Informationstheoretische Methoden für die funktionelle Anpassung von Retina-Implant-Parametern

Zur Erlangung des akademischen Grades

Doktor-Ingenieur

vom Fachbereich Elektrotechnik
der Universität Paderborn
genehmigte Dissertation von

Dipl.-Ing. Marcus Eger

aus Rinteln

Referent:	Prof. Dr. rer. nat. G. Hartmann
Korreferent:	Prof. Dr.-Ing. R. Eckhorn
Mündliche Prüfung:	30. November 2001

D 14–171

Information Theoretical Methods for the Functional Adjustment of Retina Implant Parameters

Dissertation by

Dipl.-Ing. Marcus Eger

in partial fulfillment
of the requirements for the degree

Doctor of Engineering

accepted by the Faculty of Electrical Engineering,
University of Paderborn

Advisors:	Prof. Dr. rer. nat. G. Hartmann
	Prof. Dr.-Ing. R. Eckhorn
Oral defense:	30. November 2001

D 14–171

Acknowledgements

First of all, I thank my advisor, Professor Reinhard Eckhorn, who committedly identified with the theory underlying this dissertation and who provided the possibility for my research. Already in the seventies he introduced information theory to neuroscience. His pioneering spirit challenged me to be open for unconventional ideas and certainly has influenced my thinking.

I am greatly indebted to Professor Georg Hartmann from the University of Paderborn for his unbureaucratic willingness to support my graduation at his engineering faculty. He cared about this inter-disciplinary dissertation that fits in neither with medicine, physics, or electrical engineering, but shares elements of all of them. Professor Klaus Meerkötter, also from Paderborn, contributed some very valuable remarks, especially on mathematical derivations.

The experiments that provided the data for the analyses performed would never have been possible without the help of my valued colleagues. In the first place there is my dear friend and brother in the spirit Marcus Wilms. Further on, our group leader Dr. Thomas Schanze, who held the connection to the running German retina implant projects. We discussed many topics, with tea and without, which had stimulating effects on all of us.

I would like to thank also Dr. Lutz Hesse for carrying out lots of excellent eye surgeries and Dr. Bernd Nebeling for his medical assistance. Further, the technical team of the department, Cornelia Czellner, Wilfried Gerber, Andreas Rentzos, and Uwe Thomas for their support. Our secretary Sigrid Thomas was always there, a special help in bureaucratic matters. She reminded me of things I always tend to forget.

Despite of much project-related work I had the chance of a 6 weeks' research stay in the USA last year, which Professor Reinhard Eckhorn promoted. This time had an inspiring impact on my research. Many motivating conversations widened my horizon. It was the personal hospitality of Professor John Donoghue (Brown, Providence) and Barry Richmond M. D. (NIH, Bethesda) which made this a special time.

Thanks a lot for reading the manuscript Martin Kosmalski, Martin Vogel, Erwin Wagner, Eddy Wölbern, and especially Thomas Eger. Thanks to my parents, not only for sending me huge poppy-seed stollen at times when I did not have a minute left to prepare anything. The folk sessions and cross-country jogs with my brothers kept me sane. Yet, some more of them would not have done harm ...

Thanks to God who got me through threatening times of sickness and to whom I attribute many good ideas.

Marburg, November 2001

Marcus Eger

This dissertation is dedicated to my beloved wife
Claudia
who devotedly supported me in a unique way
in the process of finishing the thesis
while taking exams herself.
I am thankful to God that our relationship grew
rather than suffered from this joint experience.

Contents

Abstract

The development of retina implants requires the adjustment of device and stimulation parameters with respect to perception. At the present state of the project psychophysical experiments with humans are ethically not acceptable except for key studies. Here, animal experiments provide valuable data because informative activity in the primary visual cortex is a prerequisite for visual percepts. To assess efficiency, reliability, and selectivity of the electrical stimulation of retinal ganglion cells, two information theoretical methods have been developed.

The *three step method* characterizes multi-input and -output neuronal systems in case of continuous stimulation. In three steps a lower bound of transinformation T is provided: (1) Estimation of the deterministic response to isolate components carrying stimulus information. (2) Coordinate transformation using PCA yields a linear independent representation. (3) Partial values of T are individually calculated by Shannon's formula assuming normality or by density estimation. The approach allows to evaluate the degree to which stimulus features are encoded in the cortical response. A second method applies to rapid series of transient stimuli. While the three step method potentially captures all stimulus aspects, the second method only assesses non-temporal aspects. It quantifies the ability of the system to discriminate between different discrete stimuli that potentially evoke temporally overlapping responses. The performance of both algorithms is investigated using simulated data.

