Classification of Sleep Apnea Events by Means of Radial Basis Function Networks

Thomas Zemen, Markus Clabian, Helmut Pfützner

Institute of Fundamentals and Theory of Electrical Engineering, Bioelectricity & Magnetism Lab.
Vienna University of Technology, Gußhausstraße 27/351, A-1040 Vienna, Austria
http://www.gte.tuwien.ac.at/

Abstract - **Sleep apneas, cessations of breathing during sleep, which lead to shortened life expectation (adults) and indicate SIDS-risks (premature born babies) must be diagnosed reliably. Therefore a monitoring system is established and attempts are made for automatic scoring to reduce clinical efforts. This paper describes the detection and classification of sleep apnea events by means of radial basis function networks (RBFN). A monitoring device, based on electric field plethysmography, which detects respiratory and cardiac activity is presented. In addition, heart rate and blood oxygen saturation are recorded. Pre-processing is performed by several feature extraction algorithms including time dependent Fourier transform, cepstrum transform, linear predictive coding, linear predictive coding cepstrum (LPCC) and wavelet transform. RBFNs are trained with learning vector quantization (LVQ) and self organizing map (SOM). To achieve high classification rates, optimization of number and spread of neurons are performed. Best results with classification rates of 64% $\pm$ 3.4% (adults) and 62.6% $\pm$ 6.8% (babies) is obtained using LPCC and data reduction through principle components analysis (PCA). The net consisted of 20 hidden neurons and is trained with LVQ. All results are derived using 10-fold cross validation.**

1 Introduction

Sleep disorders are more common than expected (concerning about 5% of the total population) and must be diagnosed reliably to avoid consequential damages. Therefore monitoring systems are established, and attempts are made for automatic scoring of apnea events to reduce clinical efforts as well as the number of medical staff.

Apnea events are defined as a cessation of normal breathing and can be broken down to a high number of classes. In this paper, classification is restricted to the following four major event types:

- Central apnea (CA) - due to missing nervous stimulus,
- obstructive apnea (OA) - due to obstruction of the upper airway,
- sigh (SI) - a deep inspiration followed by an apnea (only considered in the case of babies), and
- normal breathing (NB).

As a basis for the training of neural networks (NNs) as a diagnostic tool, the above classifications were performed by medical staff via a polysomnography system which is the standard of clinical practice. First corresponding applications of NNs have been reported in [1][2][3]. The application of NNs has proved to be an effective approach, however, in most cases, the classification rates did not exceed 60% [2][3] which could be explained by several specific problems, such as nonstationary character of physiological signals and extensive individual differences of signal patterns.

This paper presents first attempts to detect and classify apnea events by means of radial basis function networks (RBFNs).

Monitoring System - A portable device is used to monitor the respiratory and cardiac activity via electric field plethysmography (EFPG) [4], the blood oxygen saturation and the heart rate.

The EFPG is a well known method to detect cardiac and respiratory activity. An electric field is applied to the thorax, its configuration is changing with varying blood and air content. As a specific advantage, only four electrodes are needed to detect a signal which contains information on both the cardiac and the respiratory component.

The basic harmonic of the respiratory component shows a frequency range of $f_\mathrm{R} = 0.15\,\mathrm{Hz} \ldots 1.3\,\mathrm{Hz}$ ($9\,\mathrm{min}^{-1} \ldots 78\,\mathrm{min}^{-1}$) during normal breathing and $0\,\mathrm{Hz}$ for apneas. The cardiac activity covers a range of $f_\mathrm{C} = 0.65\,\mathrm{Hz} \ldots 2.8\,\mathrm{Hz}$ ($45\,\mathrm{min}^{-1} \ldots 168\,\mathrm{min}^{-1}$), the amplitude being lower by about 15 dB compared to the respiratory component. Apart from this frequency overlap, the instantaneous ratio of cardiac and respiratory frequency $f_\mathrm{C}/f_\mathrm{R}$ ranges between 2.5 and 5.8. With the above preconditions it proves to be sufficient to sample the EFPG signal s with $\nu_\mathrm{s} = 20\,\mathrm{Hz}$. A closer discussion of this topics can be found in [5].

