CLASSIFICATION OF VISUAL SENSATIONS GENERATED ELECTRICALLY IN THE VISUAL FIELD OF THE BLIND

Cédric Archambeau[1], Jean Delbeke[2], Michel Verleysen[1,*]

[1]*Université catholique de Louvain – Microelectronics Laboratory*
Place du Levant 3, B-1348 Louvain-la-Neuve, Belgium
[2]*Université catholique de Louvain – Neural Rehabilitation Engineering Laboratory*
Avenue Hyppocrate 54, B-1200 Bruxelles, Belgium

Abstract: Within the framework of the OPTIVIP project, an optic nerve based visual prosthesis is being developed in order to restore partial vision to the blind. In this paper, we concentrate on the classification problem of visual sensations generated by multiple electrical stimulations. We propose to use probabilistic neural networks in order to perform Bayesian classification. Statistical sampling techniques are utilized in order to reduce the bias on the estimated performances and assess the sensitivity of the method. In noisy environment the Parzen window estimator seems more reliable than finite Gaussian mixtures. *Copyright © 2003 IFAC*

Keywords: probabilistic neural networks, probabilistic models, probability density function, classification, parameter estimation, cross-validation, electrical stimulation, optic nerve

1. INTRODUCTION

Since the late eighties cochlear implants have rehabilitated deaf patients for whom there was no other potential treatment. Spurred by this success, several multidisciplinary teams were established during the past decade aiming at restoring partial vision to the blind and improving their quality of life. The principle consists in implanting a neural prosthesis either intra-ocularly or intra-cranially and bypass, by electrical stimulation, neurons that have become non-functional.

In the previous European project MIVIP (*MI*crosytem based *VI*sual *P*rosthesis), the feasibility of the optic nerve based visual prosthesis was investigated. The electrical stimulation of the optic nerve was demonstrated (Veraart, *et al.*, 1998), in particular for pathologies for which the optic nerve remains partly intact and functional. Retinitis pigmentosa (RP), a leading cause of blindness in the western world, is an example among others. As a consequence, further research was undertaken: a

microelectronic prototype was built (Doguet, *et al.*, 2000), a considerable amount of data was gathered concerning the visual sensations evoked in a blind RP volunteer and an attempt was made to coarsely understand the underlying physiological process (Delbeke, *et al.*, 1999).

Within the framework of the European project OPTIVIP (*OPTI*misation of an *I*mplantable *V*isual *P*rosthesis), the optic nerve based visual prosthesis has been further developed and optimised. One of the main challenges is to understand, decode and model the physiological process linking the stimulating parameters to the visual sensations produced in the visual field of a blind volunteer. A black-box prediction model has been developed by Archambeau, *et al.* (2001), which showed satisfactory prediction accuracy for single contact stimulations, i.e. stimulations generated by a single electrode contact and leading to a single visual sensation. The primary aim of such prediction model is to allow latter reconstruction of basic shapes extracted from video images after some edge

detection and, as a result, supply meaningful visual information to the blind. In addition, the achievable accuracy and reconstruction resolution is being evaluated.

In this paper, we provide a tool to decompose multiple stimulations, i.e. complex stimulations generated by multiple contact activations and leading to (apparently) simultaneous multiple visual sensations, to single ones. A proper decomposition of multiple stimulations will enable us to explore the combining rules of complex stimulations by advanced data analysis, to simplify greatly the stimulation strategy and to increase extensively the available experimental data, which is extremely valuable for improving the black-box prediction models.

This paper is organised as follows. In section 2, we address the problem of classifying visual sensations generated electrically. The next two sections describe the structure of probabilistic neural networks (PNN) (Specht, 1990). The PNN is defined as a 3-layer feed-forward neural network capable of approximating Bayes' classifier. The first and the second hidden layers are described in section 3. They consist in estimating the probability density function (PDF) of the different classes. Two alternatives are considered: the Parzen window estimator, a non-parametric PDF estimator, and finite mixture models, a semi-parametric PDF estimator. The third layer of the PNN is introduced in section 4. It makes the optimal decision according to the Bayes' law. In section 5, we propose a method to estimate statistically the classification performances and finally in section 6 we apply those techniques on the classification of visual sensations and discuss our results.
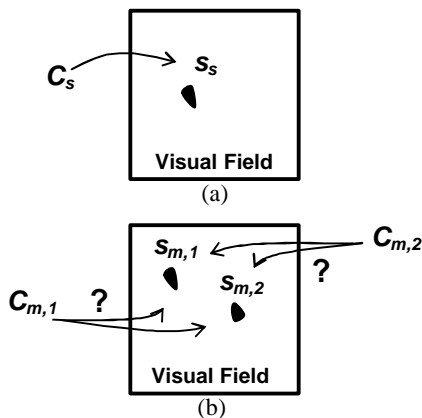

(a)

(b)

Fig. 1. Definition of the correspondence problem between the visual sensations and the active electrode contacts, (a) for the single stimulation case and (b) the complex stimulation case.

