Hybrid Wavelet–Support Vector Classifiers

D. Strauß and G. Steidl

264/01
July 2001

Hybrid Wavelet–Support Vector Classifiers

Daniel Strauß and Gabriele Steidl
University of Mannheim
Faculty of Mathematics and Computer Science
D–68131 Mannheim
Germany
strauss@keynumerics.com
steidl@math.uni-mannheim.de

Abstract

The Support Vector Machine (SVM) represents a new and very promising technique for machine learning tasks involving classification, regression or novelty detection. Improvements of its generalization ability can be achieved by incorporating prior knowledge of the task at hand.
We propose a new hybrid algorithm consisting of signal–adapted wavelet decompositions and SVMs for waveform classification. The adaptation of the wavelet decompositions is tailormade for SVMs with radial basis functions as kernels. It allows the optimization of the representation of the data before training the SVM and does not suffer from computationally expensive validation techniques.
We assess the performance of our algorithm against the background of current concerns in medical diagnostics, namely the classification of endocardial electrograms and the detection of otoacoustic emissions. Here the performance of SVMs can significantly be improved by our adapted preprocessing step.

1991 *Mathematics Subject Classification.* 90C90, 65T60, 90C20, 90C59, 90C30

Keywords. Support vector machines, radial basis functions, reproducing kernel Hilbert spaces, wavelets, adapted filter banks, frames, waveform recognition

1 Introduction

The Support Vector Machine (SVM) is a novel type of learning machine and represents a very promising tool for solving pattern recognition tasks. The SVM is well motivated from statistical learning theory and minimizes an upper bound on the generalization error [49, 6]. However, improvements of its performance can still be achieved by using prior knowledge of the pattern recognition task at hand. In [32] invariances in SVMs were incorporated by the so–called 'virtual support vectors', a SVM adjusted scheme of the known technique of learning with 'virtual training examples' [26]. Invariances and prior knowledge about the locality in images were used in [34] for the construction of appropriate kernels for SVMs, see also [3].
Wavelet decompositions allow for a time–scale analysis of functions and have gained a lot of interest in digital signal processing. Recently, overcomplete wavelet decompositions have proven to be an efficient data representation for SVMs for object detection tasks [23, 24]. Independently from SVMs, a lot of feature extraction schemes which employ fast orthonormal wavelet decompositions based on filter banks are known, see, e.g., [16, 25, 46]. Many wavelet

based feature extraction schemes deal with an adaptation of the wavelet decomposition tree, e.g., the local discriminat basis algorithm of Saito and Coifman [29]. For a recent comparison paper we refer to [28]. In the cited wavelet based feature extraction schemes and the experiments in [29, 28] standard wavelets from wavelet theory were employed. These wavelets are originally designed to optimize some properties, e.g., smoothness conditions, which are not needed in pattern recognition in general. In recent studies one of the authors has shown that a filter bank adaptation based on the lattice structure, a well accepted adaptation scheme for signal coding [10], can significantly improve the performance of a wavelet based feature extraction for waveform classification tasks [42, 41].
Wavelet based feature extraction schemes aim in the first place at a reduction of the dimensionality of the input space to tackle with the 'curse of dimensionality', i.e., proliferation of parameters, which results in immense resources and/or overfitting. SVMs do not depend on dimensionality in general. Nevertheless, recent work has shown that SVMs can indeed suffer in high dimensional spaces where many features of the input space are irrelevant and dimensionality reduction due to feature selection leads to an enhanced SVM performance [55, 14].
In this paper we couple the idea of SVM learning and adapted wavelet representations for the extraction of low dimensional feature vectors for binary waveform classification. Our adaptation strategy is based on a discriminat functional which is well motivated from the paradigm of 'large margin classifiers' underlying the SVM approach. For investigating our approach on real world data, we adopt current tasks from medical diagnostics. In particular, we deal with a rate independent arrhythmia recognition in electrocardiology, where we incorporate local instabilities in time in our learning task and with the detection of otoacoustic emissions in audiology, where we need a shift–invariant classification scheme.
This paper is organized as follows. In the Sections 2 and 3 we provide the necessary material concerning SVM classifications and wavelet decompositions by parameterized paraunitary filter banks. Section 4 presents the adaptation of the parameterized filter banks to SVMs with kernels arising from *radial basis functions* (RBFs). In Section 5 we introduce RBFs with compact support which have not been used for the construction of SVMs up to now. The Sections 6 and 7 deal with applications of our algorithm in medical diagnostics. Section 6 gives an assessment of the algorithm for the waveform recognition in electrocardiology. Section 7 presents a shift–invariant approach to classify signals in audiology. The conclusions of the paper are given in Section 8.

2 Support Vector Machine Classification

In this section we provide the tools concerning the support vector machine classification with respect to the applications we have in mind. Our approach is based on the pioneering work of Vapnik [49] and the new book of Christianini and Shawe–Taylor [6], where the reader can find a detailed introduction in terms of statistical learning theory.
Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$ containing the data to be classified. We suppose that there exists an underlying unknown function t, the so–called *target function*, which maps $\mathcal{X}$ to the binary set $\{-1, 1\}$. Given a training set

$$\mathcal{A} := \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \{-1, 1\} : i = 1, \dots, M\} \tag{1}$$

of M associations we are interested in in the construction of a real valued function f defined on $\mathcal{X}$ such that $\operatorname{sgn}(f)$ is a 'good approximation' of t which classifies the training data correctly,

i.e., $\operatorname{sgn}(f(\mathbf{x}_i)) = t(\mathbf{x}_i) = y_i$ for all $i = 1, \ldots, M$. Here

$$\operatorname{sgn}(f(\mathbf{x})) := \begin{cases} 1 & \text{if } f(\mathbf{x}) \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

We will search for f in some reproducing kernel Hilbert spaces which we will introduce next. By $L^2(\mathcal{X})$ we denote the Hilbert space of real valued square integrable functions on $\mathcal{X}$ with inner product $\langle f, g \rangle_{L^2} = \int_{\mathcal{X}} f(x)g(x)\,\mathrm{d}x$. Let $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ be a positive definite symmetric function in $L^2(\mathcal{X} \times \mathcal{X})$. Following [30], we call a function $K \in L^2(\mathcal{X} \times \mathcal{X})$ positive definite iff for any finite set of elements $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathcal{X}$, the matrix $(K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^n$ is positive definite. In this paper we are only interested in functions K arising from RBFs. In other words, we assume that there exists a real valued function k on $\mathbb{R}$ so that

$$K(\mathbf{x}, \mathbf{y}) = k(||\mathbf{x} - \mathbf{y}||_2), \tag{2}$$

where $||\cdot||_2$ denotes the Euclidean norm on $\mathbb{R}^d$. In our applications we will use Gaussian kernels and Wendland's compactly supported RBFs [54]. The latter were not applied in connection with classification tasks up to now.
For a given K, there exists a *reproducing kernel Hilbert space*

$$\mathcal{H}_K := \overline{\operatorname{span}\{K(\tilde{\mathbf{x}}, \cdot) : \tilde{\mathbf{x}} \in \mathcal{X}\}}$$

of real valued functions on $\mathcal{X}$ with inner product determined by

$$\langle K(\tilde{\mathbf{x}}, \mathbf{x}), K(\bar{\mathbf{x}}, \mathbf{x}) \rangle_{\mathcal{H}_K} := K(\tilde{\mathbf{x}}, \bar{\mathbf{x}}) \tag{3}$$

which has reproducing kernel K, i.e.,

$$\langle f(\cdot), K(\tilde{\mathbf{x}}, \cdot) \rangle_{\mathcal{H}_K} = f(\tilde{\mathbf{x}}) \quad (f \in \mathcal{H}_K).$$

By *Mercer's Theorem*, K can be expanded in a uniformly convergent series on $\mathcal{X} \times \mathcal{X}$

$$K(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{\infty} \eta_j \varphi_j(\mathbf{x}) \varphi_j(\mathbf{y}), \tag{4}$$

where $\eta_j \geq 0$ are the eigenvalues of the integral operator $T_K : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ with $T_K f(\mathbf{y}) := \int_{\mathcal{X}} K(\mathbf{x}, \mathbf{y}) f(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and where $\{\varphi_j\}_{j \in \mathbb{N}}$ are the corresponding $L^2(\mathcal{X})$–orthonormalized eigenfunctions.
We introduce a so–called *feature map* $\Phi : \mathcal{X} \to \ell^2$ by

$$\Phi(\cdot) := \left(\sqrt{\eta_j}\varphi_j(\cdot)\right)_{j \in \mathbb{N}}.$$

Let ℓ^2 denote the Hilbert space of real valued quadratic summable sequences $a = (a_i)_{i \in \mathbb{N}}$ with inner product $\langle a, b \rangle_{\ell^2} = \sum_{i \in \mathbb{N}} a_i b_i$. By (4), we have that $\Phi(\mathbf{x})$ $(\mathbf{x} \in \mathcal{X})$ is an element in ℓ^2 with

$$||\Phi(\mathbf{x})||_{\ell^2}^2 = \sum_{j=1}^{\infty} \eta_j \varphi_j^2(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}) = k(0).$$

We define the *feature space* $\mathcal{F}_K \subset \ell^2$ by the ℓ^2–closure of all finite linear combinations of elements $\Phi(\mathbf{x})$ $(\mathbf{x} \in \mathcal{X})$

$$\mathcal{F}_K := \overline{\operatorname{span}\{\Phi(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}}.$$

Then $\mathcal{F}_K$ is a Hilbert space with $||\cdot||_{\mathcal{F}_K} = ||\cdot||_{\ell^2}$. The feature space $\mathcal{F}_K$ and the reproducing kernel Hilbert space $\mathcal{H}_K$ are isometrically isomorph with isometry $\iota : \mathcal{F}_K \to \mathcal{H}_K$ defined by

$$\iota(\mathbf{w}) := f_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle_{\ell^2} = \sum_{j=1}^{\infty} w_j \sqrt{\eta_j} \varphi_j(\mathbf{x}). \tag{5}$$

In particular, we have that

$$||f_{\mathbf{w}}||_{\mathcal{H}_K} = ||\mathbf{w}||_{\mathcal{F}_K}. \tag{6}$$

Let us turn to our classification task. For a given training set (1) we intend to construct a function $f \in \mathcal{H}_K$ which minimizes

$$\lambda \sum_{i=1}^{M} (1 - y_i f(\mathbf{x}_i))_+ + \frac{1}{2} ||f||^2_{\mathcal{H}_K}, \tag{7}$$

where

$$(\tau)_+ = \begin{cases} \tau & \text{if } \tau \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that we can also look for functions of the form $f = h + b$ ($h \in \mathcal{H}_K$) with a so–called bias term $b \in \mathbb{R}$. We omit the bias term b here, because its explicit consideration does not lead to an improvement of our numerical results.
The unconstrained optimization problem (7) is equivalent to the following constraint optimization problem: find $f \in \mathcal{H}_K$ and u_i $(i = 1, \ldots, M)$ to minimize

$$\lambda \left(\sum_{i=1}^{M} u_i \right) + \frac{1}{2} ||f||^2_{\mathcal{H}_K}, \tag{8}$$

subject to

$$\begin{aligned} y_i f(\mathbf{x}_i) &\geq 1 - u_i \quad (i = 1, \ldots, M), \\ u_i &\geq 0 \quad (i = 1, \ldots, M). \end{aligned}$$