As already mentioned, monitoring included also the following two parameters for more reliable classifications:

- blood oxygen saturation c_ox - determined by an optical sensor, and
- heart rate f_C - determined by means of a non-standardized electro cardio gramm (ECG), using the EFPG electrodes in order to keep the number of electrodes to a minimum.

An example of a CA with the described signals is given in figure 1.

Data Material - Two groups of patients are investigated:

- 50 adults, suffering from the sleep apnea syndrome (SAS) which prevents them from deep refreshing sleep, and
- 30 premature born babies exhibiting respiratory disturbances which are assumed to be closely connected to the sudden infant death syndrome (SIDS).

2 Methods

The signal processing structure used to classify sleep apnea is shown in figure 2. The pre-processing stage is followed by a principal component analysis (PCA) to reduce the dimensionality of each feature vector $\mathbf{c}_i$. The principal components (PCs) $\mathbf{v}_i$ are further standardized and combined to the input vector $\mathbf{x}$ of the radial basis function network (RBFN).

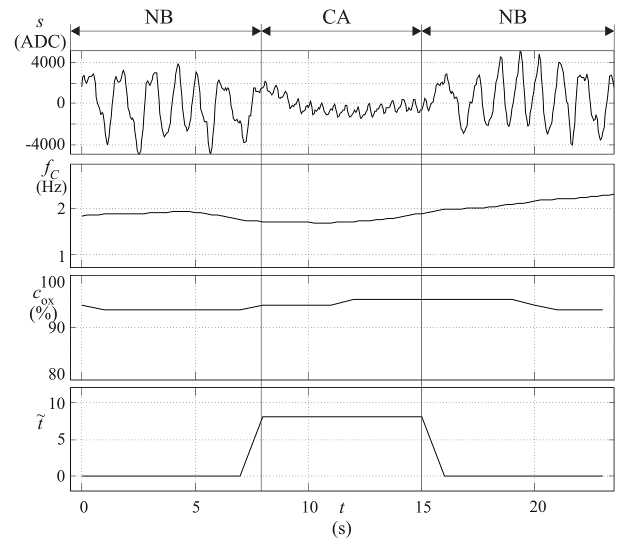

Figure 1: Registered signals during a central apnea (CA), its duration is marked. The blood oxygen saturation shows response with some delay. s EFPG signal, f_C heart rate, c_ox blood oxygen saturation, $\tilde{t}$ classification signal.

Several pre-processing algorithms, different RBFN structures and learning algorithms are evaluated to maximize the recognition rate R of the RBFN.

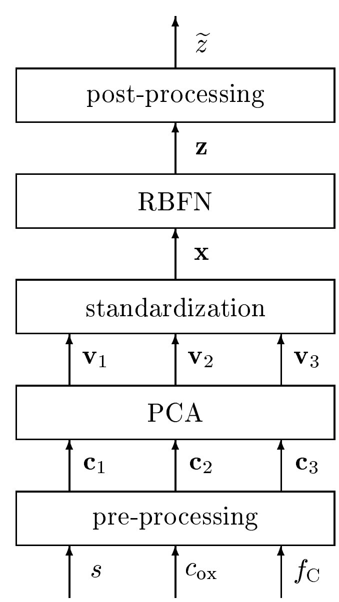

Figure 2: Signal processing structure. s EFPG signal, c_ox blood oxygen saturation, f_C heart rate, $\mathbf{c}$ feature vector, $\mathbf{v}$ principal component vector, $\mathbf{x}$ input vector, $\mathbf{z}$ output vector, $\tilde{z}$ output classification.

Pre-processing

The following pre-processing methods are applied to the EFPG signal s, which is windowed by a Hamming function

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n/L) & 0 \leq n \leq L, \\ 0 & \text{sonst} \end{cases} \tag{1}$$

with a length of six seconds (equivalent to $L = \nu_\mathrm{s} 6\,\mathrm{s} = 120$ samples), shifted by $T = 1\,\mathrm{s}$ ($\rho_\mathrm{s} = \nu_\mathrm{s} T = 20$ samples).