The methods are applied to experimental data recorded in the visual cortex of the anesthetized cat and yield the following results. (1) *Temporal*: Single stimuli evoke precise spike responses with 300 µs standard error. The decorrelated fraction of the population response lasts about 40 ms. Rapid sequences of stimulus events convey a maximum of T at a mean stimulus rate of 20–40 Hz suggesting a temporal resolution of 25–50 ms. Stimulus efficiency is highest 10–20 ms or long ($>$150 ms) after a facilitating stimulus. The temporal stimulus aspect accounts for 50–80% of total T. (2) *Intensity*: Injected charge is a weak parameter and requires at most three quantization steps. Multi unit activity encodes intensity better than local field potentials (3% vs. 15% of total T) yet shows an increasing variance with growing response strength. This recommends scarcely over-threshold stimuli. (3) *Spatial*: The values of T range within 20–100 bit/s per electrode for both electrical and visual stimulation. Not all recording positions are equally informative. With electrical stimulation the cortical profiles of T show a pronounced peak of < 1 mm radius suggesting a resolution of 2°. In case of visual stimulation the profiles are highly structured. T saturates when stimulating the retina with 7 electrodes per mm^2 confirming the resolution of $<2^\circ$.

Originality and Contributions

The animal experiments that provided the data for the investigations in this dissertation were jointly conducted, primarily with Dr. T. Schanze and M. Wilms.

The author wrote about 50.000 lines of code for the generation and application of electrical stimuli (in `c++`, Borland), the basic preprocessing of the data (in `IDL 5.4`, Research Systems Inc.), and high level statistical analysis (in `R`, Ihaka and Gentleman (1996)). The maintenance and improvement of the programs for the data preprocessing was done in teamwork with M. Wilms.

Not only in the process of developing the new methods that make up most part of Chapters 2 and 3, I have greatly appreciated the discussion with my advisor R. Eckhorn and his constructive comments. He is coauthor of the respective papers.

1 Introduction

The *adjustment of retina implant parameters* is the goal behind this dissertation. A retina implant is an electronic prosthetic device intended for people being blind due to photoreceptor degeneration. As a replacement for the malfunctional light sensitive cells, the implant receives information about the visual scene by an attached camera and transforms it into suitable electrical impulses which excite still intact retinal ganglion cells.

Background. The basic idea behind a neuroprosthesis is not new. Already in the late eighteenth century Luigi Galvani and Allesandro Volta conducted pioneering experiments eventually demonstrating that it was possible to stimulate neuronal tissue by means of electricity. Since then, many devices – now partly in clinical use – have been developed to supplement neurological functions. Examples are heart and brain pace makers, bladder control devices, "stand and transfer" prostheses, and cochlear implants for the stimulation of the hearing nerve in deaf people (Loeb, 1989; Agnew and McCreery, 1990; Eckmiller et al., 1995; Finn, 1997). Prostheses for the restoration of vision are still at an early stage. Clinically used functional prostheses do not exist, yet. However, the first report of a working visual prosthesis based on a cortical implant was reported last year (Dobelle, 2000).

Most probably, the idea of a visual prosthesis was born in the context of the findings of electrically excited "phosphenes". Phosphenes are light sensations caused by non-light stimuli. A shining example is the closed eye sensation of a round object at a position opposite to where one touches the globe of one's eye. Psychophysical experiments with humans equipped with two electrodes on forehead and temporal bone revealed that the electrically evoked phosphenes depend on background illumination and amplitude and frequency of the intermittent electrical stimulus signal (van de Grind et al., 1973; Kato et al., 1983). A frequency of about 20 Hz has the lowest stimulation threshold, i.e., it is optimal in the sense that a smallest current amplitude is needed to evoke a response.

In 1968 Potts and Buffum introduced a technique for the simultaneous recording of electrically evoked cortical response potentials using scalp electrodes. This approach is important for this thesis because here, cortical response signals are related to potential perception. In the same year, the success of the first cortically invasive experiment not embedded in an otherwise necessary surgery was published by Brindley and Lewin (1968). After the implantation of 80 telemetricly controlled platinum electrodes above the right visual cortex of a totally blind woman, Brindley

and Lewin reported that the subject experienced sensations of light in the left half of the visual field in response to electrical stimulation. Single electrode stimulation usually produced small single spots at a constant position. The electrodes were 2.4 mm apart and could well be discriminated. Multi-site stimulation had additive effects. With enhanced stimulation strength the percept got increasingly bright and sometimes additional spots at very different positions appeared, disturbing the otherwise regular mapping between stimulation position and phosphene location.