Every function $f \in \mathcal{H}_K$ corresponds uniquely to a sequence $\mathbf{w} \in \mathcal{F}_K$. Thus, by (5) and (6), the optimization problem (8) can be rewritten as follows: find $\mathbf{w} \in \mathcal{F}_K$ and u_i $(i = 1, \ldots, M)$ to minimize

$$\lambda \left(\sum_{i=1}^{M} u_i \right) + \frac{1}{2} ||\mathbf{w}||^2_{\mathcal{F}_K}, \tag{9}$$

subject to

$$\begin{aligned} y_i \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathcal{F}_K} &\geq 1 - u_i \quad (i = 1, \ldots, M), \\ u_i &\geq 0 \quad (i = 1, \ldots, M). \end{aligned} \tag{10}$$

In general the feature space $\mathcal{F}_K \subset \ell^2$ is infinitely dimensional. For a better illustration of (9) we assume for a moment that $\mathcal{F}_K \subset \mathbb{R}^n$. Then the function $\tilde{f}_{\mathbf{w}}(\mathbf{v}) := \langle \mathbf{w}, \mathbf{v} \rangle_{\mathcal{F}_K}$ defines a hyperplane $H_{\mathbf{w}} := \{\mathbf{v} \in \mathcal{F}_K : \tilde{f}_{\mathbf{w}}(\mathbf{v}) = 0\}$ in $\mathbb{R}^n$ through the origin and an arbitrary point

$\mathbf{v}_i \in \mathcal{F}_K$ has the distance $|\langle \mathbf{w}, \mathbf{v}_i \rangle_{\mathcal{F}_K}|/||\mathbf{w}||_{\mathcal{F}_K}$ from $H_{\mathbf{w}}$. Note that $\tilde{f}_{\mathbf{w}}(\Phi(\mathbf{x})) = f_{\mathbf{w}}(\mathbf{x})$. Thus, the constraints $y_i \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathcal{F}_K}/||\mathbf{w}||_{\mathcal{F}_K} \geq 1/||\mathbf{w}||_{\mathcal{F}_K} - u_i/||\mathbf{w}||_{\mathcal{F}_K}$ $(i = 1, \ldots, M)$ in (10) require that every $\Phi(\mathbf{x}_i)$ must at least have the distance $1/||\mathbf{w}||_{\mathcal{F}_K} - u_i/||\mathbf{w}||_{\mathcal{F}_K}$ from $H_{\mathbf{w}}$. If there exists $\mathbf{w} \in \mathcal{F}_K$ so that (10) can be fulfilled with $u_i = 0$ $(i = 1, \ldots, M)$, then we say that our training set is linearly separable in $\mathcal{F}_K$. Of course, for Gaussian kernels or kernels arising from Wendland's compactly supported radial basis function every finite training set is linearly separable in $\mathcal{F}_K$, see, e.g., [2] and [38]. Then the optimization problem (9) can be further simplified to: find $\mathbf{w} \in \mathcal{F}_K$ to minimize

$$\frac{1}{2}||\mathbf{w}||_{\mathcal{F}_K} \tag{11}$$

subject to

$$y_i \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathcal{F}_K} \geq 1 \quad (i = 1, \ldots, M).$$

Given $\mathcal{H}_K$ and $\mathcal{A}$, the optimization problem above has a unique solution $f_{\mathbf{w}^*}$. In our hyperplane context $H_{\mathbf{w}^*}$ is exactly the hyperplane which has maximal distance γ from the training data, where

$$\gamma := \frac{1}{||\mathbf{w}^*||_{\mathcal{F}_K}} = \frac{1}{||f_{\mathbf{w}^*}||_{\mathcal{H}_K}} = \max_{\mathbf{w} \in \mathcal{F}_K} \min_{i=1,\ldots,M} \left\{ \frac{|\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathcal{F}_K}|}{||\mathbf{w}||_{\mathcal{F}_K}} \right\}. \tag{12}$$

The value γ is called the *margin* of $f_{\mathbf{w}^*}$ with respect to the training set $\mathcal{A}$. In this context, the solutions of the optimizations problems (9) and (11) are called *soft margin* and *hard margin SVM classifiers*, respectively. See Figure 1 for an illustration of the hard margin case for $n = 2$.

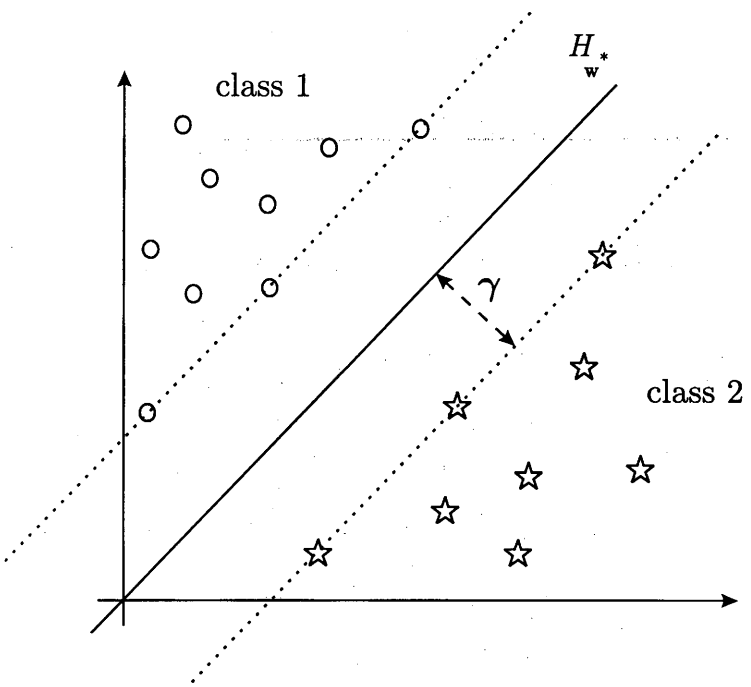

Figure 1: The separation of two classes by an optimal hyperplane $H_{\mathbf{w}^*}$ with margin γ.

Remark 1. There exists an important relation between the margin of the SVM classifier and its *generalization error*, i.e., the probability that $\operatorname{sgn}(f_{\mathbf{w}^*}(\mathbf{x})) \neq y$ for a randomly chosen example $(\mathbf{x}, y) \in \mathcal{X} \times \{-1, 1\}$, which motivates the SVM approach and our further investigations.

By [6, Theorem 4.18], the generalization error of the hard margin SVM classifier decreases if the margin γ increases. In other words: *the larger the margin γ the better generalization of the SVM can be expected.*
Note that there exist also estimates for the generalization error of soft margin SVM classifiers which involve the margin of the unknown target function, see, e.g., [6, Theorem 4.21] and [38].

Next we consider the solution of (11), where we follow mainly the lines of [51]. Here the notation 'support vector' comes into the play.
By the *Representer Theorem* ([20, 51]), the minimizer of (9) has the form

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j K(\mathbf{x}, \mathbf{x}_j). \tag{13}$$

Setting $\mathbf{f} := (f(\mathbf{x}_1), \dots, f(x_M))^T$, $\mathbf{K} := (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^M$ and $\mathbf{c} := (c_1, \dots, c_M)^T$ we obtain that

$$\mathbf{f} = \mathbf{K}\mathbf{c}.$$

Note that $\mathbf{K}$ is positive definite. Further, let $\mathbf{Y} := \operatorname{diag}(y_1, \dots, y_M)$ and $\mathbf{u} := (u_1, \dots, u_M)^T$. By $\mathbf{0}$ and $\mathbf{e}$ we denote the vectors with M entries 0 and 1, respectively. Then the optimization problem (9) can be rewritten as

$$\min_{\mathbf{u},\mathbf{c}} \ \lambda \mathbf{e}^T \mathbf{u} + \frac{1}{2}\mathbf{c}^T \mathbf{K}\mathbf{c} \tag{14}$$

subject to

$$\begin{aligned} \mathbf{u} &\geq \mathbf{e} - \mathbf{Y}\mathbf{K}\mathbf{c}, \\ \mathbf{u} &\geq \mathbf{0}. \end{aligned}$$

The dual problem with Lagrange multipliers $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^T$ and $\boldsymbol{\beta} = (\beta_1, \dots \beta_M)^T$ reads

$$\max_{\mathbf{c},\mathbf{u},\boldsymbol{\alpha},\boldsymbol{\beta}} L(\mathbf{c}, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

where

$$L(\mathbf{c}, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) := \lambda \mathbf{e}^T \mathbf{u} + \frac{1}{2}\mathbf{c}^T \mathbf{K}\mathbf{c} - \boldsymbol{\beta}^T \mathbf{u} + \boldsymbol{\alpha}^T \mathbf{e} - \boldsymbol{\alpha}^T \mathbf{Y}\mathbf{K}\mathbf{c} - \boldsymbol{\alpha}^T \mathbf{u}$$

subject to

$$\frac{\partial L}{\partial \mathbf{c}} = \mathbf{0}, \ \frac{\partial L}{\partial \mathbf{u}} = \mathbf{0}, \ \boldsymbol{\alpha} \geq \mathbf{0}, \ \boldsymbol{\beta} \geq \mathbf{0}.$$

Now $\mathbf{0} = \frac{\partial L}{\partial \mathbf{c}} = \mathbf{K}\mathbf{c} - \mathbf{K}\mathbf{Y}\boldsymbol{\alpha}$ yields

$$\mathbf{c} = \mathbf{Y}\boldsymbol{\alpha}. \tag{15}$$

Further we have by $\frac{\partial L}{\partial \mathbf{u}} = \mathbf{0}$ that $\boldsymbol{\beta} = \lambda \mathbf{e} - \boldsymbol{\alpha}$. Thus, our optimization problem becomes

$$\max_{\boldsymbol{\alpha}} \left(-\frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Y}\mathbf{K}\mathbf{Y}\boldsymbol{\alpha} + \mathbf{e}^T \boldsymbol{\alpha} \right) \tag{16}$$

subject to

$$\mathbf{0} \leq \boldsymbol{\alpha} \leq \lambda \mathbf{e}.$$

This quadratic programming (QP) problem is usually solved in the SVM literature. For a moderate number of associations some standard QP routines can be used and for a large number of associations, e.g., $|\mathcal{A}| > 4000$, specifically designed large scale algorithms should be applied, e.g., SVM*light* [17].
The *support vectors* (SVs) are those training patterns $\mathbf{x}_i$ for which α_i does not vanish. Let I denote the index set of the support vectors $I := \{i \in \{1, \dots, M\} : \alpha_i \neq 0\}$ then by (13) and (15), the function f has the sparse representation

$$f(\mathbf{x}) = \sum_{i \in I} c_i K(\mathbf{x}_i, \mathbf{x}) = \sum_{i \in I} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x})$$

which depends only on the SVs. With respect to the margin we obtain by (12) and (3) that

$$\gamma = (||f||_{\mathcal{H}_K})^{-1} = (\mathbf{c}^T \mathbf{K} \mathbf{c})^{-1/2} = \left(\sum_{i \in I} y_i \alpha_i f(\mathbf{x}_i) \right)^{-1/2}.$$

Due to the Kuhn–Tucker conditions [11] the solution f of the QP problem (14) has to fulfill

$$\alpha_i (1 - y_i f(\mathbf{x}_i) - u_i) = 0 \quad (i = 1 \dots, M).$$

In case of hard margin classification with $u_i = 0$ this implies that $y_i f(\mathbf{x}_i) = 1$ $(i \in I)$ so that we obtain the following simple expression for the margin

$$\gamma = \left(\sum_{i \in I} \alpha_i \right)^{-\frac{1}{2}}. \tag{17}$$

Remark 2. By [6, Theorem 6.8], the number of SVs can also be used to give an upper bound of the generalization error. *The fewer the number of support vectors the better generalization of the SVM can be expected.*