Fourier Transform - the EFPG signal is a mixture of two components caused by cardiac and respiratory activity. Both components are periodic with varying frequency and amplitude. The basic frequency of the cardiac component f_C is at least 2.5 higher than f_R. The time dependent Fourier transform (TDFT) is therefore expected to be a suitable pre-processing algorithm. It is calculated according to

$$S(r\rho_\mathrm{s}, k) = \sum_{m=0}^{L-1} s(n)\, w(m)\, e^{-j(2\pi/N)km} \tag{2}$$

with

$$n = r\rho_\mathrm{s} - L/2 + m \tag{3}$$

for $-\infty < r < \infty$, $0 \leq k \leq N-1$. r represents an integer time index with dimension unity [6][7]. The number of frequency components is chosen with $N = 64$. Only the first $N' = 33$ are used because they cover the frequency range of interest from 0 Hz to 5 Hz.

The vector representation of the feature vector is given by

$$\mathbf{c}_\mathrm{TDFT}(r) = 20 * \log \begin{pmatrix} S(r\rho_\mathrm{s}, 1) \\ S(r\rho_\mathrm{s}, 2) \\ \vdots \\ S(r\rho_\mathrm{s}, N') \end{pmatrix}. \tag{4}$$

Cepstrum Transform - (CT) calculated according to

$$\hat{s}(r\rho_\mathrm{s}, k) \simeq \frac{1}{L} \sum_{l=0}^{L-1} \hat{S}(r\rho_\mathrm{s}, l) e^{j(2\pi/L)l(r\rho_\mathrm{s} - L/2 + k)} \tag{5}$$

with

$$\hat{S}(r\rho_\mathrm{s}, k) = \log(S(r\rho_\mathrm{s}, k)) \tag{6}$$

allows deconvolution and frequency invariant filtering [7]. Despite this useful properties, it is by itself not well suited for pre-processing because of the high number of components of the resulting feature vector

$$\mathbf{c}_{CT}(r) = \begin{pmatrix} \hat{s}(r\rho_\mathrm{s}, 1) \\ \hat{s}(r\rho_\mathrm{s}, 2) \\ \vdots \\ \hat{s}(r\rho_\mathrm{s}, L) \end{pmatrix}, \tag{7}$$

that is equal to the window length L. But as shown later, the CT can be combined with linear predictive coding (LPC) to avoid this disadvantage.

Linear Predictive Coding - (LPC) models the signal as impulse response of an all-pole filter. Therefor the assumption

$$x(n) = \sum_{k=1}^{p} a_k x(n-k). \tag{8}$$

is made, that the current sample of the signal could be represented as a linear combination of p previous samples. The transfer function of an all-pole filter is given by

$$H(z) = \frac{b_0}{\sum_{k=0}^{p} a_k z^{p-k}}; \qquad a_0 = 1. \tag{9}$$

The characteristic polynom of (9) is equivalent to (8). The coefficients a_k are calculated by the Levinson-Durbin recursion [8][9][10], which minimizes a quadratic error function. Already a low number p of coefficients a_k approximate the strongest frequency components of the signal s appropriate (compared to TDFT). The feature vector is calculated according to

$$\mathbf{c}_\mathrm{LPC}(r) = \begin{pmatrix} a_1(r\rho_\mathrm{s}) \\ \vdots \\ a_p(r\rho_\mathrm{s}) \\ b_0(r\rho_\mathrm{s}) \end{pmatrix}. \tag{10}$$

Linear Predictive Coding Cepstrum - (LPCC) combines CT, which allows deconvolution and frequency invariant filtering, with LPC which reduces the dimension of the resulting feature vector. LPCC coefficients d_k can be recursively calculated from LPC coefficients a_k by

$$d_1 = -a_1, \tag{11}$$

$$d_n = -a_n - \frac{1}{n} \sum_{k=1}^{n-1} k d_k a_{n-k} \quad n = 2 \ldots p \tag{12}$$