2. CLASSIFICATION PROBLEM

The microelectronic prototype used in OPTIVIP controls a four contact cuff electrode, wrapped around the right eye's optic nerve of a volunteer. Consider a single electrode contact stimulation C_s. The electrical stimulation C_s produces a visual sensation s_s in the visual field of the blind volunteer (figure 1a). Single visual sensations are called

phosphenes. Next, consider a complex stimulation formed by two simultaneous electrode contact stimulations $C_{m,1}$ and $C_{m,2}$. It has been experimentally established that each contact stimulation produces one single phosphene. As will be discussed in section 6, this means that we can perform classification of multiple phosphenes based on the knowledge of the class PDF of single ones. The classification problem can therefore be stated as finding, for each perceived phosphene, the corresponding electrode contact among the active ones (figure 1b).

3. PROBABILITY DENSITY ESTIMATION

When we have a sufficient number of data points to estimate the probability density functions (PDF) of the different classes, we may be inclined to perform a Bayesian classification, which can be stated as the optimal classification scheme. In this section, we recall two techniques to estimate PDFs: Parzen windowing and the finite mixture models.

3.1 Parzen Windowing

The Parzen windowing PDF estimation (Parzen, 1967) is a non-parametric method. Such a method does not assume any functional form of the PDF, but allows its shape to be entirely determined by the data. It consists in placing a well-defined kernel function on each data point $\mathbf{x}_n$ and determining a common width s, also defined as a smoothing parameter. In practice, Gaussian kernels are often used. The total PDF is then defined as the sum of all the Gaussian kernels (Silverman, 1986):

$$p(\mathbf{x}) = \frac{1}{N} \cdot \sum_{n=1}^{N} \frac{1}{\left(2\boldsymbol{p}\boldsymbol{s}^2\right)^{d/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{2\boldsymbol{s}^2}\right), \quad (1)$$

where N is the number of data and d their dimension.

The main advantage of Parzen windowing over simple histograms is that the shape of the PDF is not discontinuous at the arbitrary bin edges and does not depend on an arbitrary choice of the origins, which can both lead to a false representation of the PDF.

3.2 Finite Mixture Model

Finite mixture models are semi-parametric PDF estimation methods. As non-parametric techniques, they do not assume the a priori shape of the PDF to estimate. However, unlike the previous method, the number of kernel functions, also denoted components, is limited to a number $M << N$ in order to avoid a prohibitive increase of the number of parameters with the size of the data set. In practice Gaussian kernels are often used. A Gaussian mixture model can be defined as a linear combination of the Gaussian component densities (McLachlan and Peel, 2000):

$$p(\mathbf{x}) = \sum_{m=1}^{M} \boldsymbol{p}_m \cdot \frac{1}{\left(2\boldsymbol{p}\boldsymbol{s}_m^{\,2}\right)^{d/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_m\|^2}{2\boldsymbol{s}_m^{\,2}}\right), \quad (2)$$

where $\mathbf{c}_m$ and s_m are the centres and the widths of the Gaussian kernels respectively and p_m the mixing proportions, which are non-negative and must satisfy the following constraint:

$$\sum_{m=1}^{M} \boldsymbol{p}_m = 1. \qquad (3)$$

Such mixture models can approximate any continuous PDF, provided the model has a sufficient number of components and provided the parameters of the model are chosen correctly (Bishop, 1995).