3 Wavelet Decompositions by Parameterized Paraunitary Filter Banks

In this section we give a short introduction to paraunitary filter banks and corresponding so called discrete–time wavelets. The first part mainly builds up on [40]. Here we prefer the use discrete–time wavelets in ℓ^2 instead of wavelets in $L^2(\mathbb{R})$ since this approach is straightforward for digital signal processing. A broader introduction to the topic can be found in [50, 8].
Let $H_0(z) := \sum_{k \in \mathbb{Z}} h_0[k] z^{-k}$ be the z–transform of the analysis lowpass filter and $H_1(z) := \sum_{k \in \mathbb{Z}} h_1[k] z^{-k}$ the z–transform of the analysis highpass filter of a two–channel filter bank with real–valued filter coefficients. Throughout this paper, we use a capital letter to denote a function in the z–transform domain and the corresponding small letter to denote its time–domain version. A filter bank with analysis filters H_0 and H_1 is called *paraunitary* (sometimes also referred as *orthogonal*) iff

$$H_0(z^{-1}) H_0(z) + H_1(z^{-1}) H_1(z) = 2, \tag{18}$$
$$H_0(z^{-1}) H_0(-z) + H_1(z^{-1}) H_1(-z) = 0. \tag{19}$$

The corresponding synthesis filters are given by

$$G_0(z) = H_0(z^{-1}), \; G_1(z) = H_1(z^{-1}).$$

The *polyphase matrix* of a paraunitary filter bank

$$\mathbf{H}_{\mathrm{pol}}(z) := \begin{pmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{pmatrix}$$

with entries from the polyphase decomposition

$$H_i(z) = H_{i0}(z^2) + z^{-1} H_{i1}(z^2) \quad (i = 0, 1)$$

satisfies the relation

$$\mathbf{H}_{\mathrm{pol}}^T(z^{-1}) \mathbf{H}_{\mathrm{pol}}(z) = \mathbf{I}_2, \tag{20}$$

where $\mathbf{I}_2$ denotes the 2×2 identity matrix.
We are interested in finite impulse response (FIR) filters of order $2L + 1$ with real–valued coefficients

$$H_i(z) := \sum_{k=0}^{2L+1} h_i[k] z^{-k} \quad (h_i[k] \in \mathbb{R}). \tag{21}$$

For these filters, according to [47], [40, Theorem 4.7], the corresponding polyphase matrix $\mathbf{H}_{\mathrm{pol}}(z)$ can be decomposed into

$$\mathbf{H}_{\mathrm{pol}}(z) = \Big(\prod_{l=0}^{L-1} \begin{pmatrix} \cos\vartheta_l & \sin\vartheta_l \\ -\sin\vartheta_l & \cos\vartheta_l \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix}\Big) \begin{pmatrix} \cos\vartheta_L & \sin\vartheta_L \\ -\sin\vartheta_L & \cos\vartheta_L \end{pmatrix}, \tag{22}$$

where $\vartheta_L \in [0, 2\pi)$ and $\vartheta_l \in [0, \pi)$ $(l = 0, \ldots, L-1)$. Further we assume that the highpass filter has at least *one vanishing moment*, i.e., $H_1(1) = 0$. By (18) and (19) this implies that the lowpass filter satisfies $H_0(-1) = 0$. Using these properties and (20) it is easy to check that

$$\mathbf{H}_{\mathrm{pol}}(1) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Since we have by (22) that

$$\mathbf{H}_{\mathrm{pol}}(1) = \begin{pmatrix} \cos(\sum_{l=0}^{L} \vartheta_l) & \sin(\sum_{l=0}^{L} \vartheta_l) \\ -\sin(\sum_{l=0}^{L} \vartheta_l) & \cos(\sum_{l=0}^{L} \vartheta_l) \end{pmatrix},$$

we obtain

$$\sum_{l=0}^{L} \vartheta_l \equiv \frac{\pi}{4} \pmod{2\pi}.$$

Let ϑ_L be the residue of $\frac{\pi}{4} - \sum_{l=0}^{L-1} \vartheta_l$ modulo 2π in $[0, 2\pi)$. Then the space

$$\mathcal{P}^L := \{\boldsymbol{\vartheta} = (\vartheta_0, \ldots, \vartheta_{L-1}) : \vartheta_l \in [0, \pi)\}$$

can serve to parameterize all two–channel paraunitary filter banks (21) with at least one vanishing moment of the highpass filter. To emphasize this parameterization we will use the superscript $\boldsymbol{\vartheta}$ later. A parameterization with more than one vanishing moment of the highpass filter can be realized by the method of Zou et al. [56]. Filter banks having the

representation in (22) can efficiently be implemented by the *lattice structure* [48]. A special and even more efficient implementation is called the *two multiplier lattice* [47]. Note that the parameterization can also be implemented by lifting steps [9] which are frequently used for designing biorthogonal filter banks. However, we rely on the implementation based on the lattice structure due its availability in already existing very efficient architectures, e.g., [19], which we will especially need when dealing with algorithms for low–power devices.

For a fixed $\vartheta \in \mathcal{P}^L$ let $G_0 = G_0^\vartheta$ and $G_1 = G_1^\vartheta$ be the synthesis filters of a two–channel paraunitary filter bank implemented by the lattice structure. When cascading these filters in an octave band tree, the synthesis filters of an equivalent parallel structure on level $j = 1, \dots, J$ are given by

$$Q_{j,0}(z) = \prod_{m=0}^{j-1} G_0(z^{2^m}), \tag{23}$$

$$Q_{j,1}(z) = G_1(z^{2^{j-1}}) \prod_{m=0}^{j-2} G_0(z^{2^m}). \tag{24}$$

According to [50] we introduce the *discrete–time scaling sequences* and *wavelets* on scale j by the impulse responses $\mathbf{q}_{j,0} = (q_{j,0}[k])_{k\in\mathbb{Z}}$ and $\mathbf{q}_{j,1} = (q_{j,1}[k])_{k\in\mathbb{Z}}$ of the filters (23) and (24), respectively. Let $\mathbf{q}_{j,i}^m := \big(q_{j,i}[k - 2^j m]\big)_{k\in\mathbb{Z}}$ $(i = 0, 1)$ denote the translation of $\mathbf{q}_{j,i}$ by $2^j m$ samples. The paraunitarity of the filter bank implies the following orthogonality relations of the scaling sequences and wavelets:

$$\begin{aligned} \langle \mathbf{q}_{j,0}^0, \mathbf{q}_{j,0}^m \rangle_{\ell^2} &= \delta[m], \\ \langle \mathbf{q}_{i,1}^m, \mathbf{q}_{j,1}^n \rangle_{\ell^2} &= \delta[i-j]\delta[m-n], \\ \langle \mathbf{q}_{j,0}^0, \mathbf{q}_{j,1}^m \rangle_{\ell^2} &= 0, \end{aligned} \tag{25}$$

$(m, n \in \mathbb{Z};\ i, j = 1, \dots, J)$, where $\delta[m] = 1$ if $m = 0$ and $\delta[m] = 0$ otherwise. We introduce the spaces $\Omega_{0,0} := \ell^2$ and

$$\Omega_{j,0} = \overline{\operatorname{span}\{\mathbf{q}_{j,0}^m \,:\, m \in \mathbb{Z}\}}, \quad \Omega_{j,1} = \overline{\operatorname{span}\{\mathbf{q}_{j,1}^m \,:\, m \in \mathbb{Z}\}}.$$

Note that $\{\mathbf{q}_{j,i}^m : m \in \mathbb{Z}\}$ forms an orthonormal basis of $\Omega_{j,i}$ $(i = 0, 1)$. Further we have by (23) that $\Omega_{J,0} \subset \cdots \subset \Omega_{2,0} \subset \Omega_{1,0} \subset \Omega_{0,0}$ and by (24) and (25) that

$$\Omega_{j-1,0} = \Omega_{j,0} \oplus \Omega_{j,1},$$

where $\oplus$ denotes the orthogonal sum. Thus, the space ℓ^2 can be decomposed as $\ell^2 = \Omega_{J,0} \oplus \bigoplus_{j=1}^J \Omega_{j,1}$ and the set

$$\big\{\mathbf{q}_{J,0}^m, \mathbf{q}_{j,1}^m \,:\, j = 1, \dots, J; m \in \mathbb{Z}\big\} \tag{26}$$

constitutes an orthonormal basis for ℓ^2. With respect to this basis an arbitrary sequence $\mathbf{x} \in \ell^2$ can be decomposed as

$$\mathbf{x} = \sum_{m\in\mathbb{Z}} d_{J,0}[m]\, \mathbf{q}_{J,0}^m + \sum_{j=1}^{J} \sum_{m\in\mathbb{Z}} d_{j,1}[m] \mathbf{q}_{j,1}^m \tag{27}$$

with the *wavelet coefficients*

$$d_{j,i}[m] = \langle \mathbf{x}, \mathbf{q}_{j,i}^m \rangle_{\ell^2} \quad (i = 0, 1).$$

We set

$$\mathbf{d}_j := (d_{j,1}[m])_{m \in \mathbb{Z}} \quad (j = 1, \ldots, J).$$

We will need norms of these wavelet coefficient vectors later.
Some applications, e.g., our classification task in Section 6 require a shift–invariant multilevel decomposition (27). This can be achieved by replacing the orthonormal wavelet basis (26) by the tight wavelet frame

$$\{2^{-J/2}\tilde{\mathbf{q}}_{J,0}^m, 2^{-j/2}\tilde{\mathbf{q}}_{j,1}^m \,:\, j = 1, \ldots, J; m \in \mathbb{Z}\},$$

where $\tilde{\mathbf{q}}_{j,i}^m := (q_{j,i}[k-m])_{k \in \mathbb{Z}}$ $(i = 0, 1)$. Then $\mathbf{x} \in \ell^2$ can be decomposed as

$$\mathbf{x} = \sum_{m \in \mathbb{Z}} \tilde{d}_{J,0}[m]\, \tilde{\mathbf{q}}_{J,0}^m + \sum_{j=1}^{J} \sum_{m \in \mathbb{Z}} \tilde{d}_{j,1}[m] \tilde{\mathbf{q}}_{j,1}^m \tag{28}$$

with the coefficients

$$\tilde{d}_{j,i}[m] = \frac{1}{2^j} \langle \mathbf{x}, \tilde{\mathbf{q}}_{j,i}^m \rangle_{\ell^2} \quad (i = 0, 1).$$

We set

$$\tilde{\mathbf{d}}_j := \left(\tilde{d}_{j,1}[m]\right)_{m \in \mathbb{Z}} \quad (j = 1, \ldots, J).$$

Overcomplete expansions can be implemented by oversampled paraunitary filter banks [7, 1] The highly redundant expansion (28) corresponds to a nonsubsampled filter bank, i.e., we have no multirate operations at all. In this special case, the subbands are obtained by pure linear time–invariant (LTI) filters given by (23) and (24), respectively. Note that an implementation as cascaded two–channel building blocks requires to insert $2^j - 1$ zeros between the nonzero coefficients of the filters at the levels $j = 2, \ldots, J$. This procedure is also known as 'algorithm a trous' [36] and is equivalent to the so–called 'cyclic spinning' [5]. The computation of the frame coefficients vectors $\tilde{\mathbf{d}}_j$ requires $\mathcal{O}(N \log N)$ arithmetic operations instead of $\mathcal{O}(N)$ for the computation of $\mathbf{d}_j$ $(j = 1, \ldots, \log_2 N)$. There exist nearly shift–invariant approaches with lower arithmetic complexity, e.g. [21]. The incorporation of these algorithms into our adaptation scheme may be a future point of research.