Current situation. The present state of vision prosthetics does not seem to have improved dramatically in proportion to the time since. However, the cortical implant has become wearable (Dobelle, 2000) and several groups in the USA, Japan, and Germany have emerged, intensely working at a functional retina implant (Humayun et al., 1999; Ito et al., 1999; Zrenner et al., 1999; Grumet et al., 2000; Eckhorn et al., 2001). Since the fundamental study of feasibility ("Machbarkeitsstudie", Eckmiller et al., 1995) two dualistic approaches have been supported by the German government (Zrenner, 2001). While the SUBRET project aims at the restoration of vision based on light sensitive micro photodiodes implanted sub-retinally, the EPIRET consortium is developing an epi-retinal prosthesis that receives the information about the visual scene from a camera distant from the implanted chip. The advantage of the SUBRET idea is that the electrically elicited activation of retinal cells is in the close vicinity of the respective light information. Moreover, the signal processing capabilities of still intact sub-retinal cells can be utilized. However, most probably a post-processing and amplification of the photodiode output will be necessary for efficient stimulation (Zrenner et al., 2001). In contrast, the EPIRET chip is equipped with current sources that will be optimized with respect to stimulation efficiency. Yet it shortcuts potential retinal signal processing and (still) lacking an on-chip camera it depends on information transmission from an external transmitter and camera system, possibly incorporated in special glasses (Groß et al., 1997; Buß et al., 2000).

Open questions. Apart from challenges in the fields of surgery techniques, energy consumption, biocompatibility, and tissue safety (Guenther et al., 1999; Hesse et al., 2000; Kohler et al., 2001) important tasks comprise functional tests and the adjustment of stimulation parameters for estimating achievable resolutions and optimizing information transmission to the visual cortex. Some practical questions yet to be answered are the following: what is the maximum but still useful density of retinal electrodes? What is the highest stimulus impulse rate per electrode that is yet efficient? Furthermore, how many current amplitudes are to be "implemented", and which values are best suitable? How much information can be transmitted to the visual cortex at all? Invasive psychophysical experiments with humans in direct dialogue with the researcher are the best sources of knowledge to answer these questions, but for ethical reasons they are only possible for pioneering key studies (Brindley and Lewin, 1968; Humayun et al., 1999). Thus, recordings of stimulus evoked cortical responses in animal experiments are a necessary replacement –

with the underlying notion that measurable activity in the primary visual cortex is a fundamental requirement for perception. The experiments referred to in this thesis utilize the cat as the choice animal model because of its well known visual system being relatively similar to that of humans.

The following concepts are prerequisites for useful visual percepts: (1) *efficiency*, i.e., weak stimulation currents are sufficient to elicit perceptually relevant cortical activity, (2) *reliability*, i.e., evoked activity patterns are reproducible and have low variability, (3) *selectivity*, i.e., they are stimulus specific uniquely reflecting different stimulus properties, and (4) *stability*, i.e., the selectivity is stationary, the evoked percept is reproducible on a larger time scale. To some degree the first point has been examined by Dawson and Radtke (1977); Humayun et al. (1996, 1999); Hesse et al. (2000); Zrenner et al. (1999); Stett et al. (2000); Eckhorn et al. (2001), mostly with the aim to find efficient stimulus parameters though systematic in vivo investigations of the effect of different parameter settings are still lacking. To my knowledge, the other issues of reliability, selectivity, and stability have not been investigated so far.

Information theory. The idea to assess and maximize the information between electrical retinal stimulation patterns and registered neuronal activity is not equivalent to the evaluation of perception. However, the arrival of stimulus information at the visual cortex will be a prerequisite for perception. Here, the proof of stimulus evoked cortical activity (*efficiency*) is not sufficient – it is crucial to show that this activity is selective, reliable, and stable. Information theory aptly unifies the concepts of *selectivity* and *reliability*: the sizes of the alphabets of information source and sink correspond to the degree of selectivity, the signal-to-noise ratio – or more general, the signal and error distributions – find their equivalent in the concept of reliability. The issue of *stability* requires long term measurements based on functioning implanted devices and thus, is presently out of scope. In this thesis the issue of *efficiency* is only marginally addressed. However, as I will focus on information theoretical analyses, the static system properties have to be taken into account as well. For this reason, systematic analyses of stimulation efficiency are also tackled in the thesis, in particular the quantification of the response strength dependent on the stimulus current amplitude and temporal distance to a preceding stimulus.

Information theory was originally developed to quantitatively describe technical information channels (Shannon, 1948) but soon became popular in many other fields including neuroscience (Werner and Mountcastle, 1965; Eckhorn and Pöpel, 1974; MacKay and McCulloch, 1952; Optican and Richmond, 1987). Yet, the applicability of currently available information theoretical methods to in vivo neuronal information channels – especially in vertebrates – is constrained: neurophysiological experiments are necessarily relatively short and stationarity over the period of the experiment cannot be guaranteed. The complexity of biological systems is high and subject to many state variables such as day and night rhythms, condition of anesthesia and metabolism, just to mention a few.

This dissertation focuses on the analysis of the visual system of the anesthetized cat. Clearly, it is not like a technical system with a single input and output, but with many possibly correlated inputs of which only a few are considered relevant. The same applies to the outputs: it is not possible to register the full state vector of the animal's visual cortex. Signals are recorded only at a few locations in one or two of many visual cortical areas. Moreover, even at zero stimulus the recorded signals are highly structured and correlated due to spontaneous activity reflecting many neuronal state changes. This spontaneous activity does not terminate at stimulation. Since it cannot be explained by the stimulus it is usually defined as noise. These thoughts shall suffice to illustrate some of the challenges the thesis is facing.