Next, let us define the likelihood function L :

$$\mathsf{L} = \prod_{n=1}^{N} p(\mathbf{x}_n). \qquad (4)$$

Maximizing the likelihood function is then equivalent to finding the most probable PDF estimate provided the data set $\{\mathbf{x}_n\}_{n=1}^{N}$. In order to avoid the complexities of a non-linear optimisation scheme for computing the maximum likelihood estimate (MLE), a practical method is obtained by applying the expectation-maximization (EM) algorithm (Dempster, *et al.*, 1977). The EM is a two stage iterative method. First, in the *E-step*, the expected value of some "unobserved" data is computed, using the current parameter estimates and the observed data. Subsequently, during the *M-step*, the expected values computed in the E-step are used to compute the MLE and the model parameters are updated. Each iteration t of the EM can be summarized as follows (McLachlan and Peel, 2000):

E-step:

$$\boldsymbol{t}_l^{(t)}(\mathbf{x}_n) = \frac{\boldsymbol{p}_l^{(t)} p_l^{(t)}(\mathbf{x}_n)}{\sum_{m=1}^{M} \boldsymbol{p}_m^{(t)} p_m^{(t)}(\mathbf{x}_n)}, \qquad (5)$$

where:

$$p_l^{(t)}(\mathbf{x}_n) = \frac{1}{\left(2\boldsymbol{p}\boldsymbol{s}_l^{2(t)}\right)^{d/2}} \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{c}_l^{(t)}\|^2}{2\boldsymbol{s}_l^{2(t)}}\right).$$

M-step:

$$\mathbf{c}_l^{(t+1)} = \frac{\sum_{n=1}^{N} \boldsymbol{t}_l^{(t)}(\mathbf{x}_n)\mathbf{x}_n}{\sum_{n=1}^{N} \boldsymbol{t}_l^{(t)}(\mathbf{x}_n)}, \qquad (7)$$

$$\boldsymbol{s}_l^{2(t+1)} = \frac{\sum_{n=1}^{N} \boldsymbol{t}_l^{(t)}(\mathbf{x}_n)\|\mathbf{x}_n - \mathbf{c}_l^{(t+1)}\|^2}{d \cdot \sum_{n=1}^{N} \boldsymbol{t}_l^{(t)}(\mathbf{x}_n)}, \qquad (8)$$

$$\boldsymbol{p}_l^{(t+1)} = \frac{1}{N} \cdot \sum_{n=1}^{N} \boldsymbol{t}_l^{(t)}(\mathbf{x}_n). \qquad (9)$$

Note that in equation (5) $\boldsymbol{t}_l^{(t)}$ corresponds to the posterior probability that $\mathbf{x}_n$ is generated by component l provided that the data point $\mathbf{x}_n$ is known. The convergence properties of the EM algorithm have been discussed by Xu and Jordan (1996).

4. CLASSIFICATION

Once we have computed the PDF $p(\mathbf{x}_n|C_i)$ of the different classes, we can perform an optimal classification based on the Bayes' law:

$$P(C_i|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|C_i) \cdot P(C_i)}{p(\mathbf{x}_n)}. \qquad (10)$$

In this equation $P(C_i|\mathbf{x}_n)$ stands for the posterior probability that data point $\mathbf{x}_n$ belongs to class C_i provided $\mathbf{x}_n$ is known, whereas $P(C_i)$ stands for the priors. They can simply be estimated by counting the number of samples in each class. At this point it should be stressed that an optimal classification implies a minimization of the classification error, but it does not entail an error-free classification.

5. PERFORMANCE ESTIMATION

The two previous sections provide the building blocs for performing optimal classification by PNNs. As in the classical neural networks theory, in order to estimate the model performances, the data set has to be divided in a learning set, for which the model parameters are estimated, a validation set, for which the optimal model is selected, and a test set, on which the performances of the optimal model are evaluated.

However, for some applications the data acquisition turns out to be fastidious and time consuming, especially when we are not dealing with an automatic data acquisition procedure and when a human being is involved, as for example in OPTIVIP. Therefore, we have favoured the approach of avoiding the definition of a test set. This enables us to have a greater amount of data at our disposal during the learning step and thus enhance the approximation quality of the PDF. Moreover, in practice, there is only a slight difference between the classification errors of the optimal model evaluated on the validation set and the test set.