In the following we emphasize the dependence of the wavelet coefficients on the chosen angles $\boldsymbol{\vartheta} \in \mathcal{P}^L$ by the superscript $\boldsymbol{\vartheta}$.

4 Adaptation to Waveforms

Recent studies [52, 42, 41] have shown that the so–called *multilevel concentrations* $||\cdot||_{\ell^p}^p$ $(1 \leq p < \infty)$ of coefficient vectors of wavelet–like decompositions in distinct levels provide reliable feature vectors for waveform classification tasks. This mainly steams from the fact that such feature vectors help to tackle with local instabilities in time. They are therefore more robust than the consideration of specific Heisenberg cells in the time–scale domain. The multilevel concentration is a robust but global feature and can be insensitive to slight morphological dissimilarities of waveforms belonging to distinct classes when non–adapted decompositions are utilized [42, 41]. Here we present an adaptation of wavelet decompositions

that is tailormade for SVM classifiers. We restrict our attention to orthonormal wavelet decompositions. The generalization to frame decompositions is straightforward.

Let a set of M waveforms $\mathbf{x}_i \in \Omega_{0,0} \subset \mathbb{R}^N$ be given which belong to two distinct classes with corresponding labels $y_i \in \{-1, 1\}$ $(i = 1, \dots, M)$. By M_+ and M_- we denote the sets of indices $i \in \{1, \dots, M\}$ with $y_i = 1$ and $y_i = -1$, respectively. In general, the length N of our waveforms will be large, e.g., $N = 512$. We intend to reduce this length while emphasizing the discriminating features of the signals to support classification tasks.
For this we consider wavelet decompositions of our signals as in the previous section. Let J be maximal depth of the wavelet decomposition. Further let $\{j_1, \dots, j_d\} \subset \{1, \dots, J\}$ be the indices of those wavelet coefficient vectors we are interested in. The choice of the relevant levels can be determined by some validation technique or by prior information about the waveforms, e.g., a known pre–filtering.
For a fixed waveform $\mathbf{x} \in \Omega_{0,0}$ with wavelet coefficient vectors $\mathbf{d}_j^{\vartheta}$ we define the function $\boldsymbol{\xi}_{\mathbf{x}} : \mathcal{P}^L \to \mathbb{R}^d$ by

$$\boldsymbol{\xi}_{\mathbf{x}}(\vartheta) = (\xi_1(\vartheta), \xi_2(\vartheta), \dots, xi_d(\vartheta)) := \left(||\mathbf{d}_{j_1}^{\vartheta}||_{\ell^p}^p, ||\mathbf{d}_{j_2}^{\vartheta}||_{\ell^p}^p, \dots, ||\mathbf{d}_{j_d}^{\vartheta}||_{\ell^p}^p \right) \tag{29}$$

and set $\boldsymbol{\xi}_i(\vartheta) := \boldsymbol{\xi}_{\mathbf{x}_i}(\vartheta)$ $(i = 1, \dots, M)$.
Note that for nonsubsampled filter banks, $||\tilde{\mathbf{d}}_{j_k}^{\vartheta}||_{\ell^p}$ and thus $\tilde{\boldsymbol{\xi}}_{\mathbf{x}}(\vartheta)$ do not change if the signal is shifted. In this case our approach becomes completely translation–invariant.
We want to find ϑ so that

$$\mathcal{A}(\vartheta) := \left\{ (\boldsymbol{\xi}_i(\vartheta), y_i) \in \mathcal{X} \subset \mathbb{R}^d \times \{\pm 1\} : i = 1, \dots, M \right\}$$

is a 'good' training set for a SVM. By Remark 1, we can expect a better generalization ability of the SVM if the margin becomes large. Consequently, we try to find $\hat{\vartheta}$ so that

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}^L} \left\{ \min_{i \in M_+, j \in M_-} \left|\left| \Phi(\boldsymbol{\xi}_i(\vartheta)) - \Phi(\boldsymbol{\xi}_j(\vartheta)) \right|\right|_{\mathcal{F}_K} \right\}. \tag{30}$$

By definition of the inner product in $\mathcal{F}_K$ and (2) it follows that

$$\begin{aligned} \left|\left| \Phi(\boldsymbol{\xi}_i(\vartheta)) - \Phi(\boldsymbol{\xi}_j(\vartheta)) \right|\right|_{\mathcal{F}_K} &= ||\Phi(\boldsymbol{\xi}_i(\vartheta))||_{\mathcal{F}_K}^2 + \left|\left|\Phi(\boldsymbol{\xi}_j(\vartheta))\right|\right|_{\mathcal{F}_K}^2 - 2\left\langle \Phi(\boldsymbol{\xi}_i(\vartheta)), \Phi(\boldsymbol{\xi}_j(\vartheta)) \right\rangle_{\mathcal{F}_K} \\ &= 2k(0) - 2k\left(\left|\left| \boldsymbol{\xi}_i(\vartheta) - \boldsymbol{\xi}_j(\vartheta) \right|\right|_2 \right). \end{aligned}$$

We suppose that $k(t)$ is monotonely decreasing in $|t|$. Then (30) can be rewritten as

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}^L} \left\{ \min_{i \in M_+, j \in M_-} \left|\left| \boldsymbol{\xi}_i(\vartheta) - \boldsymbol{\xi}_j(\vartheta) \right|\right|_2 \right\}. \tag{31}$$

Note that the geometry in feature spaces induced by kernels was investigated in [3, 33]. For most waveform recognition tasks the sets M_+ and M_- can be reduced by averaging the patterns of the respective classes (or subsets thereof) or by expert selection of representative subsets. Here we use the first approach and introduce the notation

$$\boldsymbol{\xi}_{\pm} := \frac{1}{|M_{\pm}|} \sum_{i \in M_{\pm}} \boldsymbol{\xi}_i.$$

Instead of (31) we search for $\hat{\vartheta}$ with

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}^L} \left\{ \left|\left| \xi_+(\vartheta) - \xi_-(\vartheta) \right|\right|_2 \right\}. \tag{32}$$

Solving this optimization problem analytically seems to be infeasible. The optimization functional involves the multilevel concentration and seems not to allow for sophisticated optimization strategies. In particular, hill climbing methods in $\mathcal{P}^L$ are doomed to fail due to local minima of the optimization functional. We introduce a discrete grid

$$\mathcal{P}_T^L := \{\vartheta = (\vartheta_0, \ldots, \vartheta_{L-1}) : \vartheta_l \in D\}, \quad D := \left\{ \frac{\pi\sigma}{T} : \sigma = 0, \ldots, T-1 \right\}$$

in $\mathcal{P}^L$ and solve (32) by evaluating the optimization functional at each grid point.
There exist possibilities to reduce the complexity of these computations. For example we can compress the parameter space $\mathcal{P}_T^L$ as follows: Given a positive number $\mu < 1$, we can further reduce the parameter space by selecting a maximal subset $\mathcal{P}_\mu^L$ of $\mathcal{P}_T^L$ so that

$$\frac{1}{d} \sum_{i=1}^{d} \left| \langle \mathbf{q}_{j_i,1}^0(\vartheta), \mathbf{q}_{j_i,1}^0(\vartheta^*) \rangle_{\ell^2} \right| \leq \mu \quad (\vartheta, \vartheta^* \in \mathcal{P}_L^\mu;\ \vartheta \neq \vartheta^*). \tag{33}$$

In other words, the distinct wavelets in (33) satisfy the *strengthened Cauchy–Schwarz inequality*, that is, the smaller μ the more orthogonal the wavelets become. In a way, μ steers the redundancy of our parameter space. If we restrict our interest to smooth wavelets, the consideration of only one level in (33) is sufficient due to their self–similarity across levels. Such a compression of the parameter space can be significant even for a large μ. For instance, when working with $T = 32$, $L = 2$ and $\mu = 0.98$ we have achieved a compression of $|\mathcal{P}_\mu^L|/|\mathcal{P}_T^L| \approx 0.65$ in [42].

5 Radial Basis Functions with Compact Support

In the following sections we will apply two kind of RBFs $k(||\mathbf{x}||_2)$, namely Gaussian RBFs and Wendland's compactly supported RBFs. The Gaussian RBF K with

$$k(t) = e^{-\frac{t^2}{2s^2}},$$

is a positive definite function in $\mathcal{C}^\infty(\mathbb{R}^d)$ for all dimensions d and all $s \in \mathbb{R}_+$. Here $\mathcal{C}^n(\mathbb{R}^d)$ denotes the space of n times continuously differentiable functions on $\mathbb{R}^d$. This kernel is also well accepted for constructing SVMs and provides excellent results for real world applications, see, e.g., [49, 35].
Recently positive definite RBFs with compact support have been constructed by various authors to solve interpolation problems with scattered data $\{\mathbf{x}_i\}_{i=1}^M$, see [31] and the references therein. In the context of scattered data interpolation, in particular in connection with hierarchical interpolation methods [12], RBFs with compact support have several advantages such as a sparse interpolation matrix $(K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^M$ and a sparse representation of the interpolating function $\sum a_j K(\mathbf{x}, \mathbf{x}_j)$ at $\mathbf{x} \in \mathbb{R}^d$.
RBFs with compact support have not been used for constructing SVMs up to now. One reason for this maybe that any positive definite RBF with compact support has to be designed in dependence on the space dimension d, i.e., there does not exist an universal RBF like the Gaussian which is compactly supported, positive definite and smooth for all space dimensions.

RBFs with compact support are only suited for relatively small space dimensions. Using our multilevel concentration approach the training data are at most of length $d = \log_2 N$. In our applications we have $d = 8$. For spaces of such a low dimension the following RBFs proposed by Wendland [54] can easily be calculated and evaluated.
For $\alpha \in \mathbb{R}$ and $m \in \mathbb{N}_0$ we define $(\cdot)_m$ and $[\cdot]_m$ by

$$(\alpha)_m := \frac{\Gamma(\alpha + m)}{\Gamma(\alpha)}, \quad [\alpha]_m := (\alpha - m + 1)_m,$$

where Γ denotes the Gamma function. Let

$$w_m(x) := (1 - x)_+^m.$$

Then Wendland's RBFs have the representation

$$k_{m,n}(x) = \sum_{i=0}^{n} \beta_{i,n}^{(m)} x^i w_{m+2n-i}(x) \quad (n \in \mathbb{N}),$$

where the coefficients satisfy the recursion

$$\begin{aligned} \beta_{0,0}^{(m)} &= 1, \\ \beta_{t,n+1}^{(m)} &= \sum_{i=t-1}^{n} \beta_{i,n}^{(m)} \frac{[i+1]_{i-t+1}}{(m+2n-i+1)_{i-t+2}} \quad (0 \le t \le n+1), \end{aligned}$$

if the term for $i = -1$ for $t = 0$ is ignored.
Let $\lfloor x \rfloor := \max\{i \in \mathbb{Z} : i \le x\}$. For $m = \lfloor d/2 \rfloor + n + 1$, Wendland has proved that $k_{m,n}$ defines a positive definite function in $C^{2n}(\mathbb{R}^d)$. Further, $k_{m,n}$ is of the form

$$k_{m,n}(x) = \begin{cases} p(x) & 0 \le x \le 1, \\ 0 & x > 1, \end{cases}$$

with a polynomial p of degree $\lfloor d/2 \rfloor + 3n + 1$. There does not exist a positive definite RBF in $C^{2n}(\mathbb{R}^d)$ of the above form with a polynomial p of lower degree.
In the remainder of this paper, we will deal (up to multiplications with constants) with the RBFs

$$\begin{aligned} k_{6,1}(x) &= (7x + 1)(1 - x)_+^7 \in C^2(\mathbb{R}^8) \\ k_{7,2}(x) &= 1/3 \left(80x^2 + 27x + 3\right)(1 - x)_+^9 \in C^4(\mathbb{R}^8). \end{aligned}$$

and their dilations $k_{m,n}(\cdot/s)$ $(s \in \mathbb{R}_+)$.