(see [11]). The feature vector is defined as

$$\mathbf{c}_{\mathrm{LPCC}}(r) = \begin{pmatrix} d_1(r\rho_\mathrm{s}) \\ \vdots \\ d_p(r\rho_\mathrm{s}) \\ b_0(r\rho_\mathrm{s}) \end{pmatrix} = \mathrm{LPCC}\{s(r\rho_\mathrm{s}), p\}. \tag{13}$$

Wavelet Transform - (WT) [12] constructs a given signal from time shifted and scaled wavelets, in contrast to the FT that uses a sine-wave. The wavelet ψ has finite duration, an example is given in figure 3. The discrete WT used in this application results in a continued high- and lowpass-filtering that resembles a tree structure as in figure 4. The resulting wavelet coefficients of the highpass filtering are called details $c_{\mathrm{D1}}, \ldots, c_{\mathrm{D5}}$, that of the lowpass filtering approximations $c_{\mathrm{A1}}, \ldots, c_{\mathrm{A5}}$. This can be visualized in figure 5 with the reconstructed signals $s_{\mathrm{D1}}, \ldots, s_{\mathrm{D5}}$, and s_{A5} by upsampling the wavelet coefficients.

The feature vector is computed according to

$$\mathbf{c}_{WT}(r) = \begin{pmatrix} \mathrm{LPCC}\{s_{\mathrm{D1}}(r\rho_\mathrm{s}), 3\} \\ \mathrm{LPCC}\{s_{\mathrm{D2}}(r\rho_\mathrm{s}), 3\} \\ \mathrm{LPCC}\{s_{\mathrm{D3}}(r\rho_\mathrm{s}), 3\} \\ \mathrm{LPCC}\{s_{\mathrm{D4}}(r\rho_\mathrm{s}), 3\} \\ \mathrm{LPCC}\{s_{\mathrm{D5}}(r\rho_\mathrm{s}), 3\} \\ \mathrm{LPCC}\{s_{\mathrm{A5}}(r\rho_\mathrm{s}), 3\} \end{pmatrix}, \tag{14}$$

in order to retain the strongest frequency components of each signal.

To select the best pre-processing algorithm for the EFPG signal, a RBFN with LVQ layer consisting of eight neurons is trained with each feature vector. Best results are achieved with the LPCC algorithm; $p = 12$ coefficients proved to be sufficient to represent the cardiac as well as the respiratory activity.

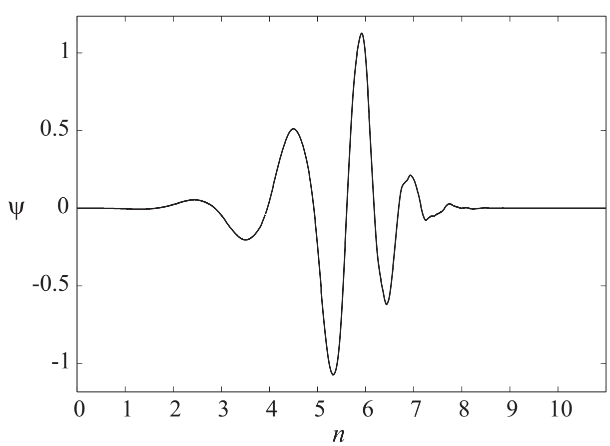

Figure 3: Daubechies wavelet of order 6.

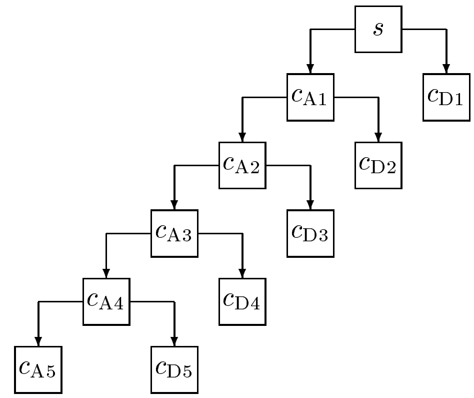

Figure 4: Wavelet Decomposition Tree.