Nonetheless, strategies suitable on certain conditions already exist and the methods that have been developed in the dissertation are partly based on them. However, it was necessary to develop new techniques or extend existing ones to widen the range of constraints with respect to the applicability of the methods, here. For example, there is an elegant method that allows to quantify the transmitted information between a (colored) Gaussian stimulus and multiple trains of registered action potentials (stimulus reconstruction technique, Bialek et al., 1991; Rieke et al., 1998). The problem is, that for the assessment of the efficiency and reliability of evoked responses to certain stimulus parameters, the restriction to single channel Gaussian stimuli is not acceptable.

Methods. In the following I will give a short overview concerning the motivation for the newly developed methods within this thesis. Detailed descriptions will be found in the introductory sections of Chapter 2 and 3. The quantification of the transmitted information from electrical stimulus patterns to cortical activity is associated with the *forward* concept of encoding: the better a certain stimulus aspect is encoded in the neuronal activity, the more information can be transmitted when this stimulus parameter is used in a potential code (Theunissen and Miller, 1995). Linked by Bayes' formula of probabilities there also exist *backward* approaches (e.g., the stimulus reconstruction technique, mentioned above) that focus on certain response deflections and deduce the responsible stimulus features (Borst and Theunissen, 1999). Here, the aim is to assess the efficiency of certain stimulus parameters, which votes for a forward approach.

The constraint of only short stationary data sections requires to use a model-based (*indirect*) mutual information estimator. *Direct* methods based on amplitude and time discretization of correlated analog signals are unsuitable due to an exploding number of joint symbols and the need to estimate their probabilities. In contrast, a model-based method isolates signal components not described by the deterministic model, defines them as noise and estimates their distributions. Finally, the quantified deterministic fraction and the noise have to be employed to yield the transinformation.

In addition to the categories forward–backward and direct–indirect, in neurophysiological experiments two different stimulation paradigms are common. The

simpler paradigm refers to *discrete stimulation* events with inter-stimulus intervals larger than the memory of the system. Here, the degree of encoding of a varied stimulus parameter can be tested by computing the amount of information that an evoked response conveys about the preceding stimulus. Usually, the response consists of a waveform and not a scalar value which may require dimensional reduction, e.g., by using principal component analysis. In this case, for the quantification of the transinformation well established methods exist (Optican and Richmond, 1987; Richmond and Optican, 1987). The second paradigm utilizes *continuous stimulation* signals. Different from the discrete case, a response value is potentially influenced by many preceding stimulus values depending on the memory of the system. To quantify the transmitted information often a model-based analysis is performed. It takes the temporal interferences into account and thus, investigates their encoding in the response. After separating signal and noise their respective powers are related in the frequency domain to obtain a frequency dependent signal-to-noise ratio. Finally, Shannon's information capacity formula yields a partial value of transinformation for each frequency bin (Shannon, 1948, 1949; Borst and Theunissen, 1999).

The approaches outlined above are perfect when they are used within their limitations: the discrete stimulation paradigm in combination with principal component analysis works if the inter-stimulus intervals are large enough. If they decline below the system's memory, response deflections begin to overlap rendering the straight forward analysis useless. Chapter 3 addresses this problem and presents a new method that allows to assess the encoding of non-temporal stimulus parameters by means of a modified principal component analysis. The continuous stimulation paradigm connected to a model-based approach as described above may yield correct partial transinformation values. Yet, if the stimulation signal deviates from being (colored) Gaussian or the system is highly nonlinear, the contributions of the frequency bins may be dependent and therefore, they must not be summed up to yield a total value of transinformation. For analog stimulus and response signals a forward model approach was not available except for one that uses the average response to multiple stimulus repetitions as model and thus, merely yields an upper bound of transinformation (Borst and Theunissen, 1999; Roddey et al., 2000). To fill the gap, Chapter 2 presents a new *three step method* that provides a lower bound of transmitted information and employs a linear multi-kernel model to support multi-dimensional stimulus signals without the restriction of Gaussianity. The additivity of partial transinformation values is achieved by decorrelation of stimulus and response using principal component analysis. This method captures the transmitted information about temporal stimulus aspects and optionally, in addition information about specific stimulus conditions. Both methods have been tested on the basis of artificial data.

Outline of analyses. The original motivation for the development of new methods was linked to the open questions concerning the functional adjustment of retina implant parameters: how efficient are certain stimulus parameters, i.e., what is the

achievable resolution, and which values are most suitable with respect to an optimized information transmission to the visual cortex? In cooperation with my colleagues more than 120 experiments and surgeries were conducted, of which about 25 yielded valuable data for information theoretical analyses. Five of these experiments provided the extensive examples analyzed in detail in Chapter 4. The other experiments still contain abundant data to further substantiate the obtained results.