6 Classification of Endocardial Electrograms

Sudden cardiac death is a major public health concern worldwide. According to American estimates, sudden cardiac death claims more than 350.000 lives in the USA every year, 80% up to 90% being due to *ventricular tachycardia* (VT) [13, 22], i.e., a fast disorder of the heart beat which stems from the major heart chamber, the ventricle. The implantable cardioverter–defibrillator (ICD) is an automated antitachycardia device and accepted to be the most effective therapy for preventing sudden cardiac death due to VTs [43]. The ICD is a

permanently implanted device which continually monitors the electrical activity of the heart by an endocardial electrogram (EE), i.e., a bioelectric signal from the inner heart is analyzed. Usually, the information of the EE utilized by an ICD is the heart rate. However, the rate is of limited reliability in some clinical situation, e.g., excitement or physical exertion where the (physiological) sinus rhythm (SR) may have an abnormal fast rate which can exceed a predefined threshold such that a malignant rhythm is detected. Although additional detection enhancements are used in third generation ICD–systems, inappropriate ICD therapy occur in up to 13% of the patients who received such a device [53].

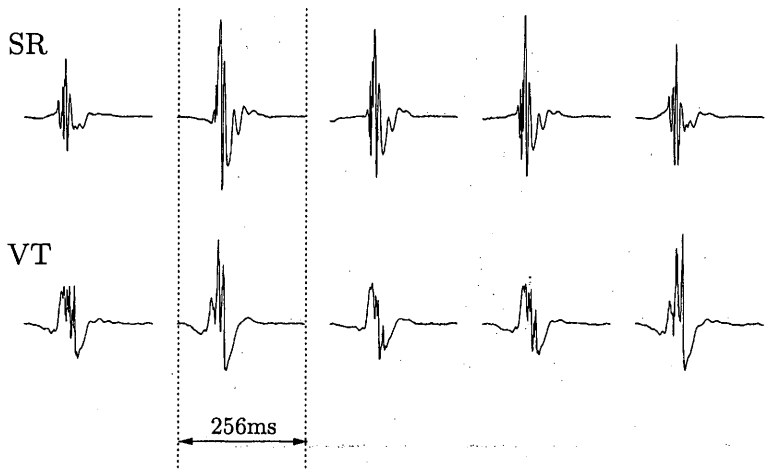

Figure 2: Consecutive beats of SR and VT within a time–frame of 256ms.

Here morphological dissimilarities in the EE of individual beats, due to different activation patterns, can be used for discriminating the physiological from the pathological rhythm since they are rate–independent. Such morphological methods are patient dependent and there is an ongoing interest in an efficient evaluation of morphological criteria [37]. Template matching methods are well known approaches. The *correlation waveform analysis* (CWA) with *best fit alignment* [44] offers an excellent performance and is widely accepted [39]. We will use this method as reference tool for performance comparisons to our algorithm.

Pattern recognition schemes for EEs have to tackle the following problems: an excellent generalization performance although the training data is very sparse and a highly efficient implementation for classifying the waveforms in view of the limited energy resources of an implantable device. Furthermore, in view of the current interest in EE compression algorithms for ICDs [4], the use of perfect reconstruction filter banks is also desirable for EE analysis since coding conditions can be added without difficulty.

The signals analyzed in this section were obtained during electrophysiological examinations at the University Hospital of Homburg, Germany. Bipolar EEs were obtained from the apex of the right ventricle in 10 patients with inducible monomorphic ventricular tachycardia (VT). The EEs were amplified, filtered (10–500Hz), and digitized (2kHz, 12 bit resolution). Data segments of 10s duration were recorded during SR and VT. Consecutive beats were selected as morphological patterns of SR (240 beats) and VT (240 beats) within a time–frame of 256ms and normalized. Consequently, we have to deal with waveforms $\mathbf{x} \in \Omega_{0,0} \subset \mathbb{R}^N$, where $N = 512$ and with a decomposition depth of the wavelet tree of $J = 8$. The set of beats was controlled by an expert to exclude ectopic beats and artifacts. Since false classifications can be of fatal consequences, we allow no errors at all and work with $\lambda = \infty$ in (9) to implement

a hard margin SVM. For our numerical experiments we use $p = 1$ in (29). The discrete parameter space was constructed with $T = 30$ and $L = 2$. This setting was also used in [41], where one of the authors has successfully classified antegrade and retrograde atrial activations patters by neural networks. We label examples of SR by 1 and of VT by -1. Figure 2 shows five original beats of SR and VT within their time–frame.

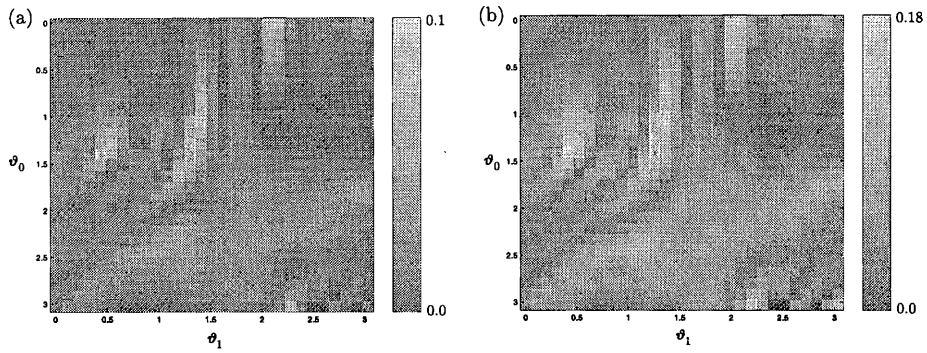

Figure 3: Distribution of the difference $\left|\left|\boldsymbol{\xi}_+(\boldsymbol{\vartheta}) - \boldsymbol{\xi}_-(\boldsymbol{\vartheta})\right|\right|_2$ (a) and of the corresponding margin $\gamma(\boldsymbol{\vartheta})$ of the SVM (b) in dependence on $\boldsymbol{\vartheta}$.

Figure 3(a) presents the distribution of the differences $\left|\left|\boldsymbol{\xi}_+(\boldsymbol{\vartheta}) - \boldsymbol{\xi}_-(\boldsymbol{\vartheta})\right|\right|_2$ in $\mathcal{P}_T^L$ for an individual patient ($|M_+| = |M_-| = 1$). In Figure 3(b) we show the corresponding distribution of the margin $\gamma(\boldsymbol{\vartheta})$ in $\mathcal{P}_T^L$ for a SVM with Wendland's function $k_{7,2}(\cdot/5)$. As noticeable, the margin is directly related to $\left|\left|\boldsymbol{\xi}_+(\boldsymbol{\vartheta}) - \boldsymbol{\xi}_-(\boldsymbol{\vartheta})\right|\right|_2$. The angles $\hat{\boldsymbol{\vartheta}}$ provide a multilevel concentration that results in the largest margin of the SVM classifier.
In order to illustrate our adaptation strategy, we consider for a moment only the multilevel concentrations on the levels $j = 4$ and $j = 5$, i.e., $\xi_{\mathbf{x}}(\boldsymbol{\vartheta}) = (||\mathbf{d}_4^{\boldsymbol{\vartheta}}||_{\ell^1}, ||\mathbf{d}_5^{\boldsymbol{\vartheta}}||_{\ell^1})$.

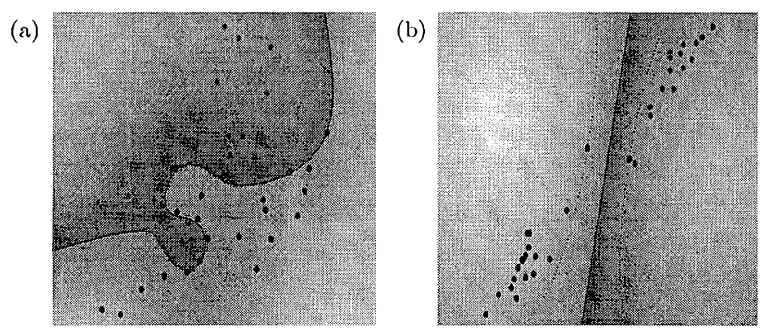

Figure 4: The decision curve $f(\boldsymbol{\xi}) = 0$ for SVM applied to the non–adapted multilevel concentrations with $\boldsymbol{\vartheta}_D$ (a) and to the adapted multilevel concentrations in $\mathbb{R}^2$. The darker points denote SR.

This allows us to visualize the 'decision curve' $f(\boldsymbol{\xi}) = 0$ in $\mathbb{R}^2$. We have used a number of $M = 36$ training examples for this experiment with $|M_+| = |M_-| = 18$. For the SVM we have employed a Gaussian kernel with scaling factor $s = 1$. Figure 4(a) shows the decision curve and the 36 training patterns when utilizing the lattice angles $\boldsymbol{\vartheta}_D = (1.4653, 0.49984)$ which correspond to the Daubechies wavelet D_3 with three vanishing moments. We use the well known Daubechies wavelet as representative for a non–adapted wavelet. In Figure 4(b)

we have taken the same settings but applied the optimal angles $\hat{\vartheta}$. Clearly, we expect that the SVM classifier performs more reliable for the adapted approach. Here the examples of the distinct classes laying far apart.
In the following we provide some assessments of the classification performance of our method using the whole data set of all ten patients. Here two decomposition levels as in the previous example are not sufficient. To rule out that filtering effects may lead to a loss of information we consider the full wavelet decomposition tree, i.e., $d = 8$. Of course, a couple of other data dependent strategies to choose $j_1, \ldots, j_d$ in (29) is possible. But, instead of engineering the best possible classification scheme for SR and VT, we are mainly interested in a comparison of our method to SVMs applied to the original time–domain waveforms and to SVMs applied to non–adapted multilevel concentrations.