The heart rate f_C and the blood oxygen saturation

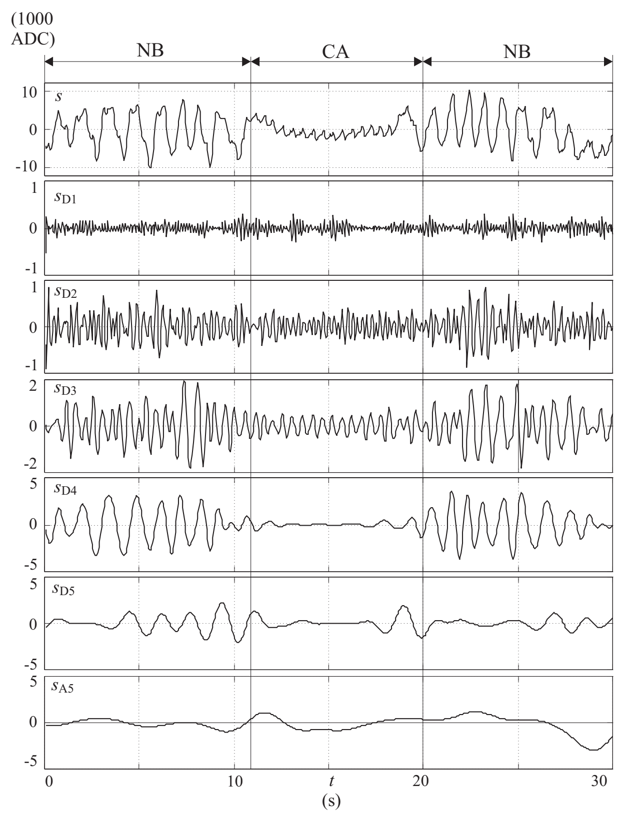

Figure 5: Wavelet decompositions of a CA with a Daubechies wavelets of order 10. The continued separation in a high and lowpass signal is clearly visible.

c_{ox} are represented as time vectors

$$\mathbf{c}_{\text{OX}}(r) = \begin{pmatrix} c_{\text{ox}}(r+5) \\ c_{\text{ox}}(r+10) \\ \vdots \\ c_{\text{ox}}(r+30) \end{pmatrix} - c_{\text{ox}}(r) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{15}$$

and

$$\mathbf{c}_{\text{HR}}(r) = \begin{pmatrix} \overline{f_{\text{C}}}(r+5) \\ \overline{f_{\text{C}}}(r+10) \\ \vdots \\ \overline{f_{\text{C}}}(r+30) \end{pmatrix} - \overline{f_{\text{C}}}(r) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{16}$$

with

$$\overline{f_{\text{C}}}(n) = \frac{1}{n} \sum_{k=\nu_{\text{C}}T(n-1)}^{\nu_{\text{C}}Tn-1} f_{\text{C}}(k); \quad \nu_{\text{C}} = 4\,\text{Hz}. \tag{17}$$

Principal Component Analysis - To reduce the dimensionality of the feature vectors, a PCA is performed for $\mathbf{c}_{\text{LPCC}}(r)$, $\mathbf{c}_{\text{OX}}(r)$ and $\mathbf{c}_{\text{HR}}(r)$ separately [13]. The number of principal components (PCs) of each feature vector is chosen to cover 90% of data variation. The formal notation

$$\mathbf{v}(r) = \text{PCA}(\mathbf{c}(r), k) \tag{18}$$

describes the calculation of the first k PCs of the vector set $\mathbf{c}(r)$. The resulting PCs for each feature vector are:

$$\mathbf{v}_{\text{OX}}(r) = \text{PCA}(\mathbf{c}_{\text{OX}}(r), 3), \tag{19}$$

$$\mathbf{v}_{\text{HR}}(r) = \text{PCA}(\mathbf{c}_{\text{HR}}(r), 3), \tag{20}$$

$$\mathbf{v}_{\text{LPCC}}(r) = \text{PCA}(\mathbf{c}_{\text{LPCC}_{12}}(r), 8). \tag{21}$$