The analyses were categorized according to *temporal*, *intensity*, and *spatial* aspects of information and with regard to the two stimulation paradigms (*discrete* and *continuous*) because the associated methods are conceptually different. In the following I will give a brief outline of the analyses performed.

A rather technical prerequisite deals with the localization and elimination of the influence of the electrical stimulus artifact due to volume conduction between stimulation and recording electrodes. Then, the temporal precision of evoked spike activity is examined and corresponding analog signals representing neuronal population activity. The stimulated retinal neurons are exemplarily identified with respect to X and Y pathway. To uncover informative response sections in case of available responses to many identical stimulus repetitions, an information theoretical method is developed and applied that may be used as an extension of the widely used peri-stimulus time histogram: this method implies a serial decorrelation based on Schmidt's orthogonalization procedure (Rice, 1966). It does not cover an own Chapter but comprises a separate publication (see Appendix 4.5.3 in Chapter 4 and Eger and Eckhorn (2002c)). Subsequent analyses are devoted to mean–variance relations and static characteristic curves: response strength and variance dependent on stimulus intensity and the duration of the preceding inter-stimulus interval. Then, the spectral distributions of transinformation – the partial transinformation profiles – are investigated, further the dependence of the total value of transinformation on mean stimulus rate, the number of independent stimulation electrodes, and the stimulus amplitude quantization. Finally, spatial profiles of transinformation are examined. The degree of redundancy and synergy between different recording positions is assessed by comparing the profiles with cumulative and exclusive profiles of transinformation. The results of the analyses will be presented and discussed in Chapter 4 and the consequences for the retina implant project will be outlined in Conclusion.

Guideline for the reader. This dissertation deals with a deeply inter-disciplinary field with many aspects. Some of them are dealt with in detail. Others have been omitted, not because they are considered irrelevant but simply in order to limit the extent of the work. There are numerous excellent works about the functional and physiological properties of the visual system (e.g., Nicholls et al., 1992; Kandel et al., 1995). Therefore, this thesis does without a corresponding introductory Chapter. To provide some aid for the reader interested in neurophysiology, a glossary that briefly explains important neurophysiological concepts and terms has been included at the end.

This dissertation comprises two Chapters (2 and 3) already accepted for publication (Eger and Eckhorn, 2002b,a) and one Chapter (4) that is planned to be submitted in similar form. The Chapters are closely related but may as well be read independently. Thus, repetitions could not always be avoided. A supplementary discussion and concepts of two developed devices necessary for the experiments, that did not find room in the respective Chapters were added to the Appendix A.

2 Three Steps for the Analysis of Neuronal Information Transmission in Multi-Input and -Output Systems

We present a new method to characterize multi-input and -output neuronal systems using information theory. Motivated by the development of retina implants we aim at an optimization of electrical stimuli with respect to high spatial, temporal, and intensity resolution. Informative activity in the primary visual cortex is a prerequisite for visual percepts. To obtain a lower bound of transinformation we take three steps: (1) Estimation of the deterministic response to isolate components carrying stimulus information. The deviation of the original response from the deterministic estimate is defined as noise. (2) Coordinate transformation using PCA yields an uncorrelated representation. (3) Partial transinformation values are calculated independently either by Shannon's formula assuming normality or based on density estimation for arbitrary distributions. We investigate the performance of the algorithms using simulated data and discuss suitable parameter settings. The approach allows to evaluate the degree to which stimulus features are encoded. Its potential is illustrated by analyses of neuronal activity in cat primary visual cortex evoked by electrical retina stimulation.

2.1 Introduction

Information theory has originally been developed for a quantitative description of technical communication channels disturbed by noise (Shannon, 1948). Later, it has also been widely applied in the field of neuroscience (e.g., MacKay and McCulloch, 1952; Werner and Mountcastle, 1965; Stein, 1967; Eckhorn and Pöpel, 1974, 1975; Eckhorn et al., 1976; Optican and Richmond, 1987; Bialek et al., 1991; Gershon et al., 1998; Rieke et al., 1998; Borst and Theunissen, 1999; Buracas and Albright, 1999; Wiener and Richmond, 1999; Reich et al., 2000; deCharms and Zador, 2000). It provides a general formalism for the quantification of the actual and potential information transfer across a black box communication channel. The amount of transinformation[1] T depends on the properties of the input signal, the channel's transfer

[1] We use the term *transinformation* instead of *mutual information* because in our case the investigated channel has a directed information flow.

characteristics and the noise affecting the information transmission. To maximize T an optimal code for the message to be transmitted is required but often difficult to find. In the special case of a continuous channel disturbed by additive Gaussian noise, the spectrum of the signal has to be adjusted to inversely match that of the noise ("water filling" concept in Shannon, 1949; Cover and Thomas, 1991). An analogous task is tackled here for the development of a retina implant. The aim is to quantify the information transmission via the retino-cortical pathway during electrical stimulation of retinal ganglion cells (Eger et al., 2001; Eger and Eckhorn, 2001a).[2] The quantification of the transmitted information between dynamic electrical stimulus patterns and the recorded cortical activity is a prerequisite for the choice of appropriate stimulation parameters.

