single example, which is shown in fig. 14. The example shows a noisy pneumotach signal, its "optimal" warped approximation and two low-pass filtered versions of the signal. Actually, in order to avoid computing N^* for each signal interval, the adaptive warped polynomial filter uses a fixed value of N in each interval (see section 2.4.2), the optimal value of which is determined experimentally. Consequently, the adaptive filter is actually only sub-optimal.

The results in fig. 14 clearly show that it is not possible to adequately filter the pneumotach with non-adaptive filters. Indeed, when the cut-off frequency of such a filter is low (curve C) noise is adequately removed, but the signal is heavily distorted; of course the signal distortion can be reduced by increasing the cut-off frequency, but only at the expense of a lower output SNR. In fact, the optimal compromise (curve D) is far from perfect: the output signal is still noisy and some signal distortion remains. In contrast, the adaptively filtered signal (curve B) does not suffer from these problems, because its cut-off frequency (curve E) adapts to the local signal bandwidth. In fact the filter reacts *immediately* to changes in the signal's morphology, and this has proved a major advantage over adaptive least mean squares and recursive least squares filters, especially in biomedical applications.

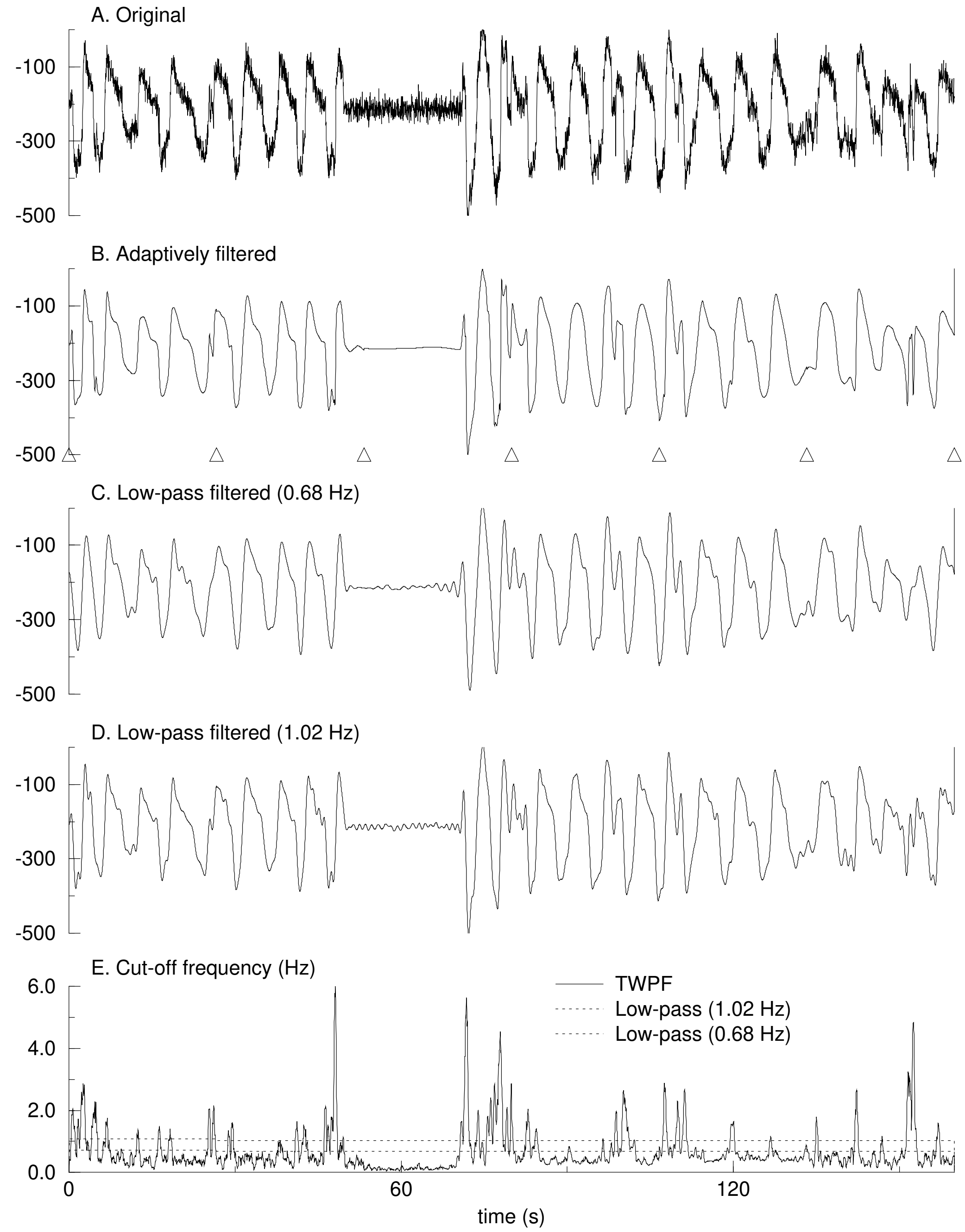

Figure 14: Filtering a noisy pneumotach signal (sampled at 30Hz) using the warped polynomial adaptive filter and two different low-pass filters. The cut-off frequencies of the interval-adaptive and the non-adaptive filters are shown at the bottom.

Because of its superiority over other adaptive filters, the adaptive warped polynomial filter method has been applied in a Spanish commercial ECG transmission system. This system transmits analogue ECGs over the public telephone system using pulse width