The input vector for the RBFN

$$\mathbf{x}(r) = \begin{pmatrix} \mathbf{v}_{\text{LPCC}}(r) \\ \mathbf{v}_{\text{LPCC}}(r+6) \\ \mathbf{v}_{\text{LPCC}}(r-6) \\ \mathbf{v}_{\text{OX}}(r) \\ \mathbf{v}_{\text{HR}}(r) \end{pmatrix} \tag{22}$$

has a total of 30 components. $\mathbf{x}$ is standardized to a mean value of zero and a variance of one.

According to the above-defined classes NB, CA, OA and SI, the target vector is determined via a 1 of 4 coding scheme, with a formal description by

$$\mathbf{t}_1(r) = \begin{pmatrix} \delta(\widetilde{t}(r) - c_1) \\ \vdots \\ \delta(\widetilde{t}(r) - c_4) \end{pmatrix} \tag{23}$$

with $c_1 = 0$, $c_2 = 1$, $c_3 = 8$ and $c_4 = 9$.

Radial Basis Function Network

RBFNs are established with two layers (figure 6). Neurons of the first layer are activated by the Euclidean distance between input and code book vector (CV) $\boldsymbol{\mu}_k$ weighted with a Gaussian function with spread σ_k:

$$\gamma_k(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}_k\|^2}{2\sigma_k^2}\right). \tag{24}$$

CVs are determined by two training algorithms: Learning vector quantization (LVQ) and self organizing map (SOM) [14][15][16][17][18]. Neurons of the second layer are activated by the inner product of input and weight vector, applying a linear activation function. Weights are determined by the Levenberg-Marquard algorithm [16]. The complete RBFN can be described by

$$z_i(\mathbf{x}) = \sum_{k=1}^{K} w_{ik}\gamma_k(\mathbf{x}) + w_{i0}. \tag{25}$$

SOM - an unsupervised training algorithm - is used in one and two dimensional (quadratic) topology, while LVQ is supervised, not needing a special topology. Both training algorithms are compared, and σ_k is optimized.

The performance of the RBFNs is evaluated with a 10-fold cross validation algorithm [19] using 1000

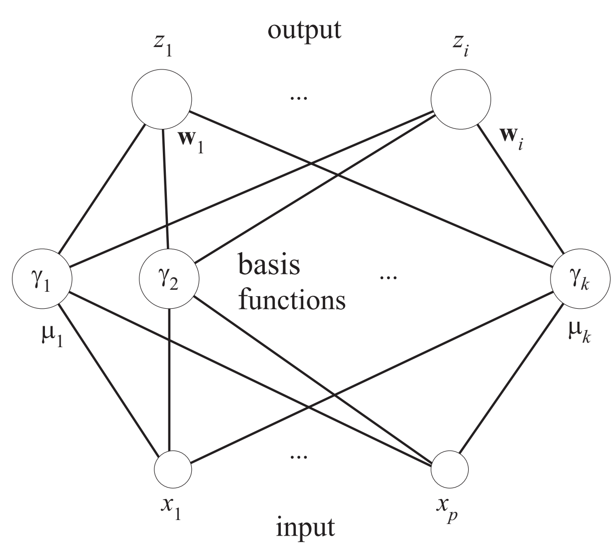

Figure 6: Structure of the RBFN. Each neuron in the hidden layer represents a basis function γ_k with the code book vector (CV) $\boldsymbol{\mu}_k$. The weights $\mathbf{w}_i$ are displayed as connections from the basis functions to the linear output neurons.

vectors including 176 events, equally distributed over the four (or three) classes for training and testing. Segmentation is performed based on vectors and based on events, that consist of about 20 vectors.

Committee of neural networks - V NNs can build a committee according to

$$\mathbf{z}_{\mathrm{COM}}(\mathbf{x}) = \frac{1}{V}\sum_{i=1}^{V}\mathbf{z}_i(\mathbf{x}). \tag{26}$$

In this work individual NNs from the different realizations during the V-times cross validation are used [16].