Forward and backward approaches. There are two alternatives for the calculation of T as a direct consequence of Bayes' law of probabilities (Shannon, 1948)

$$\begin{aligned} T(S;R) &= H(R) - H(R|S) \\ &= H(S) - H(S|R)\,. \end{aligned} \tag{2.1}$$

In the *forward approach* the conditional entropy $H(R|S)$ (irrelevance) is subtracted from the output entropy $H(R)$. $H(R|S)$ describes the remaining uncertainty of output R if input S is known and has also been called "neuronal noise".[3,4] The forward approach is adopted when the output is examined conditional on the input. In the *backward approach* the perspective is changed and the input is viewed conditional on the output. Here, T is computed as the difference between the input entropy $H(S)$ and the conditional entropy $H(S|R)$ (equivocation).

Both approaches have been widely used to relate stimuli and their evoked responses. The forward approach being in control of the stimulus side takes the perspective of the information flow. It is characterized by questions of how a certain stimulus aspect is encoded in the response. On the other hand, the backward approach focuses on the response side and asks, which stimulus properties are responsible for certain response features. Its view is similar to that of a sensory system, which has no direct knowledge of the stimulus but tries to estimate it on the basis of evoked activity and stored knowledge. Examples of these contrary views are the peri-stimulus time histogram and the spike-triggered average of the stimulus. Both, the average and the variability of the signals have to be taken into account

[2] The retina implant is to benefit patients blind due to photoreceptor degeneration. By electrical stimulation of still intact retinal ganglion cells reliable and selective cortical activity patterns shall be evoked, a prerequisite for discriminable percepts. Animal experiments provide valuable data for the evaluation of the stimulation with respect to spatial and temporal resolution. At the present state of the project psychophysical experiments with humans are ethically not acceptable except for key studies (Humayun et al., 1999).

[3] We denote the input and output with S and R indicative for stimulus and response.

[4] If S and R are continuous random variables, H indicates differential instead of discrete entropy not affecting T (Cover and Thomas, 1991, p. 224).

in information theoretical analyses. For the quantification of T several methods are common in the field of neuroscience.

Direct calculation. In a direct approach input and output signals are discretized into sequences of input and output symbols s_i and r_j. T can be calculated directly after estimating the conditional probabilities $p(r_j|s_i)$ (forward) or $p(s_i|r_j)$ (backward) using (2.10). When the symbols are serially dependent in time the history[5] has to be taken into account and far more probabilities $p(r_{j,k}|s_{i,k}, s_{i,k-1}, \ldots, r_{j,k-1}, \ldots)$ (forward) or $p(s_{i,k}|s_{i,k-1}, \ldots, r_{j,k}, r_{j,k-1}, \ldots)$ (backward) have to be determined. Except for the requirement of stationarity the direct approach is not restricted. It yields precise results if the discretization is sufficiently fine and enough data is available for the estimation of the conditional probabilities. In practice, the latter is problematic. The number of conditional probabilities to be computed may easily reach astronomic dimensions if analog signals are discretized. It may be impossible to estimate them on the basis of relatively short data records.[6] Variants of the direct method have been successfully applied to different sensory systems (e.g., Eckhorn and Pöpel, 1975, 1981; Eckhorn et al., 1976; Eckhorn and Querfurth, 1985; de Ruyter van Steveninck et al., 1997; Buracas et al., 1998; Strong et al., 1998; Reinagel and Reid, 2000). The variance of mutual information estimators has been investigated by Moddemeijer (1989, 1999) for the direct approach.

Indirect calculation. For the reasons mentioned above, indirect methods based on a model of the noiseless information channel have often been preferred. The model provides a deterministic estimate of the response based on the stimulus or vice versa. Instead of quantifying T between stimulus and response directly, one relates the original signal (stimulus or response) and its deterministic estimate. The resulting value of T can never exceed the real value. If the channel model is perfect, a signal value merely depends on the value of its estimate at the same time. This drastically decreases the number of conditional probabilities that have to be considered. If the model is deficient and conditional probabilities between different times are neglected, T is downward biased. A straightforward approach to isolate signal and noise for the calculation of T consists in repetitive recordings of responses to identical stimulation epochs. It was introduced and applied to the peripheral visual system with either movement or luminance modulation of a light spot by Gaussian white noise (GWN) as input and spike trains of retinal ganglion cells as output (Eckhorn and Pöpel, 1981). Eckhorn and Pöpel calculated T on the basis of estimated conditional probabilities. To simplify the computation of the noise entropy, the noise may be assumed to be Gaussian distributed which requires considerably less data. In this context, de Ruyter van Steveninck (1996); Borst and Theunissen (1999) presented a guideline for computing an upper bound of transinformation on the basis of signal-to-noise ratios in the frequency domain by means of Shannon's