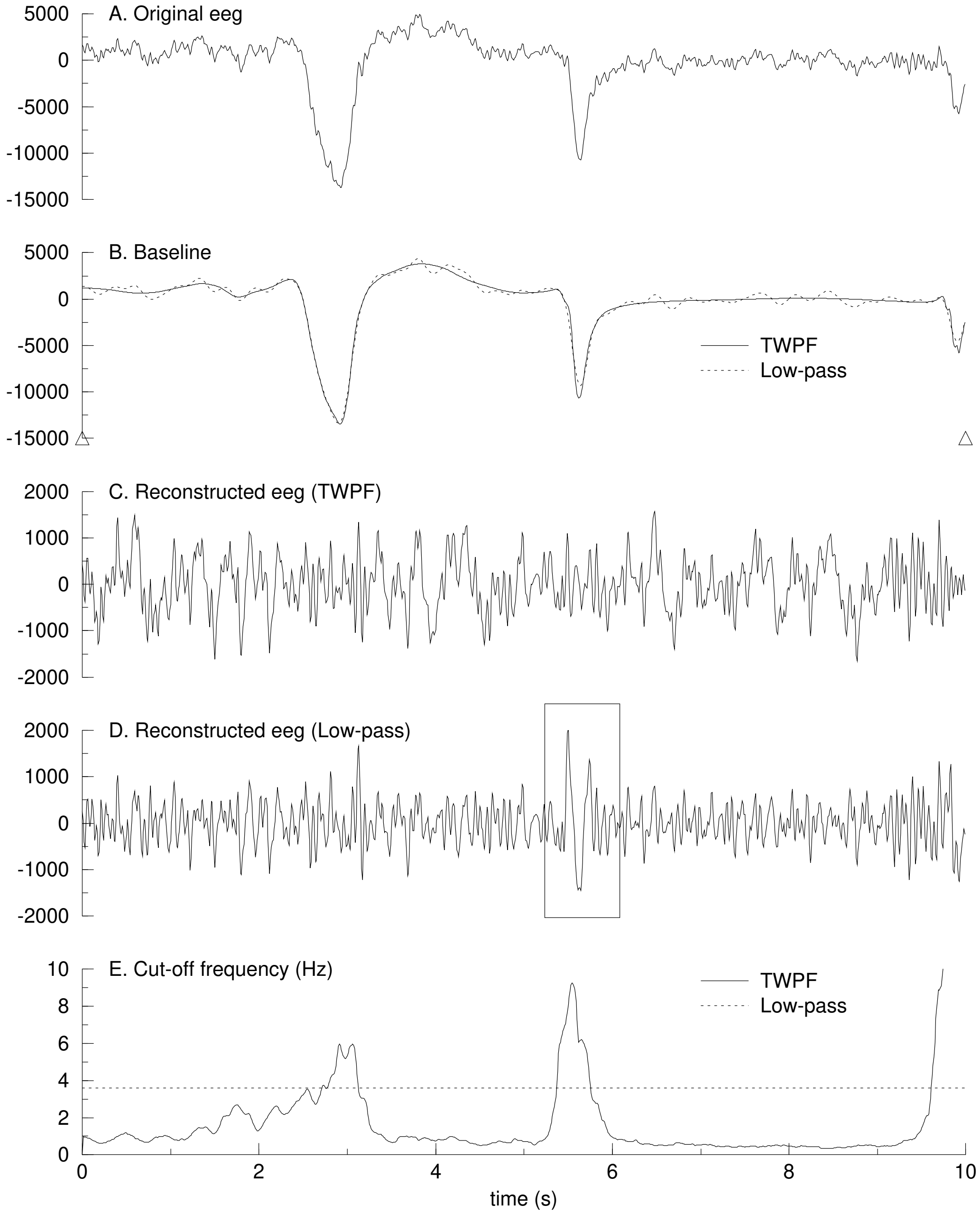

Figure 15: Base line removal from an EEG signal (sampled at 100Hz): A. the original signal; B. the estimated base lines using interval-adaptive and non-adaptive filtering; C. the EEG after interval-adaptive base line removal; D. the EEG after non-adaptive base line removal; E. the cut-off frequencies of the base line estimators.

modulation. At the receiver's end, the adaptive filter will remove both the recorded noise and the noise introduced by the communication channel. A digital ECG transmission system is planned and will also employ the filter method, this time to compress the signals and to remove the recorded noise.

3.3.2 Baseline correction of electro-encephalograms

To demonstrate that the warped polynomial adaptive filter performs well, even when the input noise is non-stationary, we now present a second application of the filter: the base line correction of electro-encephalogram (EEG) signals. Fig. 15 shows an original 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. 15, which shows the base line estimated using a low-pass FIR filter (curve B, dotted line) and the reconstructed EEG (curve D) 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, this optimal compromise completely 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. 15 contains such an artifact.

Better results can be obtained with an adaptive warped polynomial filter: the baseline estimated by the adaptive filter (curve B, solid line) is much flatter overall, except there where the base line fluctuates rapidly. As a result, most of the δ-waves are retained in the reconstructed EEG (curve C). This is confirmed by the fact that the local cut-off frequency of the adaptive filter (curve E) is typically about 1 Hz. On the other hand, this cut-off frequency is much higher than that of the FIR-filter in those areas where base line fluctuations occur. As a result, no artifacts are introduced in that area.

3.4 Adaptive transform coding of images

The theory of warped signal approximation can be extended to images, i.e., to signals of two variables. In this case, the image intensity function $f(x, y)$ is approximated by a bi-variate warped polynomial

$$f_I(x, y) \stackrel{\triangle}{=} \sum_{(i,j)\in I} C_{i,j} X^i(x, y) Y^j(x, y), \tag{13}$$

where I is a suitably chosen index set.[14]

Just as in the one-dimensional case, it is possible to express the approximation eq. (13) as a linear combination of warped orthonormal polynomials. Furthermore, these orthonormal base functions satisfy recurrences, but these recurrence are usually many times more complex than the three-term recurrence which holds in the one-dimensional case. A noteworthy exception are the weakly separable orthonormal warped polynomials [20] which satisfy three-term recurrences.

[14] The functions $X(x, y)$ and $Y(x, y)$ are valid warping functions when the transformation $(x, y) \mapsto (X, Y) : X = X(x, y), Y = Y(x, y)$ is invertible on the support of $f(x, y)$.

Unfortunately, generalizing the theory and the algorithms of section 2 to two dimensions is far from easy. First of all, no analytic approximation of the orthonormal warped polynomials has been found yet for the two-dimensional case, except for a certain type of weakly separable orthonormal warped polynomials [20]. This complicates extending the local bandwidth theory to two dimensions. Secondly, computing the orthonormal bases requires many more computations than in the one-dimensional case. For this reason, the author has not yet investigated the general two-dimensional case.

Fortunately it is possible to avoid two-dimensional approximation altogether by applying a number of one-dimensional warped transforms, first to the image rows and then to the image columns. This is explained below.

3.4.1 A practical image coding method

In traditional two-dimensional coding, the computational complexity is limited by using separable orthogonal transforms instead of general two-dimensional transforms. This means that a one-dimensional transform such as the DCT is successively applied to each of the image rows and then to each of the image columns. In practice, this approach works quite well and it has been shown that the separable 2D-DCT is close to optimal for coding stationary image blocks.

In principle, a similar approach may be used in extending adaptive warped transform coding to two dimensions. This would involve computing the optimal transform for each image row, and the corresponding coefficients A_n. Next the coefficients A_n for the rows would be arranged in an *intermediate coefficient image* and the columns of this image would be transformed like the rows of the original image. The resulting *final coefficients* would then be quantized and coded, together with the frequency functions for all rows of the original image and for all columns of the intermediate coefficient image.

Unfortunately, the author's experiments have shown that this approach does not work well. The main reason for its failure seems to be that coefficients of the same order on different rows are not strongly correlated. This implies that the column transforms are not very efficient, because there is little redundancy to exploit. Furthermore, there is a significant degree of correlation between coefficients of different order on neighboring rows. However, this correlation (and the redundancy it implies) cannot be exploited by column transforms.

In contrast, in traditional separable coding, coefficients of the same order on neighboring rows are highly correlated, while those of different order are (almost) completely uncorrelated, at least for stationary image statistics. This means that column transforms are able to exploit all of the remaining redundancy. Of course, this is only true for stationary image statistics.

In the author's opinion, the undesirable correlation properties in adaptive warped separable coding are due to the fact that different rows are transformed using different transforms, while in traditional transform coding the same transform is used on all rows. However, a convincing theoretical explanation of this phenomenon is still lacking.