Nevertheless, the data split is still arbitrarily. This can lead to a strong bias in the model parameters, which in turn can cause poor performance estimation and a bad generalisation by an inappropriate model selection. In order to reduce the bias and estimate the data sensitivity of the probabilistic model, statistical sampling methods can be used. One popular technique for achieving this is the *K*-fold cross-validation. The method can be summarized as follows (Efron, 1998):

1. The initial data set is split in K roughly equal sized subsets.
2. For $k = 1 \ldots K$:

- Define the validation data set D_V as the kth part of the split and the learning data set D_L as the remaining $K-1$ parts;
- Calculate the classification error E_k computed on D_V, for the model fitted on D_L;

3. Compute the statistical estimate of the classification error and its standard deviation:

$$\boldsymbol{m}_{CV} = \frac{1}{K} \cdot \sum_{k=1}^{K} E_k \ , \qquad (11)$$

$$\boldsymbol{s}_{CV} = \sqrt{\frac{1}{K-1} \cdot \sum_{k=1}^{K} \left| E_k - \boldsymbol{m}_{CV} \right|^2} \ . \quad (12)$$

Such statistical sampling method makes it possible to use a high proportion $(1 - \frac{1}{K})$ of the available data to train the networks, while also making use of all the data points when evaluating the cross-validation error. A disadvantage of the approach is that it requires the training process to be repeated K times, which can lead to large processing times.

6. CLASSIFICATION OF VISUAL SENSATIONS

In this section, we will demonstrate the relevance of PNNs for approximating the optimal classifier in a real life problem: the classification of phosphenes generated in the frame of the OPTIVIP project.

By electrical stimulation, visual sensations are selectively produced in the visual field of the blind RP volunteer, as drawn in figure 2. The locations depicted in each box correspond to phosphenes induced by successively activating one single electrode contact, letting the stimulation parameters vary.

The implanted cuff electrode contains four electrode contacts, wrapped around the optic nerve. Each contact C_i is named by its radial position around the nerve:

$$C_i : i = \left\{ 0°, 90°, 180°, 270° \right\} \qquad (13)$$

Experimentally, it was found that for each contact, a restricted, but dissimilar area of the visual field is accessible (figure 2). This experimental fact confirms the hypothesized retinotopic structure of the optic nerve, which postulates that the neighbouring optic nerve fibres correspond to the neighbouring photosensitive cells of the retina.
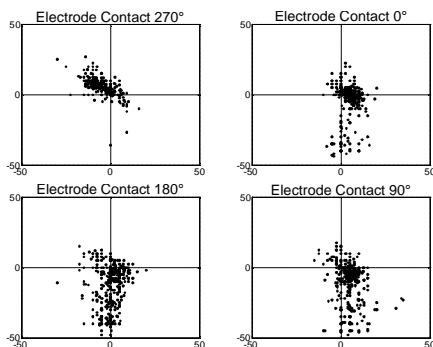
Fig. 2. Location of the phosphenes in the visual field of the blind RP volunteer, measured experimentally for each of the four electrode contacts. The positions are measured in degrees, in both horizontal and vertical directions.

Furthermore, for complex stimulation it was established that the number of perceived phosphenes is, in most cases, equal to the number of activated electrode contacts. Indeed, depending on the type of complex stimulation, this was confirmed for 80% to 90% of the data.

Based on these two experimental results, it seems reasonable to assume spatial superposition when complex stimulations are provoked. That is, when a set of single contact activations are combined in order to build one complex stimulation, we can assume the resulting visual perception as being the spatial superposition, in the visual field, of all the single visual perceptions. In practice however, a slight influence has been noticed on the exact phosphene location, but the effect is limited and localized inside the specific area of each electrode contact.

6.1 Probability Density Estimation

In order to estimate adequately the class PDFs, the optimal model parameters are selected. First, let us define the total classification error E_{tot}:

$$E_{tot} = \frac{1}{4K} \cdot \sum_{k=1}^{K} \left(E_{k,C_{0°}} + E_{k,C_{90°}} + E_{k,C_{180°}} + E_{k,C_{270°}} \right)$$

$$= \frac{1}{4} \cdot \left(\boldsymbol{m}_{CV,C_{0°}} + \boldsymbol{m}_{CV,C_{90°}} + \boldsymbol{m}_{CV,C_{180°}} + \boldsymbol{m}_{CV,C_{270°}} \right)$$

$$(14)$$

where E_{k,C_i} is the classification error for class i associated to the kth part of the data split, that is, the number of misclassifications divided by the number of data of the validation set corresponding to this class.