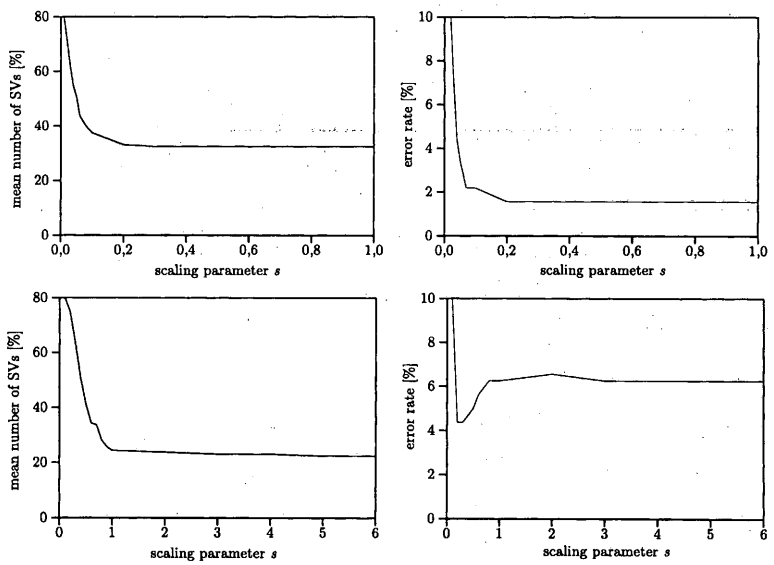

Figure 5: The mean number of SVs and the error rate of a SVM applied to the original waveforms with Gaussian Kernel (top) and to the non–adapted multilevel concentrations (D_3) with Wendland's function $k_{7,2}$ (bottom).

We separate the total data set of 480 beats in a training set of 160 beats (SR: 80, VT: 80) and a test set of 320 beats (SR: 160, VT: 160). Thus, for each patient we have a training set of $|M_-| = |M_-| = 8$ waveforms. The remaining set of 32 beats forms an independent test set for the individual patient.
In the Figures 5 and 6, we have plotted the mean number of SVs (of the ten patients) and the error rate in dependence on the scaling factor s for different RBFs. The mean number of SVs is given in percent and 100% means that all of the 16 training patters are SVs for all patients. The error rate [%] is determined for all patients in common, that is, the ratio of false classifications on the whole test set to the total number of 320 examples within this set. A comparison of Figure 5 with Figure 6 with respect to the error rate shows that our adapted algorithm with both the Gaussian and Wendland's function is superior to the original SVM and to the SVM on non–adapted multilevel concentrations. Note that by Remark 2 the number of support vectors can be considered as indicator for the generalization of the SVM. The number of support vectors of our new algorithm is approximately half as many as that of the original SVM. However, the case of SVMs on non–adapted multilevel concentrations

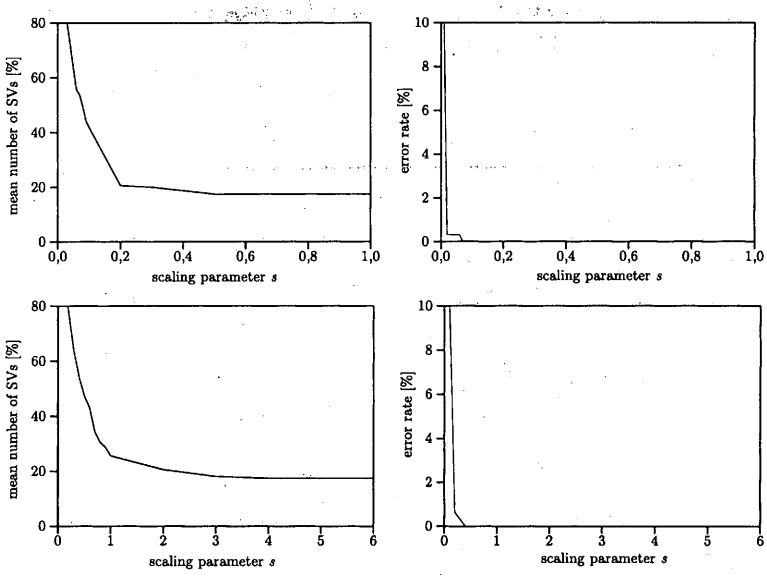

Figure 6: The mean number of SVs and the error rate of a SVM applied to the adapted multilevel concentrations with Gaussian Kernel (top) and to the adapted multilevel concentrations with Wendland's function $k_{7,2}$(*bottom*).

seems to be not in a good agreement with Remark 2. Here we have a smaller number of SVs than for SVMs on the original waveforms but a worse performance on the test set.
In Figure 7(a1) and (b1) we show the evaluations of f for SVMs with Gaussian kernel and scaling factor $s = 0.8$ applied to the original waveforms and for SVMs with $k_{7,2}$ and scaling factor $s = 5$ applied to the adapted multilevel concentrations. The corresponding number of SVs for the individual patients is given in Figure 7(a2) and (b2), respectively. We see that SVMs on adapted multilevel concentrations perform better and more reliable than SVMs on the time–domain waveforms. It is noticeable that the number of SVs found is less for our hybrid approach.
For the purpose of a performance comparison we shortly introduce the CWA. Let $\mathbf{x}^+ \in \Omega_{0,0}$ serve as template of SR which is obtained by averaging all the training waveforms of SR, i.e., $\mathbf{x}^+ = |M_+|^{-1} \sum_{i \in M_+} \mathbf{x}_i$. Let further $\mathbf{x} \in \Omega_{0,0}$ an arbitrary EE waveform to be classified. Then the CWA between the template and the waveform is based on the correlation coefficient $\rho \in [-1, 1]$ defined by

$$\rho := \frac{\sum_{i=1}^{N} (x_i^+ - \overline{x}_+)(x_i - \overline{x})}{\sqrt{\sum_{i=1}^{N} (x_i^+ - \overline{x}_+)^2} \sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2}}, \tag{34}$$

where $\overline{x}_+ := N^{-1} \sum_{i=1}^{N} x_i^+$ and $\overline{x} := N^{-1} \sum_{i=1}^{N} x_i$ For $\rho = 1$ we have a perfect match of the waveform $\mathbf{x}$ and the template. The decision of the CWA, i.e., the classification of SR and VT, is based on an appropriate threshold τ with $\tau < 1$. Now $\rho > \tau$ denotes SR and $\rho \leq \tau$ denotes VT. The performance of the CWA heavily depends on the alignment of the time–frame in which an individual beat is selected. Therefore, we can further improve the results of the CWA by the *best fit alignment* strategy. Here the template $\mathbf{x}^+$ is shifted over a specific time–frame centered around the detection point of the beat, i.e., the point where an individual beat is separated from the EE, and (34) is calculated for each point $n \in \mathbb{Z}$ of this time–frame so that

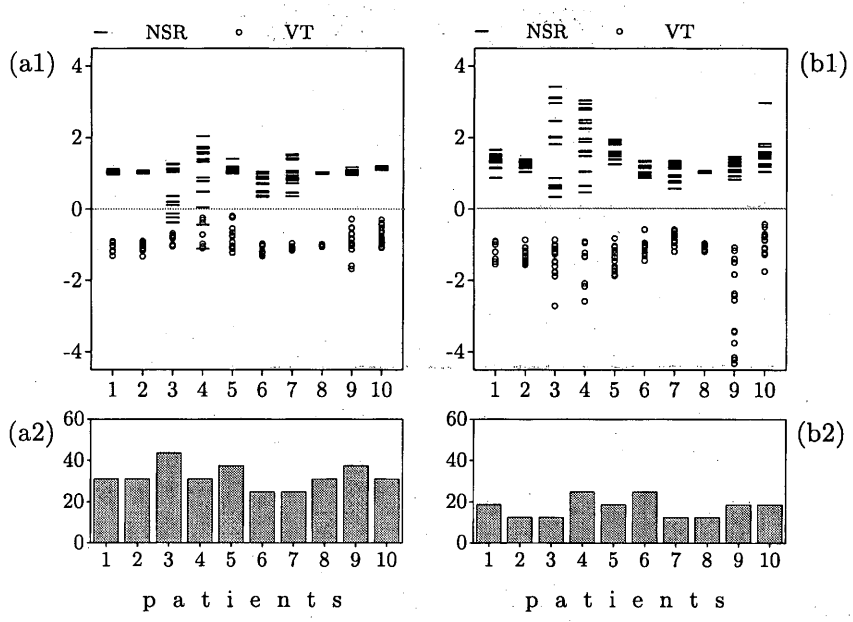

Figure 7: The evaluation of f for 10 patients by an SVM applied to the time–domain waveforms (a1) and to the adapted multilevel concentrations (b1). The corresponding number of SVs is given in (a2) and (b2), respectively.

$\rho[n]$ becomes a time dependent sequence. The decision is then based on $\eta = ||\boldsymbol{\rho}||_{\ell^\infty}$ using the same threshold criterion as before for ρ. This procedure is computationally demanding in general and not appropriate for efficient implementations. The support of the time–frame depends on the preprocessing of the data. In our experiments we have used 20ms. It is worth to emphasize that the best fit strategy has only a minor influence for our signals as the beats are already well aligned. However, in other settings it can be significant.

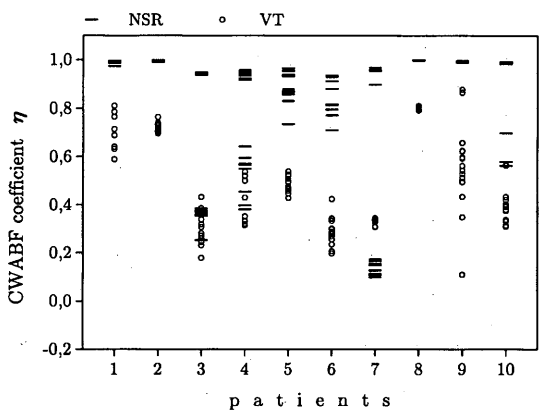

Figure 8: The results for the CWA with best fit alignment (CWABF) for 10 patients.

When averaging the 8 training waveforms of SR of the individual patient as template for the CWA, we obtain the results in the Figure 8. There is a significant overlap for a subset of patients and a threshold τ can heavily be defined for these patients. Comparing these results to those in Figure 7 we see that the CWA performs much worse than our hybrid scheme and also worse than SVMs applied to the time–domain waveforms.

7 Detection of Otoacoustic Emissions

As already mentioned, the beats analyzed in the previous section are well aligned within their time–frame. Thus, the analysis does not strongly suffer from shifts in time so that shift–invariance is of minor importance. Here computationally very efficient but time–variant orthonormal wavelet decompositions perform well. However, as in our next classification problem, there is often a need for shift–invariance in waveform recognition. In the following we use the shift–invariant frame decompositions introduced at the end of Section 3 instead of orthogonal wavelet decompositions.

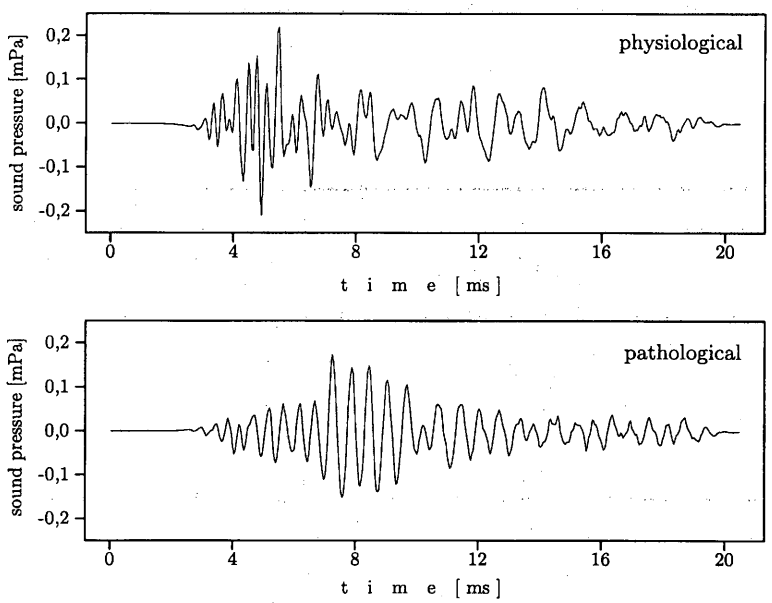

Figure 9: Two exemplary signals from the physiological and pathological group.