Fortunately, it is possible to transform the coefficients of each row into a new representation, with much better inter-row correlation properties, using the principle of sub-Nyquist uniform sampling (see section 2.6.2). Indeed, suppose that all image rows can be well approximated using optimally adapted warped polynomials of degree N_r. As explained in section 2.6.2, it is possible to exactly reconstruct these warped polynomials from

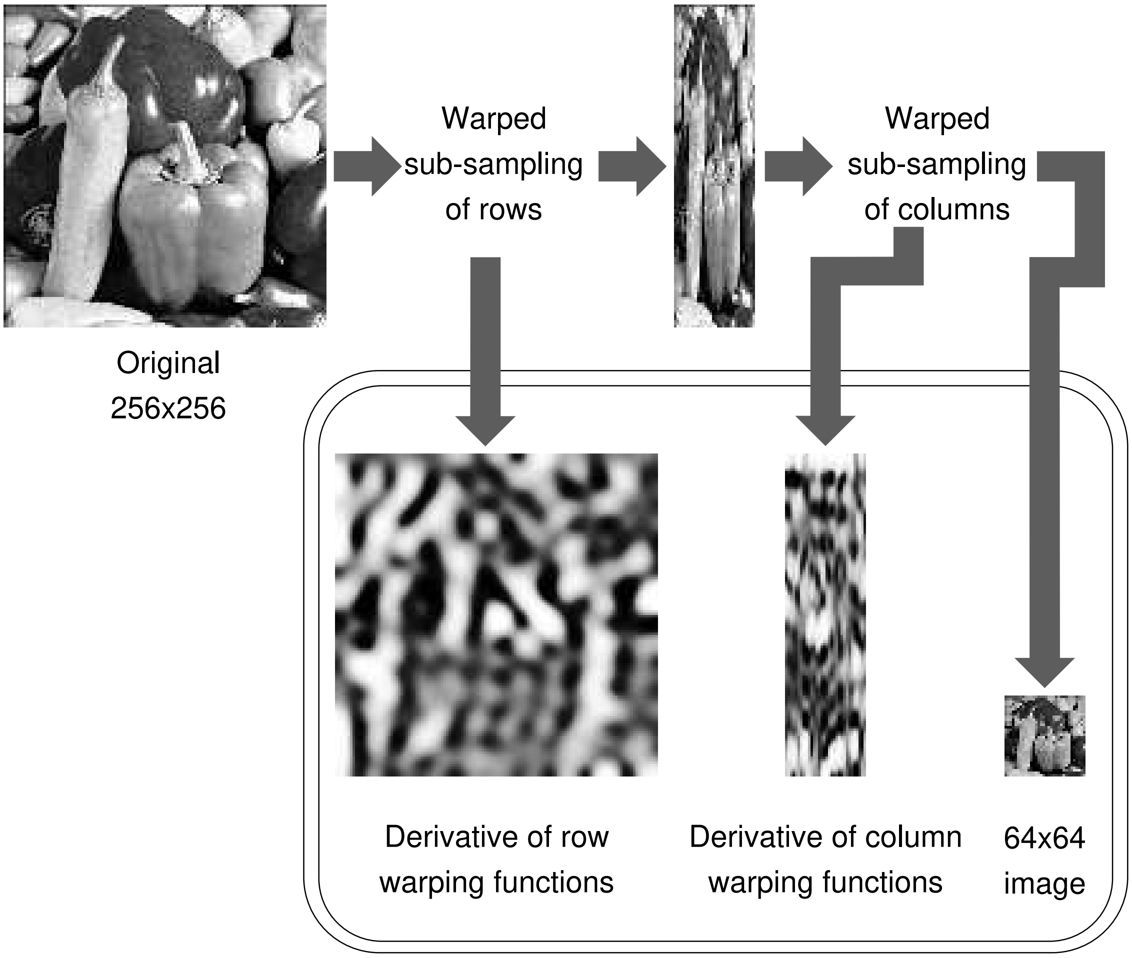

Figure 16: The principle of warped image coding using adaptive sub-Nyquist subsampling.

$N_r + 1$ uniformly taken samples. When these samples are arranged in an *intermediate image* of size $R \times (N_r + 1)$ where R is the number of rows in the original image, then this image will be (approximately) a horizontally subsampled version of the original image. Therefore, the pixels within one column of the intermediate image will be highly correlated (as in the original image), while the correlation between pixels in different columns will be low (because those pixels are far apart in the original image). Consequently, the intermediate image can be efficiently compressed using a warped column transforms.

Fig. 16 illustrates how this warped image coding (WIC) method is used to compress an image. First, the optimal frequency functions (and the corresponding warping functions) for the different rows are computed. Of course, just like in one-dimensional coding these warping functions are restricted to be polynomials to reduce the overhead involved in coding their parameters. In fact, to reduce this overhead even further, these parameters are forced to vary polynomially over the different image lines. In the second step, the warped polynomial approximations of the image rows are subsampled horizontally. This results in an intermediate image, the columns of which are subsampled in a similar manner.

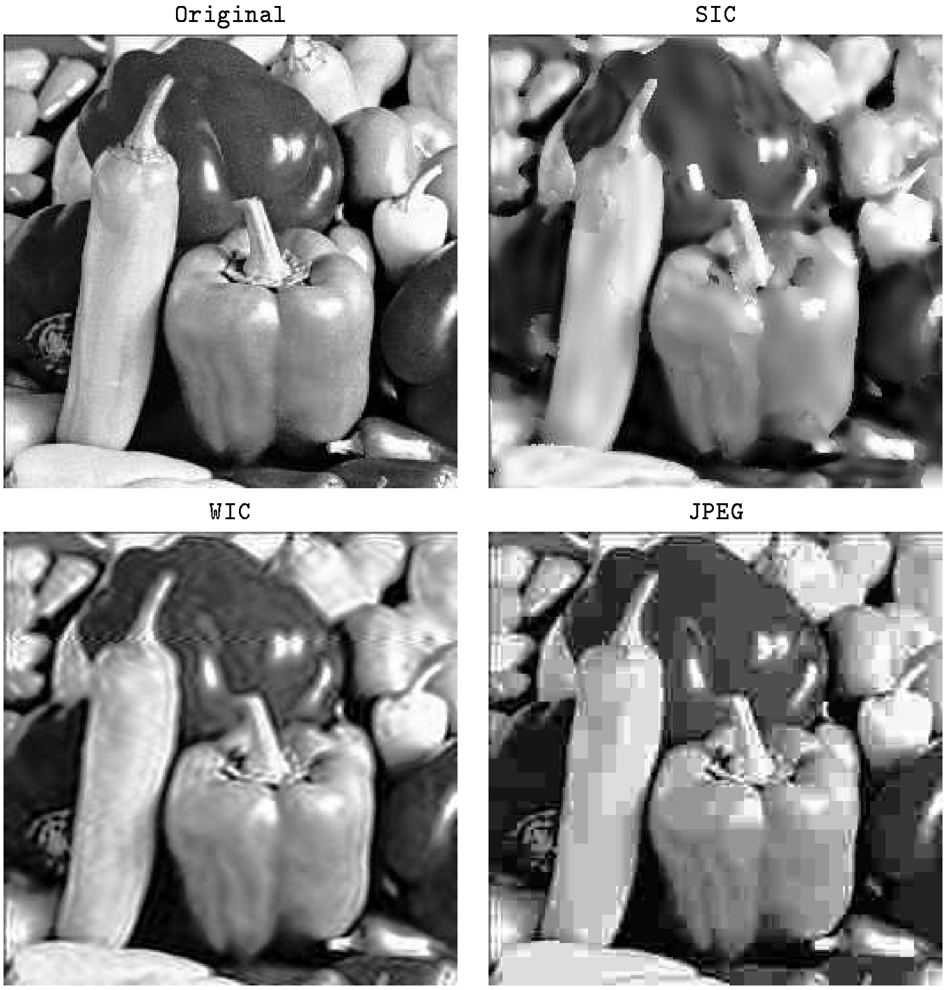

Figure 17: The original 256x256 "peppers" image and the images compressed to 0.3 bpp by SIC (RMSE= 8.95), WIC (RMSE= 6.47) and JPEG (RMSE= 7.09).