In figure 3, the estimated classification error by 10-fold cross-validation is represented. We clearly see that for Parzen windowing the minimum E_{tot} is obtained for a smoothing factor s equal to 2.5. When we use Gaussians mixtures to estimate the PDFs the optimal number of components per class is located around 5.
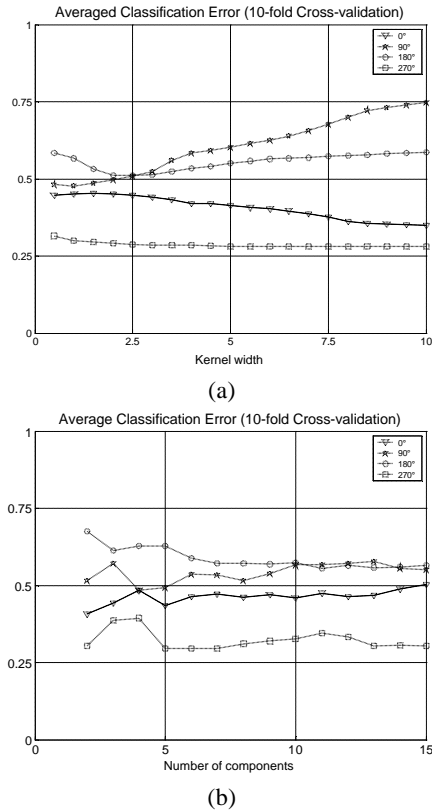
(a)

(b)

Fig. 3. Classification error for each class statistically estimated by 10-fold cross-validation. (a) depicts the classification error of the PNN using the Parzen windowing estimator and (b) the PNN using finite mixture models.

(a)

(b)

Fig. 4. Standard deviation of the classification error for each class based on (a) Parzen windowing estimator and (b) mixture models.

As we compare both methods, we see that, in practical problems that can be considered as very noisy due to the low number of available data (around 500 per class), the Parzen windowing estimator appears more reliable than iterative methods such as mixture models. The last get trapped in local minima, which leads to a higher total classification error and a greater sensitivity of the performances on the data subset selection. Indeed, looking at figure 4, where the standard deviation of the classification errors of the different classes is represented, we observe a greater variability for the Gaussian mixture models. Note that the scale is different from figure 3.

The PDF of the four electrode contacts are shown in figure 5. The optimal class PDFs are computed using the complete data set. Unfortunately it can be seen that the classes are strongly overlapping. This means that even for an optimal classification, the classification error will be relatively large.

6.2 Classification Performances

The performances of a classification scheme can be evaluated by using a confusion matrix. The confusion matrix $\mathbf{M}$ is defined as follows:

$$(m)_{ji} = \int_{\Re^d} p(\mathbf{x}|C_i) f_j(\mathbf{x}) d\mathbf{x}, \qquad (15)$$

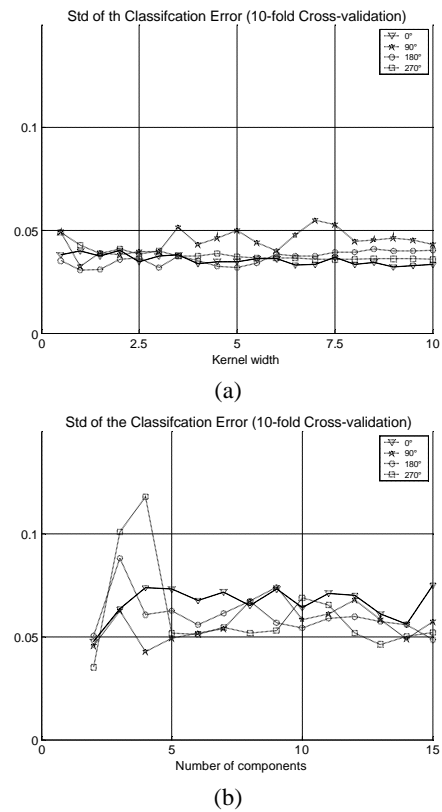where $f_j(\mathbf{x})$ is the estimated class indicator function and $p(\mathbf{x}|C_i)$ is the estimated posterior density. Thus, the confusion matrix counts, for each class i, the number of data that are correctly classified and the number of misclassifications, both labelled by their classification bin j in which they are classified.

Based on 10-fold cross-validation, the estimated classification performances are represented in table 1 when four phosphenes are perceived.