Otoacoustic emissions are sounds generated by the cochlea that can be recorded in the external ear canal, see [18, 27]. In the following, we will deal with the *transient evoked otoacoustic emission* (TEOAE) that is, in short, a brief pressure wave that emanates from the ear in response to an acoustic stimulus. These responses are typically of a low intensity or absent in individuals with mild or greater hearing loss. Consequently, the analysis of TEOAEs seems to be promising as noninvasive method for detecting hearing loss in patients. Such an objective analysis requires no patient interaction and is especially of interest in infants. The detection of otoacoustic emissions represents a challenge for automated schemes. They are of a very low intensity and hard to separate from the signal background in current measurement set ups. Up to now, the classification of individuals with and without hearing losses is to large extend based on expert knowledge, see [15] for a description of the parameters used by experts. The analysis and characterization of TEOAEs is an active and ongoing field of research. For approaches in the time–frequency domain, see, for instance, [45] and the references therein.

The signals analyzed in this section were obtained at the University Hospital of Homburg, Germany. The signals were recorded in a sound–proof cabin using a probe inserted into the outer ear canal. The probe contains a transmitter which delivers the acoustic stimulus and a miniaturized microphone. The signals were amplified, filtered (300Hz – 10kHz), digitized (25kHz, 12 bit resolution) and averaged to enhance the signal noise ratio by the standard

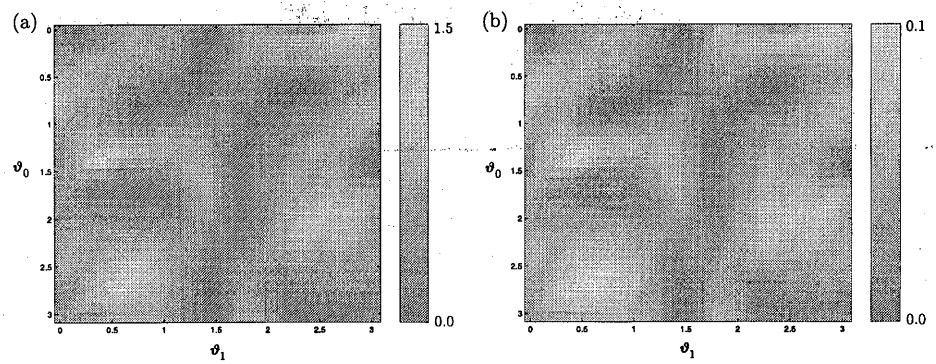

Figure 10: Distribution of the difference $\left|\left|\boldsymbol{\xi}_+(\boldsymbol{\vartheta}) - \boldsymbol{\xi}_-(\boldsymbol{\vartheta})\right|\right|_2$ (a) and of the corresponding margin $\gamma(\boldsymbol{\vartheta})$ of the SVM (b) in dependence on $\boldsymbol{\vartheta}$.

technique of the ILO88 Systems (Otodynamics, Ltd.). With this preprocessing, signals of approximately 20ms (again represented by 512 samples) duration were normalized (by their energy) and stored for subsequent analysis. The individuals were binary classified in a pathological and a physiological group. The pathological group is labeled with -1 and consists of individuals with a hearing loss at higher frequencies or with a broad band hearing loss. All individuals in this group have a hearing loss of more than 20dB above 3kHz. The physiological group is labeled with 1 and consists of individuals with a hearing loss of less than 20dB in range from 0kHz to 10kHz. Figure 9 shows an exemplary signal of the physiological and the pathological group, respectively.
We have a total of 68 individuals (and signals) in the physiological group. The pathological group consists of 55 individuals. We use twenty signals of each group as training set, i.e., $|M_+| = |M_-| = 20$. The remaining signals form an independent test set for the respective group.

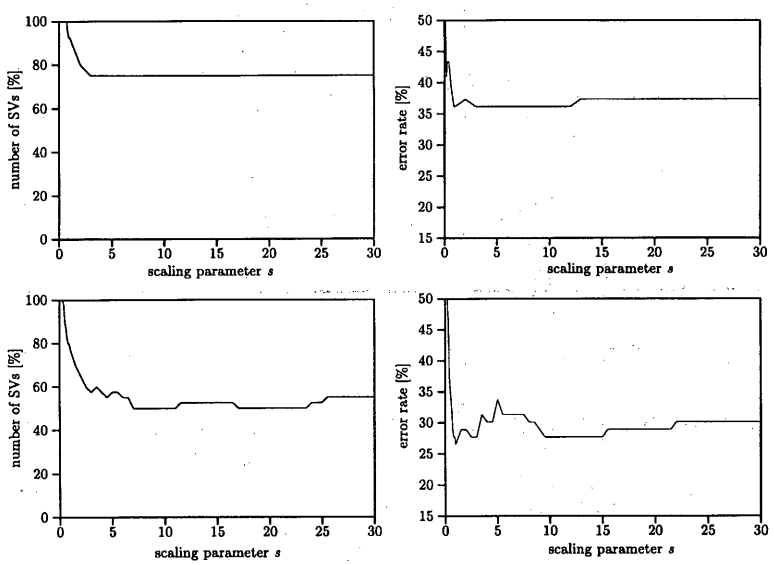

Figure 11: The mean number of SVs and the error rate of a SVM applied to the original waveforms with Gaussian Kernel (top) and to the adapted orthonormal multilevel concentrations also with a Gaussian kernel (bottom).

If present, the occurrence of TEOAEs in our signals varies over time for the individual subjects. Consequently, there is a need for a recognition scheme which is not based on locality. In other words, it must incorporate the fact the information can appear anywhere in the signal and must thus be shift–invariant.
The morphology of physiological and pathological waveforms may overlap, in particular for physiological group and individuals which only suffer from a hearing loss at higher frequencies. Therefore, the appropriate choice of the regularization parameter λ in (16) may improve the results. However, we use again $\lambda = \infty$ for implementing a hard margin SVM since it is easier to interpret and we are mainly interested in comparing SVM on the original time–domain waveform and on the shift–invariant adapted multilevel concentrations.
Figure 10 illustrates the difference $\left|\left|\boldsymbol{\xi}_+(\vartheta) - \boldsymbol{\xi}_-(\vartheta)\right|\right|_2$ for $|M_+| = |M_-| = 1$ as example. The corresponding margin $\gamma(\vartheta)$ in $\mathcal{P}_T^L$ is also shown, where we have used the Wendland function $k_{7,2}$ for the SVM. As in the previous section, only a few wavelets of the parameter space result in a large margin of the SVM classifier.

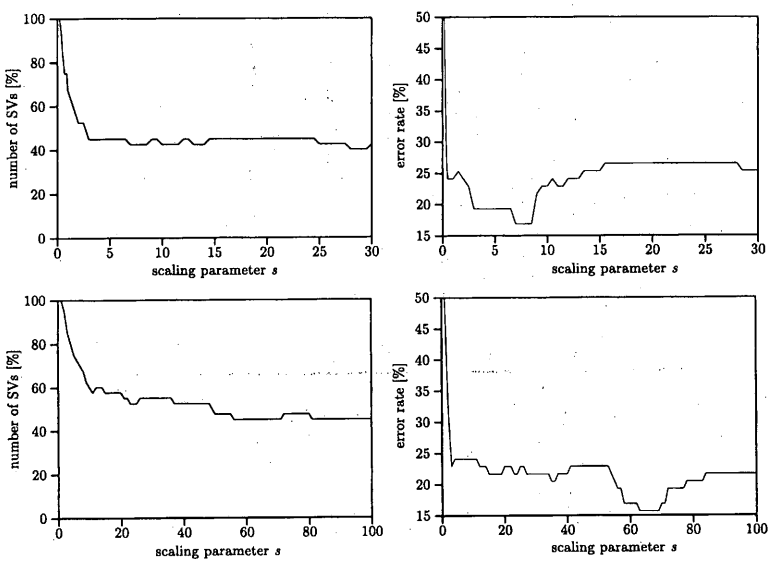

Figure 12: The mean number of SVs and the error rate of a SVM applied to to the adapted shift–invariant multilevel concentrations with the Gaussian kernel (top) and to the adapted shift–invariant multilevel concentrations with $k_{7,2}$ (bottom).

Figure 11 contain the error rate and the number of SVs in dependence on the scaling parameter s for a SVM applied to the original waveform and for our scheme, but based on orthonormal (not shift–invariant) decompositions. It is obvious that no reliable hypothesis about the underlying mapping can be deducted form the training set in the first case so that the original SVM classifier allows no satisfactory discrimination between the physiological and the pathological group. In view of the fact that the SVM is a *state of the art classifier*, this indicates how difficult it is to deduce a reliable hypothesis. Our scheme with orthonormal decompositions performs better. However, when using the shift–invariant tight frames in our scheme, the results in Figure 12 are obtained for $k_{7,2}$ and a Gaussian kernel. Here our hybrid SVM classifier performs significantly better than SVMs on the original waveforms and on the orthonormal multilevel concentrations. The number of SVs is much smaller than that of the original SVM. Clearly, with further heuristics, e.g., by discarding information or by

regularization, better results seem to be achievable.

In our applications RBFs with compact support have performed similar to Gaussian RBFs. The best performance of RBFs with compact support was achieved for large values of the scaling parameter s so that $K(\mathbf{x}_i, \mathbf{x}_j) > 0$ for all $i, j \in \{1, \ldots, M\}$. Thus, we have achieved no further advantages from the compact support of these functions.

8 Conclusions

We have presented a new method for improving the performance of SVMs. For this, we have merged ideas which have recently been developed in signal processing and machine learning to obtain a hybrid scheme based on wavelet decompositions and SVMs for waveform classification.
We have illustrated the performance of our scheme against the background of current concerns in medical diagnostics. Our hybrid scheme classified all EEs of SR and VT of the given test set correctly, a result that was neither achievable by a SVM on the original waveforms nor by the well accepted CWA with best fit alignment. Based on the very efficient lattice structure implementation, our scheme meets also the low power requirements of ICDs.
For the analysis of TEOAEs we have employed tight frames to obtain shift–invariance of our scheme. Here we archived a significant improvement of the classification performance for detecting the hearing loss in individuals compared to SVMs on the original waveforms.
We have used RBFs with compact support and have shown that these functions perform similar to the well accepted Gaussian kernels. However, further research is necessary to exploit the full power of the compact support of these functions.
We conclude that the performance of SVMs can significantly be improved by signal–adapted wavelet decompositions since it allows the inclusion of prior knowledge such as local instabilities in time and shift–invariance. The direct relation of the distances in the input space and the feature space induced by the RBF allows an optimization of the signal representation before training the SVM.
For our applications, the presented method is powerful and very promising. However, especially in the case of TEOAE analysis further investigations are needed on larger data sets. Here is still room for improvements, i.e., by the optimal choice of the decomposition levels, the straightforward incorporation of wavelet packets and the optimal choice of the regularization constant λ.

Acknowledgements: This work was done in collaboration with the University Hospital Homburg, Germany. In particular, the authors like to thank Dipl.–Ing. D. Benyoucef, OA Dr. W. Delb, and OA PD Dr. J. Jung for providing the medical data and for critically reading the manuscript.