Finally, this results in two sets of frequency function parameters and one subsampled image of size $N_r \times N_c$, which may be coded using any traditional technique.

To reconstruct the image from the coded data, the coding procedure is simply applied in reverse. This involves reconstructing the warped polynomial approximations of the columns of the intermediate image and of the rows of the original image from their sub-Nyquist samples. As we have seen in section 2.6.2, practical algorithms for doing this are numerically instable. However, this problem can be circumvented by using a slightly more complicated subsampling strategy in the coder. Indeed, it is possible to devise an orthogonal transform which converts $N+1$ coefficients of a warped polynomial approximation $s_N(t)$ into a set of $N+1$ sample values which closely resemble (but are not equal to) uniformly taken samples of $s_N(t)$. As this "subsampling" transform is orthogonal, it can be inverted without any numerical problems whatsoever [6, 7].

3.4.2 Experimental results

The example in fig. 17 illustrates the potential of warped image coding (see also [22]). It compares the WIC method to three other methods at a rate of 0.3 bit per pixel, which is rather low for this type of image. In this particular example the final subsampled image of the WIC method is of size 64×64 and it is coded by a full-image DCT. The warping functions for the line and column transformations can are represented by 800 parameters in total. As usual, the coefficients of the sub-sampled image and the warping function parameters are quantized and entropy coded.

The other methods being compared in fig. 17 are the JPEG image compression standard, the full-image DCT and, finally, segmented image coding (SIC) which is a very promising low bit rate coding technique that represents images using a contour-texture model. Of course, in all methods the bit rate of 0.3 bit per pixel takes into account all necessary over-head information (e.g., the warping function parameters in WIC).

According to fig. 17, the WIC method produces the best images at 0.3 bit per pixel (bpp), both in terms of RMS Error (RMSE) and from a subjective image quality point of view. In particular, the WIC image is sharper and less distorted by Gibbs effect than the full-image DCT. Furthermore, unlike JPEG, it does not suffer from annoying block effects. Finally, the WIC image looks less artificial than the SIC image.

Some care should be taken in interpreting these results: the bit rate and image quality of SIC depend heavily on the segmentation accuracy and on the number of contour pixels. With the development of better segmentation methods and contour coders, the SIC results will probably improve. Nevertheless, fig. 17 indicates that WIC has a clear potential to outperform SIC in low bit rate coding. On the down side, it should be mentioned that the current WIC implementation requires 10 times more computations than SIC (but the complexity of WIC can still be reduced in several ways).

3.4.3 Remarks

To the author's knowledge, only few adaptive image compression techniques (i.e., techniques using image dependent bases) have ever been proposed and even fewer have been successful. This is due to the difficulties in coding bases efficiently and due to the lack of a suitable criterion for adapting the base. Both of these problems are solved by using warped polynomials, which can be represented efficiently and for which an adaptation rule is available.

Also, the problem of adaptive coding has often been approached from a stochastical or information theoretic point of view. This is appropriate when the same base is used several times, e.g., to code a relatively large number of image blocks. The theory of warped polynomials, on the contrary, treats adaptive coding from a bandwidth point of view. This is appropriate even when a base is only used once, to code a single signal or image. In this context, warped adaptive coding is far more flexible, because it is optimal for *each image*, instead of *on average*.

4 Periodical time-warped transforms

Even though the class of real-valued warped polynomial transforms is quite general and includes many well-known transforms such as the DCT, the DLT and the DST, it does not include the discrete Fourier transform (DFT), which has complex-valued base functions. The DFT is a very important transform in signal processing; it is employed in many applications, e.g., spectral analysis, filter design and coding. From a stochastical point of view, the DFT is the optimal transform[15] for coding segments of duration T of periodical signals with period T generated by a stationary stochastic process.

Just like the DCT in the case of locally correlated signals, the DFT is usually not the optimal transform for approximating a particular segment of a particular periodical signal, even though it usually is optimal in the mean, when coding a large number of signal segments (see section 1.2). Therefore, it is worthwhile to investigate transforms which may be adapted to a given (periodical) signal. In other words, it would be nice to extend the theory of the previous sections to periodical signals. Fortunately, this is possible by introducing periodical warped trigonometric polynomial (PWTP) signal approximations, which may be viewed as generalizations of DFT approximations.

In the remainder of this section, we first summarize the theory of PWTP approximation. We will be very brief here as this theory is very similar to the theory of "ordinary" warped polynomial approximation. Next, we will discuss the first practical application of PWTP approximation, i.e., contour coding.

4.1 Approximation by periodical warped trigonometric polynomials

In the following, we will deal with periodical approximations of periodical signals. Without loss of generality, we will assume that the period of these functions is 2π.

Any periodical signal $s(\theta)$ may be approximated by a (discrete) Fourier approximation $s_N(\theta)$ of order N, which is an expansion in terms of the orthonormal complex exponentials $\exp(jn\theta)$:

$$s_N(\theta) = \sum_{n=-N}^{N} A_n \exp(jn\theta), \tag{14}$$

where A_n is the inner product of $s(\theta)$ and the base function $\exp(jn\theta)$. In practical computations, only uniformly taken sample values $s(\theta_k)$, $0 \le k \le L$ of the signal are available and the coefficients A_n are computed using the discrete inner product

$$f \cdot g = \frac{1}{L} \sum_{k=0}^{L} f(\theta_k)\overline{g(\theta_k)}, \text{ where } \theta_k = k2\pi/L. \tag{15}$$

In theoretical derivations, it is sometimes more convenient to assume that $s(\theta)$ is known for all $\theta \in [0, 2\pi]$; then the coefficients are computed using the continuous inner product

$$\langle f\,,\,g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(\theta)\overline{g(\theta)}\,dt. \tag{16}$$

A linear combination of complex exponentials $\exp(jn\theta)$, i.e., a function like $s_N(t)$ but with arbitrary coefficients A_n, is called a *trigonometric polynomial* of degree N. The concept of Fourier approximation may be generalized by introducing a warping function $\gamma(\theta)$ (which in this context is defined as a strictly increasing function that maps the

[15] In section 1.2, we have seen that the DCT is close to optimal for coding stationary signals with *local correlation*. However, the local correlation assumption is not satisfied for periodical signals; therefore the DCT is not optimal for such signals.

interval $[0, 2\pi]$ onto itself) into the approximation $s_N(\theta)$, i.e., by turning $s_N(\theta)$ into a warped trigonometric polynomial of degree N:

$$s_N(\theta) = \sum_{n=-N}^{N} C_n \exp(jn\gamma(\theta)). \tag{17}$$

Unfortunately, the functions $\exp(jn\gamma(\theta))$ are not orthonormal in general. This means that the optimal coefficients C_n, i.e., the ones that minimize the mean squared error between $s_N(\theta)$ and $s(\theta)$, can no longer be computed using a simple inner product evaluation. This problem is easily solved by introducing orthonormal base functions $\phi_n(\theta)$, $n = -N, \ldots, 0, \ldots N$ which span the same space $\mathcal{S}_N$ as the non-orthogonal functions $\exp(jn\gamma(\theta))$, $n = -N, \ldots, 0, \ldots N$. In terms of these base functions, the LMS approximation $s_N(\theta)$ is given by

$$s_N(\theta) = \sum_{n=-N}^{N} A_n \phi_n(\theta), \tag{18}$$

in which the optimal coefficients are now given by $A_n = s \cdot \phi_n$ (or $A_n = \langle s\,,\, \phi_n \rangle$ in the continuous case).