[5] indicated by index k

[6] E.g., for signals with only 8 bit amplitude resolution and serial correlation among 8 consecutive samples the number of conditional probabilities amounts to $(2^8)^{16} = 2^{128} \approx 3.4 \cdot 10^{38}$.

formula for the information capacity. However, they did not offer a solution for the problem of potential correlation between Fourier coefficients at different frequencies, e.g., due to nonlinearities of the channel and transients in the input signal. In such cases, the sum of partial transinformation values PT exceeds the real value of T, though the upper bound holds.[7]

An elegant backward approach that provides a lower bound of T in the case of stimulus evoked spike responses is the *stimulus reconstruction technique* (Bialek et al., 1991; Warland et al., 1997; Rieke et al., 1998). Here, a stimulus estimate is reconstructed from the response by means of reverse kernels. The kernels model the deterministic properties of the system and are the result of a squared error minimization between the stimulus and its reconstruction. Regarding the stimulus as "signal" and the deviation of the fit as "noise", PT values are calculated in the frequency domain based on individual signal-to-noise ratios. Contrary to the upper bound method described above, here the sum of PT does not exceed T as long as the stimulus is a stationary signal whose spectrum consists of Gaussian distributed independent Fourier coefficients. Any nonlinearities in the system or inaccuracy of the model would only increase the noise so that a lower bound results. Yet, when the stimulus frequency components are correlated T is biased upward. The stimulus reconstruction technique is restricted to a single channel stimulus whereas the upper bound method allows multiple input signals (Rieke et al., 1998; Buracas and Albright, 1999).

With respect to our research goals we need a method that unifies the advantages of the methods described above. We prefer a forward approach that provides a lower but relatively close bound of T, because we intend to conservatively assess the neuronal encoding of stimulus attributes. Further we demand its ability to handle systems with continuous multiple inputs and outputs and a broad variety of dynamic stimuli. For these reasons we developed a *three step approach* which is presented in the following sections. Each step is illustrated by analyses of visual cortex population activity evoked by electrical stimulation of the cat retina. We assess accuracy and limitations of the approach on the basis of simulated data for which the theoretical value of T is calculated analytically or numerically. Finally, we discuss the results and give rules of thumb for suitable parameter settings. A novel approach addressing some of the above demands was presented by Dimitrov and Miller (2001) shortly after submission of the present paper and will also be discussed. An outline of our method appeared as a conference abstract (Eger and Eckhorn, 2001b).

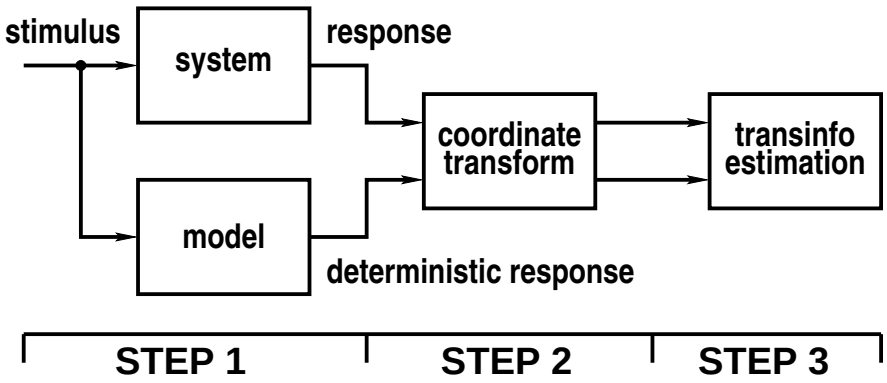

Figure 2.1: Scheme for the calculation of *T*. It consists of three steps: 1. Estimation of the deterministic system response on the basis of linear kernels. The convolution of the kernels with the stimulus provides an estimate for the deterministic fraction of the original response. 2. Coordinate transformation of original and deterministic response epochs by means of PCA provides uncorrelated coefficients. 3. Estimation of *T* as the sum of partial values *PT*.

2.2 Methods

We do not relate the stimulus and response signals directly but in a *first step* project the stimulus into the response space preserving only the relevant stimulus information. This projection is equivalent to estimating the deterministic response of the system given the stimulus. It is achieved by calculation of linear forward kernels which, convolved with the stimulus signal, provide an estimate for the deterministic fraction of the system response. We restrict our derivation to a linear model for computational reasons. The model may optionally consist of several components. For each stimulus condition to be included into the model one kernel is provided. The shape of the kernels is the result of a least squared error fit between the original and the deterministic response. In a *second step* the original and deterministic response are each divided into sequences of epochs long enough to preserve the signals' correlations. Then, we perform a linear coordinate transform of the epochs by means of principal component analysis (PCA) to yield a representation with decorrelated coefficients. In the *third step* we compute *PT* on a per-coefficient basis and summate the individual contributions to yield *T*. Fig. 2.1 illustrates the three step strategy for the calculation of *T*.