References

[1] H. BÖLCSKEI, F. HLAWATSCH, AND H. G. FEICHTINGER, *Frame-theoretic analysis of oversampled filter banks*, IEEE Trans. Signal Processing, 46 (1998), pp. 3256–3268.

[2] C. BURGES, *A tutorial on support vector machines for pattern recognition*, Data Mining and Knowledge Discovery, 2 (1998), pp. 121–167.

[3] ——, *Geometry and invariance in kernel based methods*, in Advances in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 89–116.

[4] R. J. COGGINS AND M. A. JABRI, *A low-complexity intracardiac electrogram compression algorithm*, IEEE Trans. on Biomed. Eng., 46 (1999), pp. 83–91.

[5] R. R. COIFMAN AND D. DONOHO, *Translation invariant de-noising*, in Wavelet in Statistics, A. Antoniadis, ed., Lecture Notes, Springer Verlag, New York, 1995, pp. 125–150.

[6] N. CRISTIANINI AND J. SHAWE-TAYLOR, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.

[7] Z. CVETKOVIĆ AND M. VETTERLI, *Oversampled filter banks*, IEEE Trans. on Signal Processing, 46 (1998), pp. 1245–1255.

[8] I. DAUBECHIES, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[9] I. DAUBECHIES AND W. SWELDENS, *Factoring wavelet transforms into lifting steps*, J. Fourier Anal. Appl., 4 (1998), pp. 245–267.

[10] P. H. DELSARTE, B. MACQ, AND D. T. M. SLOCK, *Signal-adapted multiresolution transforms for image coding*, IEEE Trans. on Information Theory, 38 (1992), pp. 897–903.

[11] R. FLETCHER, *Practical Methods of Optimization*, John Wiley & Sons, New York, 1990.

[12] M. S. FLOATER AND A. ISKE, *Multistep scattered data interpolation using compactly supported radial basis functions*, J. of Computational and Applied Mathematics, 73 (1996), pp. 65–78.

[13] R. F. GILUM, *Sudden cardiac death in the united states*, Circulation, 79 (1989), pp. 756–776.

[14] L. HERMES AND J. M. BUHMANN, *Feature selection for support vector machines*, in Proc. International Conference on Pattern Recognition 2000, Barcelona, Spain, 2000.

[15] S. HOTH, *Relationship between parameters of evoked otoacoustic emissions and hearing loss*, Audiologische Akustik, 1 (1995), pp. 20–29.

[16] N. INTRATOR, Q. HUYNH, Y. WONG, AND B. H. K. LEE, *Wavelet feature extraction for discrimination tasks*, in Proceedings of the Canadian Workshop on Information Theory, Toronto, 1997, pp. 83–86.

[17] T. JOACHIMS, *Making large–scale support vector machine learning practical*, in Advances in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 169–184.

[18] D. T. KEMP, *Stimulated acoustic emissions from within the human auditory system*, J. Acoust. Soc. Am., 64 (1978), pp. 1386–1391.

[19] J. T. KIM, Y. H. LEE, T. ISSHIKI, AND H. KUNIEDA, *Scalable VLSI architectures for lattice-structure-based discrete wavelet transform*, IEEE Trans. on Circuits and Systems II, 45 (1998), pp. 1031–1043.

[20] G. S. KIMELDORF AND G. WAHBA, *Some results on tchebycheffian spline functions*, J. Anal. Applic., 33 (1971), pp. 82–95.

[21] N. G. KINGSBURY, *Complex wavelets for shift invariant analysis and filtering of signals*, Submitted to: J. of Applied Computation and Harmonic Analysis, (2000).

[22] R. J. MYERBURG, K. M. KESSLER, AND A. CASTELLANOS, *Sudden cardiac death, structure, function, and time–dependence of risk*, Circulation, 85 (1992), pp. I2–I10.

[23] M. OREN, C. PAPAGEORGIOU, P. SINHA, AND E. OSUNA, *Pedestrian detection using wavelet templates*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 193–199.

[24] C. PAPAGEORIOU, M. OREN, AND T. POGGIO, *A general framework for object detection*, in Proceedings of International Conference on Computer Vision (ICCV'98), Bombay, India, 1998, pp. 555–562.

[25] S. PITTNER AND S. V. KAMARTHI, *Feature extraction from wavelet coefficients for pattern recognition tasks*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 21 (1999), pp. 83–88.

[26] T. POGGIO AND T. VETTER, *Recognition and structure from one 2d model view: Observations on prototypes, object classes, and symmetries*, AI Memo 1347, Massachusetts Institute of Technology, Cambridge, MA, 1992.

[27] R. PROBST, B. L. LONSBURY-MARTIN, AND G. K. MARTIN, *A review of otoacoustic emissions*, J. Acoust. Soc. Am., 89 (1991), pp. 2027–2067.

[28] G. RUTLEDGE AND G. MCLEAN, *Comparison of several wavelet packet feature extraction algorithms*, Submitted to: IEEE Trans. on Pattern Analysis and Machine Intelligence, (2000).

[29] N. SAITO AND R. R. COIFMAN, *Local discriminant bases*, in Wavelet Applications in Signal and Image Processing II, Proc. SPIE 2303, A. F. Laine and M. A. Unser, eds., July 1994, pp. 2–14.

[30] R. SCHABACK, *Creating surfaces from scattered data using radial basis functions*, in Mathematical Methods in Computer Aided Geometric Design III, M. Dæhlen, T. Lyche, and L. L. Schumaker, eds., Vanderbilt Univ. Press, 1995, pp. 477–496.

[31] ——, *Reconstruction of multivariate functions from scattered data*, Monograph, Institute for Numerical and Applied Mathematics, University of Göttingen, (1997).

[32] B. SCHÖLKOPF, C. BURGES, AND V. VAPNIK, *Incorporating invariances in support vector learning machines*, in Artificial Neural Networks, ICANN'96, Springer Lecture Notes in Computer Science, M. I. Jordan, M. J. Kearns, and S. A. Solla, eds., vol. 1112, Berlin, Germany, 1996, pp. 47–52.

[33] B. SCHÖLKOPF, S. MIKA, C. BURGES, P. KNIRSCH, K.-R. MÜLLER, G. RÄTSCH, AND A. J. SMOLA, *Input space vs. feature space in kernel-based methods*, IEEE Trans. on Neural Networks, 10 (1999), pp. 1000–1017.

[34] B. SCHÖLKOPF, P. Y. SIMARD, A. J. SMOLA, AND V. VAPNIK, *Prior knowledge in support vector kernels*, in Advances in Neural information processings systems, M. I. Jordan, M. J. Kearns, and S. A. Solla, eds., vol. 10, Cambridge, MA, 1998, pp. 640–646. MIT Press.

[35] B. SCHÖLKOPF, K. SUNG, C. BURGES, F. GIROSI, P. NIYOGI, T. POGGIO, AND V. VAPNIK, *Comparing support vector machines with gaussian kernels to radial basis function classifiers*, IEEE Trans. Signal Processing, 45 (1997), pp. 2758–2765.

[36] M. SHENSA, *The discrete wavelet transform: Wedding the à trous and mallat algorithms*, IEEE Trans. Signal Processing, 40 (1992), pp. 2464–2482.

[37] M. M. SHIEH, L. CLEM, L. MALDEN, P. A., AND E. FAIN, *Improved supraventricular and ventricular tachycardia discrimination using electrogram morphology in an implantable cardioverter defibrillator*, G. Ital. Cardiol., 28 (supplement 1) (1998), pp. 245–251.

[38] I. STEINWART, *On the influence of the kernel on the generalization ability of support vector machines*, Technical Report 01-01, 2001, Institute of Mathematics, University of Jena, 2001.

[39] S. A. STEVENSON, J. M. JENKINS, AND L. A. DICARLO, *Analysis of the intraventricular electrogram for differentiation of distinct monomorphic ventricular arrhythmias*, PACE, 20 (1997), pp. 2730–2738.

[40] G. STRANG AND T. NGUYEN, *Wavelets and Filter Banks*, Wellesley–Cambridge Press, Wellesley, MA, 1996.

[41] D. STRAUSS, J. JUNG, A. RIEDER, AND Y. MANOLI, *Classification of endocardial electrograms using adapted wavelet packets and neural networks*, Ann. Biomed. Eng. in press.

[42] D. STRAUSS, T. SINNWELL, A. RIEDER, Y. MANOLI, AND J. JUNG, *A promising approach to morphological endocardial signal discriminations: Adapted multiresolution signal decompositions*, Applied Sig. Process., 6 (1999), pp. 182–193.

[43] THE ANTIARRHYTHMICS VERSUS IMPLANTABLE DEFIBRILLATORS (AVID) INVESTIGATORS, *A comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from near fatal ventricular arrhythmias*, N. Engl. J. Med., 337 (1997), pp. 1576–1583.

[44] R. D. THRONE, J. M. JENKINS, S. A. WINSTON, C. J. FINELLI, AND L. A. DICARLO, *Discrimination of retrograde from antegrade atrial activation using intracardiac electrogram waveform analysis*, PACE, 12 (1989), pp. 1622–1630.

[45] G. TOGNOLA, F. GRANDORI, AND P. RAVAZZANI, *Wavelet analysis of click-evoked otoacoustic emissions*, IEEE Trans. on Biomedical Engineering, 45 (1998), pp. 686–697.

[46] L. J. TREJO AND M. J. SHENSA, *Feature extraction of event-related potentials using wavelets: An application to human performance monitoring*, Brain and Language, 66 (1999), pp. 89–107.

[47] P. P. VAIDYANATHAN, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[48] P. P. VAIDYANATHAN AND P. Q. HOANG, *Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction qmf filter banks*, IEEE Trans. Acoust. Speech, and Signal Proc., 36 (1988), pp. 81–94.

[49] V. VAPNIK, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[50] M. VETTERLI AND J. KOVAČEVIĆ, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[51] G. WAHBA, *Support vector machines, reproducing kernel hilbert spaces and the randomized GACV*, in Advanced in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 293–306.

[52] Z. WANG, Z. HE, AND J. D. Z. CHEN, *Filter banks and neural networks-based feature extraction and automatic classification of electrogastrogam*, Ann. Biomed. Eng., 27 (1999), pp. 88–95.

[53] M. WEBER, D. BÖCKER, D. BANSCH, J. BRUNN, M. CASTRUCCI, R. GRADAUS, G. BREITHARDT, AND M. BLOCK, *Efficacy and safety of the initial use of stability and onset criteria in implantable cardioverter defibrillator*, J. Cardiovasc. Electrophysiol., 10 (1999), pp. 145–153.

[54] H. WENDLAND, *Piecewise polynomial, positive definite functions and compactly supported functions radial basis functions of minimal degree*, Adv. in Comp. Math., 4 (1995), pp. 389–396.

[55] J. WESTON, S. MUKHERJEE, O. CHAPELLE, M. PONTIL, T. POGGIO, AND V. VAPNIK, *Feature selection for SVMs*, in Advances in Neural Information Processing Systems, vol. 13, 2000.

[56] H. ZOU AND A. H. TEWFIK, *Parameterization of compactly supported wavelets orthonormal wavelets*, IEEE Trans. on Signal Processing, 41 (1993), pp. 1428–1431. See also: S. H. Wang, A. H. Tewfik, and H. Zou. Correction to *Parameterization of compactly supported wavelets orthonormal wavelets*. IEEE Trans. on Signal Processing, 42 (1994), pp. 208–109.