There exist many possible sets of orthonormal PWTP base functions $\phi_n(\theta)$ which span the space $\mathcal{S}_N$. However, in practice the base functions $\phi_n(\theta)$, $n = -N, \ldots, 0, \ldots N$ are always chosen such that any subset $\phi_n(\theta)$, $n = -j, \ldots, 0, \ldots j$ spans the subspace $\mathcal{S}_j$ of $\mathcal{S}_N$. When this *inclusion property* is satisfied, approximations $s_N(\theta)$ of different degree N may be obtained simply by removing or adding terms in eq. (18); furthermore, for such a choice, the orthonormal PTWPs can be computed quickly using a three-term recurrence (see [3, p. 82]).

Note that there exist several sets of orthonormal PWTPs which satisfy the inclusion property. In particular, there exist complex valued base functions, but also real valued ones.[16] Just as in the case of ordinary warped polynomials, simple analytic approximations of these base functions have been derived [3, p. 82], even in the more general case of a weighted inner product. Furthermore, the approximations are amplitude- and frequency-modulated complex exponentials, or cosines and sines or combinations thereof and the local base function frequency $\nu_n(\theta)$ is a simple function of the warping function:

$$\nu_N(\theta) = N \frac{d\gamma}{d\theta}. \tag{19}$$

This almost complete analogy between ordinary orthonormal warped polynomials and PWTPs is no coincidence because the former can be considered special cases of the latter [3]. In fact "ordinary" orthonormal warped polynomials are PWTPs with a symmetric frequency function $d\gamma/d\theta$ and a symmetric weight function. The analogy implies that most of the theory concerning "ordinary" warped signal approximations directly applies to PWTP approximation (with some minor modifications, due to, e.g., the use of an approximation interval $[0, 2\pi]$ instead of $[-1, 1]$). In particular, the local bandwidth theory and the iterative algorithm for computing the optimal warping function (or frequency function) apply to PWTP approximations with almost no changes. For this reason, we will not repeat the theory here.

[16] This is not surprising because Fourier approximations can be expressed in terms of complex exponentials, as in eq. (14), but also in terms of $\cos(jn\theta)$ and $\sin(jn\theta)$, which are also mutually orthogonal functions.

4.2 Contour coding with periodic warped trigonometric polynomials

In several applications, the ability to represent a contour (e.g., the boundary of an object) compactly is essential. For instance, in pattern recognition such a description may be used for shape recognition or for geometric alignment of two objects; there, a compact description reduces both the computation time and the storage requirements (which is important, e.g., when a contour needs to be compared to the contours in a large data base). Efficient contour coding is also important in segmented image coding [20], which is a second-generation technique that compresses image contours and texture separately.

4.2.1 The theory of contour representation

Any contour can be represented by a parameter representation $(x(t), y(t))$, which associates a value of the real parameter t to each contour point (x, y). Therefore, a contour may be compactly represented by approximating the functions $x(t)$ and $y(t)$ by a warped polynomial or a PWTP and then using the coefficients of this approximation as contour descriptors. In the case of a closed contour, $x(t)$ and $y(t)$ are periodical functions, which are best represented by a warped *trigonometric* polynomial. On the other hand, open contours should be approximated using *ordinary* warped polynomials, because their $x(t)$ and $y(t)$ are not periodical. In this section, we only deal with closed contours and without loss of generality, we will assume that the period of $s(t)$ is 2π. However, most of the results equally apply to open contours.

In principle, $x(t)$ and $y(t)$ may be represented independently, i.e., using warped trigonometric polynomials with different warping functions. Indeed, simultaneous optimal coding of $x(t)$ and $y(t)$ can in general only be achieved using different warping functions. However, in practice it may be more beneficial to use the same warping function for $x(t)$ and $y(t)$, at the expense of a higher approximation order N for $x(t)$, $y(t)$, or both. This is because the overhead involved in coding the warping functions can be completely eliminated when the warping functions for $x(t)$ and $y(t)$ are the same. To explain this, we first note that the parameter representation of a contour is always redundant: indeed, all parameter representations $\left(x\left(\gamma(t)\right), y\left(\gamma(t)\right)\right)$, with $\gamma(t)$ an arbitrary warping function, trace exactly the same contour, albeit at different speeds. Now suppose that $x_N(t)$ and $y_N(t)$ are PWTP approximations of $x(t)$ and $y(t)$ with the same warping function $\gamma(t)$. Furthermore, suppose that $x_N(t)$ and $y_N(t)$ are expanded into a linear combination of (non-orthogonal) warped complex exponentials $\exp\left(jn\gamma(t)\right)$, as in eq. (17). Clearly, these expansions are of the form $x_N(t) = x'_N\left(\gamma(t)\right)$ and $y_N(t) = y'_N\left(\gamma(t)\right)$, where

$$x'_N(t) \triangleq \sum_{n=-N}^{N} C_{n,x} \exp\left(jnt\right), \tag{20}$$

and

$$y'_N(t) \triangleq \sum_{n=-N}^{N} C_{n,y} \exp\left(jnt\right). \tag{21}$$

Now as $\left(x'_N(t), y'_N(t)\right)$ traces the same contour as $\left(x_N(t), y_N(t)\right)$, this contour can be reconstructed from the coefficients $C_{n,x}$ and $C_{n,y}$, without knowing the warping function $\gamma(t)$.

Using the same warping function for both components is extremely attractive. Indeed, this not only eliminates the parameter bit rate, but it also allows the use of warping functions, which are more flexible than polynomials of low degree. Of course, on the downside, using the same warping function leads to an increased coefficient bit rate. At the moment, it is not yet known which of these two effects (reduction in parameter bit rate or increase in coefficient bit rate) is dominant in practice.

In the remainder of this document, we will only consider the use of a single warping function. In this case, $x(t)$ and $y(t)$ are approximated using the same orthonormal base functions and the notations may be simplified by introducing the complex-valued parameter representation $s(t) \stackrel{\triangle}{=} x(t) + jy(t)$. The problem is then reduced to computing the optimal PWTP approximation $s_N(t) \stackrel{\triangle}{=} x_N(t) + jy_N(t)$ of the complex-valued signal $s(t)$, and expressing it in the form of eq. (17) to determine the coefficients C_n. In the particular case where $\gamma(t) \equiv t$, the coefficients C_n are called the Fourier descriptors (FDs) of the contour. Therefore it seems appropriate to call them improved Fourier descriptors (IFDs), in general.