With the assumption of statistical independent NNs the error of the committee

$$E_{\mathrm{COM}} = \frac{1}{V}E_{\mathrm{AV}} \tag{27}$$

is reduced by the factor $1/V$ in relation to the error of an average NN

$$E_{\mathrm{AV}} = \frac{1}{V}\sum_{i=1}^{V}E_i. \tag{28}$$

E_{AV} is the average over the errors of the individual networks E_i.

The different NNs in this practical application are statistical dependent, however E_{COM} will be always lower than E_{AV}

$$E_{\mathrm{COM}} < E_{\mathrm{AV}}. \tag{29}$$

3 Results

In a general way, the achieved rates R of correct classification increased with the number of neurons in the hidden layer. Reaching its optimum for $k = 20$ neurons (figure 7). One-dimensional SOM showed higher variance of R. However, SOM needs more neurons for similar performance as LVQ (figure 8). The spread optimization process revealed that an individual spread for every neuron chosen as two times the distance to the next CV yields best generalization:

$$\sigma_k = 2 * \min_l(\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_l\|), \qquad l \neq k. \tag{30}$$

For vector based segmentation achieved with the RBFN with LVQ-layer and 20 neurons, the rate R for the baby patient group is $62.6\% \pm 6.8\%$. Segmentation based on events reduced R to $47.7\% \pm 10.7\%$

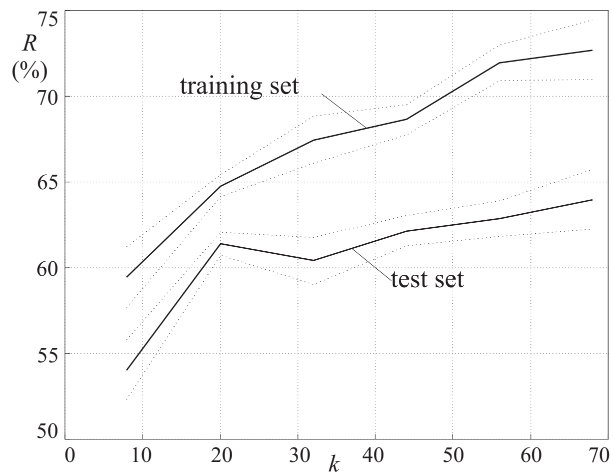

Figure 7: Recognition rate R for the RBFN with LVQ layer over varied number of neurons k.

because two vectors of the same event are more correlated than vectors of different events.

In the case of adults, a rate of $R = 64.4\% \pm 3.4\%$ is achieved for vector based segmentation. R decreased to $58.52\% \pm 4.5\%$ using event based segmentation. The fact of better rates for adults can be explained with a higher amount of training samples, reduced number of classes and also by more stationary character of physiological signals.

Building committee networks improved the rates by about 4% up to i.e. $52\% \pm 6.5\%$ for event based segmentation for the baby patient group. Post-processing algorithms that combine net results with regard to their time-dependence offer a further tool for the improvement of rates [20].

As a major conclusion, the performance of RBFNs proved to be quite similar to that of the backprop-

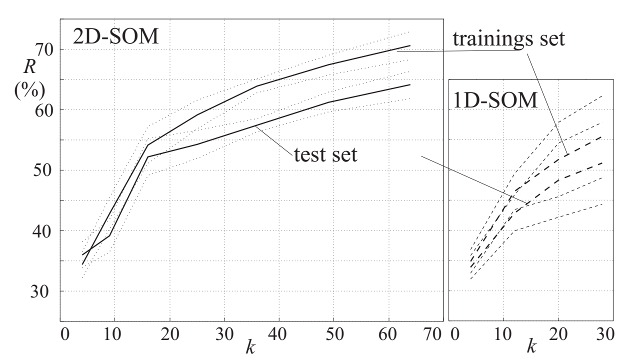

Figure 8: Recognition rate R for the RBFN with SOM layer in one and two dimensional topology over varied number of neurons k.