Table 1 Confusion Matrix for four activated contacts.
(Correct classifications are bolded)

Electrode Contact	Bin 0°	Bin 90°	Bin 180°	Bin 270°
0°	**56**%	20%	17%	09%
90°	23%	**52**%	18%	07%
180°	17%	22%	**43**%	18%
270°	13%	08%	09%	**70**%

At first sight, the performances are relatively poor (55% of correct classification on average). This is mainly due to the important overlapping of the different classes as it can be observed in figure 5. Nevertheless, in practice most of the complex stimulations are induced by activating only two or three single contacts. When fewer contacts are used,

the classification performances increase rapidly since the overlapping is reduced. Table 2 and 3 show examples of the increase of correct classifications for 3 and 2 perceived phosphenes respectively.

Table 2 Confusion Matrix for three activated contacts. (Correct classifications are bolded)

Electrode Contact	Bin 0°	Bin 180°	Bin 270°
0°	**70**%	21%	09%
180°	22%	**63**%	14%
270°	16%	09%	**76**%

Table 3 Confusion Matrix for two activated contacts. (Correct classifications are bolded)

Electrode Contact	Bin 90°	Bin 270°
90°	**89**%	11%
270°	19%	**81**%

7. CONCLUSION

In this paper, we have addressed the classification problem of visual sensations induced by electrical stimulation of the human optic nerve. We have provided a practical tool for decomposing complex stimulations into single ones that is showing satisfactory performances for further study of the underlying physiological process. Optimal classification was performed using PNNs. The actual performances of the method were statistically estimated by K-fold cross-validation, reducing the bias and estimating the sensitivity. Comparing the Parzen window estimator and finite Gaussian mixtures, we showed that in noisy environments the former was more reliable. Moreover, the class PDF estimates gave us additional insight in the physiological process relating the stimulating parameters to the perceived visual sensations.
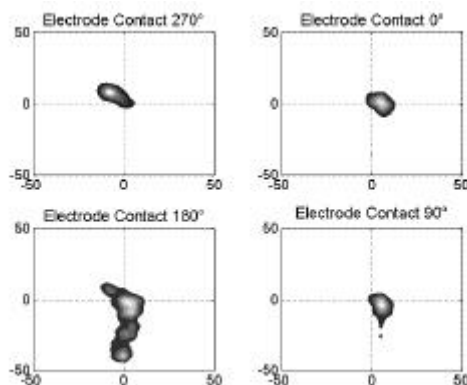

Fig. 5. Class PDF estimation of the phosphenes in the visual field of the blind volunteer, based on their location.

REFERENCES

Archambeau, C., A. Lendasse, C. Trullemans, C. Veraart, J. Delbeke and M. Verleysen (2001), Phosphene Evaluation in a Visual Prosthesis with Artificial Neural Networks, *Proc. EUNITE2001,* Puerto de la Cruz, Tenerife (Spain), December 13-14, 509–515.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, chapter 2. Oxford University Press, Oxford.

Delbeke, J., S. Parrini, G. Michaux, A. Vanlierde, C. Veraart (2000), Perception Threshold Changes in Phosphenes Generated by Direct Stimulation of a Human Optic Nerve, *Proc. IFESS2000-NP2000*, Aalborg, Denmark, June 18-19, 152–155.

Dempster, A.P., N.M. Laird, D.B. Rubin (1977), Maximum Likelihood from Incomplete Data via the *EM* Algortihm, *J. Roy. Stat. Soc. (B)*, **39**, 1–38.

Doguet, P., H. Mevel, M. Verleysen, M. Troosters, C. Trullemans (2000), An Integrated Circuit for the Electrical Stimulation of the Optic Nerve, *Proc. IFESS2000*, Aalborg, Denmark, June 18-19, 18–20.

Efron, B., and R.J. Tibshirani (1998). *An Introduction to the Bootstrap*, chapter 17. Chapman & Hall, London.

McLachlan, G., and D. Peel, (2000). *Finite Mixture Models*. Wiley, New York.

Parzen, E. (1962), On Estimation of a Probability Density Function and Mode, *Annals of Math. Statistics*, **33**, 1065–1076.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Specht, D.F. (1990), Probabilistic Neural Networks, *Neural Networks*, **3**, 109–118.

Xu, L., and M.I. Jordan (1996), On Convergence Properties of the EM Algorithm for Gaussian Mixtures, *Neural Computation*, **8**(1), 129–151.

Veraart, C., C. Raftopoulos, J.T. Mortimer, J. Delbeke, D. Pins, G. Michaux, A. Vanlierde, S. Parrini and M.C. Wanet-Defalque (1998), Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode, *Brain Res.*, **813**, 181–186.