IFDs are alway more efficient contour descriptors than FDs, in the sense that fewer IFDs are required to approximate a contour with a given accuracy. More importantly, they are invariant to warping of the parameter representation $s(t)$, i.e., the IFDs of $s(t)$ and those of $s(\gamma(t))$ differ only by a simple phase factor at the most [24,25]. This is a very desirable property in pattern recognition, because it facilitates the comparison of contours traced at different speeds.

4.2.2 Example

Fig. 18 shows two original contours, extracted from segmented images, and their approximations using 15 IFDs and 15 FDs, respectively. Note that the number of descriptors (30 real numbers) is low compared to the number of contour points. In spite of this, the IFD approximations faithfully reconstruct the major features of the original contours. In any case, the FD approximations are much less accurate than their IFD counterparts. In fact, they lack some of the important features by which a human observer differentiates one contour from another. For instance, some of the "U-bends" in the original contours are deformed into "V-bends" (e.g., in the contour on the left) or are missing (e.g., in the contour on the right). Also, "rough" contour parts are smoothed more in the FD approximations. In fact, all of these errors are typical for FD approximations.

The above example, and other examples in [24], clearly demonstrates the potential of PWTP contour approximation. However, much research remains to be done. For example, one problem to be resolved is the following: when contours are extracted from images, coordinates of neighboring contour points always differ by 1, -1, or 0. In the IFD and FD methods, this information is not used and it remains to be seen whether it can be exploited to achieve more efficient contour coding.

5 Topics of further research

The preceding sections have described many theoretical and experimental results concerning warped polynomials. However, the story of warped polynomials and their applications doesn't end here. Indeed, the theory of warped polynomial approximation can be extended

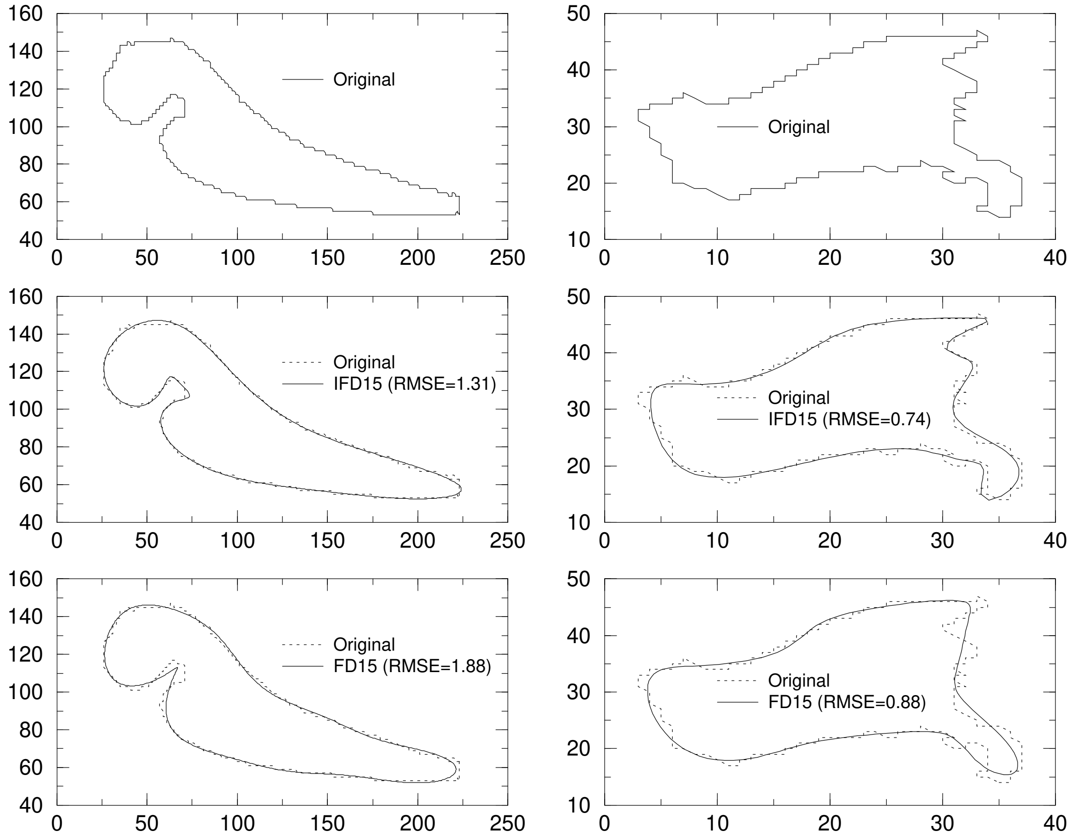

Figure 18: Approximation of two contours using 15 ordinary Fourier descriptors (FD15) and the same number of improved Fourier descriptors (IFD15). The contours were extracted from segmented versions of an mr-image (left) and of the "peppers" image (right) and consist of 587 and 152 points, respectively.

in several ways. Furthermore, warped polynomials may be used in many new applications. In this section, we discuss a few of these possible extensions and new applications.

5.1 Possible extensions of the theory and the algorithms

- **Extension to band-pass signals.** The signal approximation theory of section 2 and the methods of section 3 are best suited for representing signals which are locally low-pass. However, both the theory and the algorithms can be extended to band-pass signals. Bandpass time-warped transforms could represent chirp signals (often used in radar), other frequency-modulated signals and certain speech signals. Bandpass signals are represented by their local bandwidth and their local central frequency. Using standard modulation theory, it is not difficult to produce variants of the time-warped base functions which represent such signals with the smallest possible number of coefficients.

- **Extension to warped wavelet bases.** In the continuous case (i.e., for non-sampled signals), it is just as easy to define orthonormal warped wavelets as it is to compute orthonormal warped polynomials (e.g., see [37]). However, no simple method for computing discretely orthonormal wavelet bases and for adapting them to a particular signal has yet been published (although there is some literature on the closely related topic of signal-adaptive subband coding, e.g., [38]). Nevertheless, such bases might be useful: they would combine the advantages of warped polynomials (very flexible adaptation) and wavelets (localization in both time and frequency).

- **Extension to the multi-variate case.** As mentioned in section 3.4, the theory of warped polynomials can be extended to the multi-variate case, i.e., the approximation of functions of multiple variables by warped polynomials in the same variables. In particular, it is possible to generate bi-variate warped polynomials which are orthonormal on a general two-dimensional region (e.g., see [20] for only one example). Such base functions may be used to approximate image intensities on rectangular or non-rectangular regions. This paves the way to adaptive image coding techniques, which are more sophisticated but also more complicated than the method described in section 3.4. However, some important problems remain to be solved. For instance, the local bandwidth theory of warped polynomial approximation needs to be extended to the multi-variate case.

- **Extension to multiple warping functions.** The uni-variate orthonormal warped polynomials depend on a single warping function and a single weight function. However, it is possible to create more complicated orthonormal base functions which depend on two or more warping functions. This can be achieved by applying the techniques for generating bi-variate base functions which are orthonormal on a general two-dimensional region, to regions which are essentially one-dimensional, e.g., a number of points located on a parabola or a hyperbola.