agation networks as applied in former studies [2][3]. This indicates that restricted recognition rates are not due to imperfections of NNs. Rather they can be explained by three reasons:

- lacks of the supplied physiological signals to characterize the different events in an unambiguous way,
- insufficiencies of subjective classifications performed by the medical staff for the establishment of material for training and testing, a procedure which cannot be performed in an exact way for monitoring durations of 8 hours with reasonable expenditure of time, and
- the a-priori impossibility to attribute the high variety of physiological activities to three or four classes.

However, with respect to a diagnostic tool, exact numbers of events are not needed. Relevance is rather given for their global distribution to the different classes.

Acknowledgments - This study was supported by the Austrian Federal Ministry of Science and Traffic; Project No. 49.930/2/II/4/95). The authors want to thank the Hospital of Mödling (Prim. F. Paky) and the Hospital of Lainz (OA. Rauscher) for data classification.

References

[1] P. A. D. Wilks, M. English. A system for rapid identification of respiratory abnormalities using a neural network. Med. Eng. Phys. 17, 551-555, 1995.

[2] Markus Clabian, C. Nussbaum, Helmut Pfützner. Artificial neural networks for apnea detection. Proc. EANN 601-608, 1996.

[3] Markus Clabian, Helmut Pfützner. Determination of decisive inputs of a neural network for sleep apnea classification. Proc. EANN 171-178, 1997.

[4] Helmut Pfützner, K. Futschik, A. Doblander, G. Schenz, H. Zwick. A method for the elimination of artefacts of electric field plethysmography signals. Front. Med. Biol. Engng. 2, 53-63, 1990.

[5] Christoph Ruhsam, Helmut Pfützner, P. Nopp, H. Nakesch, H. Frais-Klbl, W. Marktl. Time frequency spectral analysis of electric field plethysmography signals. Med. Prog. thr. Technol. 21, 17-28, 1995.

[6] Christoph Ruhsam. Digitale Verarbeitung feldplethysmographischer Signale zur simultanen Erfassung kardiorespiratorischer Funktion. Dissertation an der TU Wien, Wien, 1994.

[7] Alan V. Oppenheim und Roland W. Schafer. Discrete-Time Signal Processing. Prentice-Hall, 1989.

[8] Thomas P. Krauss, Loren Shure, John N. Little. Signal Processing Toolbox User's Guide. The MathWorks, Inc., 1994.

[9] Ananthram Swami, Jerry M. Mendel, Chrysostomos L. (Max) Nikias. Higher Order Spectral Analysis Toolbox User's Guide. The MathWorks, Inc., 1995.

[10] Frank Fallside, William A. Woods. Computer Speech Processing. Prentice Hall, 1985.

[11] Manfred R. Schroeder. Direct (Nonrecursive) Relation Between Cepstrum and Predictor Coefficients. IEEE Trans. Acoust., Speech, Signal Processing, pp. 297-301, 1981.

[12] Michael Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi. Wavelet Toolbox User's Guide. The MathWorks, Inc., 1996.

[13] I. T. Jolliffe. Principal Component Analysis. Springer-Verlag, New-York, 1986.

[14] Teuvo Kohonen. Self-Organization and Associative Memory. Springer-Verlag, Berlin, 1989.

[15] Howard Demuth, Mark Beale. Neural Network Toolbox User's Guide. The MathWorks, Inc., 1994.

[16] Christopher M. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.

[17] Raúl Rojas. Theorie der neuronalen Netze. Springer-Verlag, Berlin, 1993.

[18] Monika Köhle, Neurale Netze. Springer-Verlag, Wien - New York, 1990.

[19] Arthur Flexer. Statistical Evaluation of Neuronal Network Experiments: Minimum Requirements and Current Practice, Wien, 1996.

[20] Markus Clabian. Automatische Klassifikation von Schlafapnoen mittels Neuralen Netzen. Dissertation an der TU-Wien, Wien, 1998.