- **Extension to multi-resolution coding.** The image coding method described in section 3.4 clearly offers the potential for multi-resolution coding. In principle, the subsampled image can itself be coded by the adaptive method. In this way, a pyramid of images would be obtained. These images could be complemented by error images if needed. This scheme is somewhat similar to traditional sub-band coding, but with one distinction: it is image-adaptive.

5.2 Possible new applications

5.2.1 Reducing the acquisition time in computed imaging

Accurately representing signals by a small number of coefficients is not only important in coding and filtering. It is also important in applications where signals or images need to be reconstructed from a small number of measurements. On important application of this is reducing the acquisition time in magnetic resonance (MR) imaging.

In MR imaging, the measured data are samples of the spatial Fourier transform $f(u, v)$ of the desired image. Actually, each measurement produces a whole "line" of the Fourier spectrum, e.g., all samples $f(u, v)$ corresponding to a given horizontal spatial frequency v.

Unfortunately, acquiring a complete image takes quite some time, which is undesirable for several reasons. For instance, any movement of the patient during the measurements (e.g., due to breathing or heart beats) causes motion artifacts in the images. Also, the long measurement times (typically 30 minutes to 1 hour for a set of 128 to 256 images) limit the number of patients that can be examined on a given day. Therefore, reducing the measurement time would reduce costs and would improve the quality of the images (because fewer motion artifacts would occur). Unfortunately, the time needed to measure a single line cannot be reduced because of fundamental physical reasons. Therefore, the only way to reduce the total acquisition time is to measure fewer lines.

Now suppose that for each value of u, $f(u, v)$ is represented as a linear combination of base functions with a given number $N + 1$ of unknown coefficients. Then in principle, $N + 1$ well-chosen samples of $f(u, v)$ for each value of u suffice to determine all of the unknown coefficients and hence the complete image. Furthermore, if only these samples are acquired, the total acquisition would be reduced by a factor $(L + 1)/(N + 1)$, where $L + 1$ is the number of desired lines.

Clearly, in the above method, the acquisition time is directly proportional to the minimally achievable approximation order N. The previous sections have demonstrated that this approximation order can be many times lower than L, when $f(u, v)$ (or, alternatively, the original image) is represented in terms of image-adapted orthonormal warped polynomials. This suggests the following procedure for MR imaging with a reduced acquisition time:

1. Record a low resolution training image (which can be done quickly).
2. Adapt the frequency functions of the warped polynomials to the training image, and compute the corresponding optimal base functions.
3. Determine optimal values v_n, $0 \leq n \leq N$ for the frequencies v of the lines to be measured.
4. Measure the lines $f(u, v_n)$, $0 \leq n \leq N$.
5. Compute $f(u, v)$ for all desired u and v.
6. Reconstruct the image from $f(u, v)$.

This procedure reduces the total imaging time (steps 1 and 4) at the expense of an increased computation time (steps 2-5). However, in contrast to the acquisition time, the computation time can be kept small, e.g., by using parallel processing (note that a typical MR-scanner costs about 100 million Belgian francs, therefore computing cost is a relatively minor issue in MR imaging).

To the uninformed reader the chances of success of the proposed procedure may seem slim. As the proposed method has not yet been implemented (our research group hopes to do this during the following months) it is not easy to argue with this point of view. Nevertheless, there are indications that the procedure may actually work. For instance, Weaver and Healy [39] have published preliminary, but promising results using a similar procedure. However, their procedure has some important disadvantages over the one proposed above. Indeed, Weaver and Healy were unaware of the theory of warped polynomial approximation and therefore used a computationally very demanding statistical method to compute a set of statistically "optimal" base functions.[17] In contrast, orthonormal

[17] Interestingly, Weaver and Healy's base functions resemble orthonormal warped polynomials. Also, in some respects their method of computing the base functions is similar to the author's model-dependent adaptive transform [27].

warped polynomials can be computed and adapted to a given image many times faster. Furthermore, the theory of warped polynomial approximation may be used to derive some theoretical information concerning the optimal frequencies in step 3, which may be exploited to compute the optimal frequencies more quickly. In contrast, Weaver and Healy had no theoretical knowledge concerning their base base functions.

5.2.2 Other applications

Other possible applications of warped polynomial approximation include:

- **Adaptive interpolation of signals and images.** This may be useful to correct compressed image or signal blocks which have been lost due to transmission errors.

- **Time-frequency analysis.** The frequency function $\left|\frac{d\theta}{dt}\right|$ of the optimal transform provides information about the local signal bandwidth and could therefore be useful in signal analysis. Actually, other researchers, e.g., Baraniuk and Jones [40,41], have already introduced the concept of warping in time-frequency analysis. However, the current theory of warped signal analysis is restricted to non-sampled signals. Furthermore, the strategy for adapting the warping function to a particular signal is purely numerical, because a "local-bandwidth-like" theory is still lacking.

- **Aligning geometrically distorted images.** Actually, geometric distortion is nothing but a warping effect. Therefore, the alignment problem is equivalent to a "dewarping" problem. Dewarping may be facilitated by first approximating the images by warped polynomials.

6 Conclusion

Time-warped orthogonal transforms are a new and original tool for adaptive coding and filtering of signals, curves and images. The author's research has produced the following important results:

- It has introduced a new set of base functions, which are discretely orthonormal, which can be computed quickly, and which depend on only two functions, i.e., the warping function and the weight function.

- It has related the properties of the base functions to the warping and weight functions from which they are derived.

- It has introduced a simple theory which relates the minimal approximation order required to represent a given signal by a particular time-warped polynomial transform to the local signal bandwidth and to the frequency function of the transform. The theory also shows how to find the optimal transform for a given signal, i.e., the one which requires the lowest possible approximation order.

- It has introduced a quickly converging and simple algorithm for adapting a time-warped transform to a specific signal.

In addition, the author has extended this theory to warped trigonometric transforms and to image coding. Furthermore, the theory of warped polynomial approximation offers much room for extensions, e.g., to warped wavelet bases and to warped bases with multiple warping functions.

The theoretical results on warped polynomial approximation have already been successfully applied in adaptive signal, image and contour coding methods and in an adaptive filter scheme for biophysical signals. Experimental results shows that the coding and filter methods outperform traditional, non-adaptive methods. Finally, the developed theory promises improvements in such diverse applications as the coding of arbitrarily shaped image regions, computed imaging, time-frequency analysis, non-uniform sampling and interpolation.

References

[1] 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.

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

[3] W. Philips, *Vervormde Polynomen voor de Compressie van Eén- en Tweedimensionale Signalen.* PhD thesis, Department of Electronics and Information Systems, RUG, St.-Pietersnieuwstraat 41, 9000 Gent, Belgium, 1992–1993. In Dutch.

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

[5] W. Philips, "On the data compression properties of orthogonal polynomial expansions," Tech. Rep. DG 91-05, ELIS, RUG, Gent, Belgium, 1991.

[6] W. Philips, "Adaptive image coding using warped polynomial transforms," in *Signal Processing VII: Theories and Applications*, vol. 1, (Amsterdam), pp. 588–591, Elsevier, Sept. 1994. Proceedings of EUSIPCO-94, Edinburgh, Scotland.

[7] W. Philips, "Adaptively subsampled image coding with warped polynomial transforms," in *Proceedings of ICIP94*, vol. 2, pp. 366–370, Nov. 1994. Austin, Texas.

[8] W. Philips, "Adaptive warped transform coding of signals and images," in *Proceedings of the PRORISC IEEE Benelux Workshop on Circuits, Systems and Signal Processing*, pp. 185–191, Mar. 1994. Papendal, Arnhem, the Netherlands. Best Poster Award.

[9] W. Philips, "Adaptive baseline correction of EEGs using warped polynomials," in *Proceedings of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 853–854, Sept. 1995. Montréal, Canada.

[10] W. Philips, "Adaptive noise removal from biomedical signals using warped polynomials," *IEEE Trans. Biomed. Eng.*, 1996. To be published.

[11] W. Philips, "Warped polynomial approximation for adaptive filtering of biomedical signals," in *Proceedings of the International Conference on Digital Signal Processing (DSP95)*, vol. 2, pp. 896–901, June 1995. Limassol, Cyprus.

[12] W. Philips, "Adaptive warped polynomial base line correction of EEG signals," in *Proceedings of the VII Mediterranean Conference on Medical and Biological Engineering*, p. 158, Sept. 1995. Jerusalem, Israel.

[13] W. Philips and G. Van Hoey, "A new adaptive filter for biomedical applications," 1996. IASTED International Conference on Signal and Image Processing and Applications (SIPA96). Submitted.

[14] W. Philips, "Codering en filtering van biomedische signalen." *Symposium van de Genootschap Biomedische Techniek & Gezondheidszorg*, Leuven. In Dutch, Jan. 1995.

[15] W. Philips and G. De Jonghe, "Data compression of ECGs by high-degree polynomial approximation," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 330–337, Apr. 1992.

[16] W. Philips and G. De Jonghe, "Polynomial approximation for data reduction on ECGs," in *Proc. MEDICON 89*, pp. 222–223, 1989. Patras, Greece.

[17] W. Philips, "A multi-lead ECG data compression method based on polynomial transform," in *Proc. Second IASTED International Conference on Computer Applications in Industry*, pp. 282–285, May 1992. Alexandria.

[18] W. Philips and M. Vanwormhoudt, "Low-bitrate ECG coding using time-warped polynomials," in *M.I.C. '93 Proceedings, Communication and Integration in Healthcare Informatics*, pp. 127–136, Nov. 1993. Medische Informatika Congres 1993, Gent.

[19] W. Philips, "ECG data compression with time-warped polynomials," Tech. Rep. DG 93-05, ELIS, RUG, Gent, Belgium, 1993.

[20] W. Philips, "Fast coding of arbitrarily shaped image segments using weakly separable bases," *Optical Engineering*, vol. 35, pp. 177–186, Jan. 1996. Special section on Visual Communications and Image Processing.

[21] W. Philips, "Weakly separable bases for fast segmented image coding," in *Hybrid Image and Signal Processing IV*, vol. 2238, pp. 153–163, Apr. 1994. SPIE Conference, Orlando.

[22] W. Philips, "A comparison of segmented image coding and warped polynomial image coding," in *Proceedings of the Ninth IMDSP Workshop*, Mar. 1996. Belize. To be publixhed.

[23] W. Philips, "Polynomial and warped polynomial image and object descriptions," in *Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 263–268, Oct. 1995. Vancouver, Canada. Invited paper.

[24] W. Philips, "Adaptive warped polynomial contour descriptors," in *Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, pp. 4638–4643, Oct. 1995. Vancouver, Canada.

[25] W. Philips, "Adaptive contour approximation by warped polynomials," in *Proceedings of the Second Asian Conference on Computer Vision (ACCV '95)*, vol. 3, pp. 201–205, Dec. 1995. Singapore.

[26] W. Philips, "Adaptive contour coding using warped polynomials," in *Proceedings of ICASSP96*, May 1996. Atlanta. To be published.

[27] W. Philips, "Adaptive orthogonal transform for image data compression," in *Signal Processing VI: Theories and Applications*, vol. 3, (Amsterdam), pp. 1255–1258, Elsevier, Aug. 1992. Proceedings of EUSIPCO-92, Brussels.

[28] W. Philips, "Orthogonal warped polynomials, their applications in signal processing and some open problems." *NFWO-contactgroup Special Functions and Their Applications*, Namen, Belgium, Dec. 1994.

[29] W. Philips, "Model dependent adaptive transform," Tech. Rep. DG 91-16, ELIS, RUG, Gent, Belgium, 1991.

[30] W. Philips, "Recursive computation of polynomial transform coefficients," *Electronic Letters*, vol. 27, no. 25, pp. 2337–2339, 1991.

[31] W. Philips, "A new fast algorithm for moment computation," *Pattern Recognition*, vol. 26, pp. 1619–1621, Nov. 1993.

[32] W. Philips, "Datacompressie van beelden en signalen." Presentation, IEEE Stundent Chapter, Gent (In Dutch), Nov. 1995.

[33] W. Philips and K. Denecker, "Adaptive warped polynomial filtering of biomedical signals," in *Proceedings of the PRORISC IEEE Benelux Workshop on Circuits, Systems and Signal Processing*, pp. 235–242, Mar. 1995. Mierlo, the Netherlands.

[34] W. Philips, D. Torfs, and J. Van Overloop, "An experimental comparison of segmented image coding and warped polynomial image coding," 1996. Conference on Digital Compression Technologies & Systems for Video Communications. Submitted.

[35] B. Simaeys, "Compressie van signalen met variërende bandbreedte: Wavelets vs. vervormde polynomen." Engineering thesis, ELIS-RUG, St.-Pietersnieuwstraat 41, B9000 Gent, Belgium, 1993–1994.

[36] C. Christopoulos, W. Philips, A. Skodras, and J. Cornelis, "Discrete cosine transform coding of images," in *Proc. of the International Conference on DSP and CAES*, pp. 164–169, July 1993. Nicosia, Cyprus.

[37] R. Baraniuk and D. Jones, "Shear madness: New orthonormal bases and frames using chirp functions," *IEEE Transactions on Signal Processing (Special Issue on Wavelets in Signal Procesing)*, vol. 41, pp. 3543–354, Dec. 1993.

[38] P. Delsarte, B. Macq, and D. Slock, "Signal-adapted multiresolution transform for image coding," *IEEE Trans. Inf. Theory*, vol. 38, pp. 897–904, Mar. 1992.

[39] J. Weaver and D. Healy, "Acquisition of the Karhunen-Loève expansion to reduce MR imaging times," in *Proceedings of ICIP94*, vol. 3, pp. 35–39, Nov. 1994. Austin, Texas.

[40] R. Baraniuk and D. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, pp. 2269–2282, Oct. 1995.

[41] P. Jones, S. Daly, R. Gaborski, and M. Rabbani, "Comparative study of wavelet and DCT decompositions with equivalent quantization and encoding strategies for medical images," in *Medical Imaging 1995: Image Display*, vol. 2431, pp. 571–582, Feb. 1995. SPIE Conference, San Diego.

[42] S. Mitra, S. Chakrabarti, and E. Abreu, "Nonuniform discrete Fourier transform and its applications in signal processing," in *Signal Processing VI: Theories and Applications*, vol. 2, (Amsterdam), pp. 909–912, Elsevier, Aug. 1992. Proceedings of EUSIPCO-92, Brussel.