As an extension we modified the upper bound method (Borst and Theunissen, 1999) by implementing the PCA instead of the Fourier transformation. Since serial correlation in the input is taken into account, we get an upper bound which is closer to the real value. This approach corresponds to a modification of our *first step*, in

[7] With "partial transinformation" *PT* we here denote the contribution of one sample or frequency bin to *T*. According to the independence bound of entropy $\sum PT \geq T$ (Cover and Thomas, 1991).

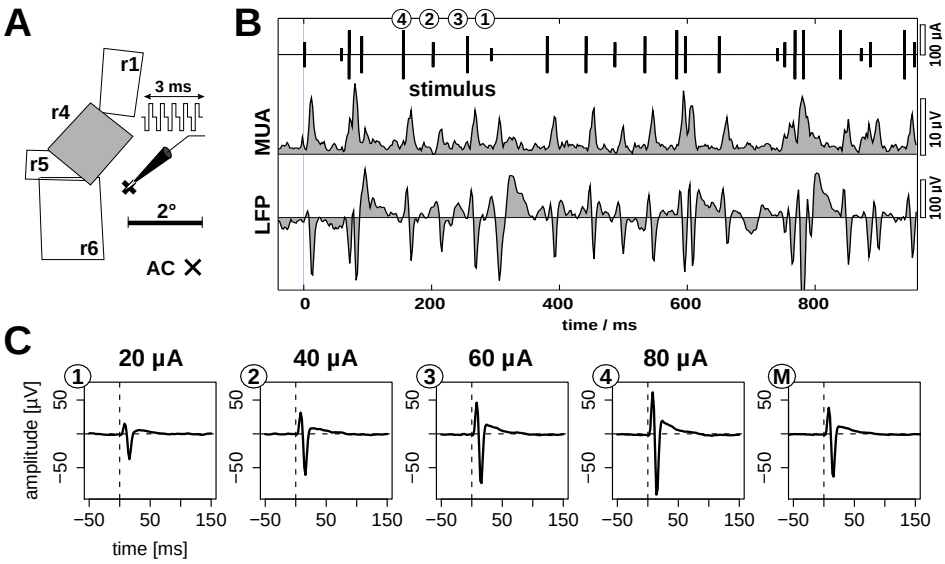

Figure 2.2: Example of local field potential (LFP) and multi unit activity (MUA) in visual cortex (V1) of anesthetized cat, in response to electrical stimulation of the retinal surface by a µ-electrode, 2° from the right area centralis (AC). **A**: Cortical receptive fields and stimulus position. The shaded field (r4) belongs to the electrode used for recordings shown in **B**: first 1000 ms of stimulus and response signals. Top: Stimulus sequence with Gamma distributed intervals (mean rate 20 imp/s). Each stimulus event was a burst consisting of 5 biphasic current impulses with 200 µs phase width as indicated in A. We distinguished between four stimulus conditions according to the current amplitude, that was randomly varied between the values 20, 40, 60, and 80 µA. For four stimulus events the stimulus conditions are symbolically depicted as encircled numbers at the very top. Middle and bottom: MUA and LFP response, respectively. **C**: Linear kernels for LFP corresponding to the 4 stimulus conditions and mean kernel (M, right).

which the linear model is replaced by the optimal model described by Roddey et al. (2000).

2.2.1 Step 1: Deterministic response model

The estimation of the deterministic response is based on linear kernels that can be understood as prototype response epochs. They characterize the system's behavior with respect to the corresponding stimulus attributes, aspects, or parameter settings, which we will denote as *conditions* (cf. Eger and Eckhorn, 2002a). The kernels are optimized with respect to the least squared error of the deviation between the response signal and the deterministic estimate.[8] When each stimulus condition is

[8]The least squared error seems to be an appropriate metric. It ensures that the deterministic estimate and the noise are uncorrelated. The implicit assumption of Gaussian noise distribution leads to an upper bound of the noise entropy and thus, assists a lower bound of transinformation.

regarded as a separate signal, the response estimate is the result of the superposition of the stimulus conditions convolved with the respective kernels. Optionally, specific stimulus conditions may be exclusively supported by the model. This is achieved by including corresponding kernels into the least squared error criterion affecting the number of degrees of freedom and also the amount of data needed for the kernel estimation. Consequently, the response estimate predicts only those response features that are due to the corresponding stimulus conditions. This allows us to investigate their encoding in the response. For example, one might assign one kernel per stimulus position, per different stimulus pattern, or per stimulus intensity step. If certain stimulus conditions are not considered in the model, the response estimate cannot reflect a potential encoding of the conditions; the model has to generalize over the conditional response variations. They are interpreted as noise because they cannot be associated with an appropriate stimulus condition known by the model. However, if indeed these conditions are not encoded in the response then the performance of the model is not decreased. This argument illustrates that the selective inclusion of certain stimulus conditions into the model may provide a quantitative measure of their encoding in the response. Fig. 2.2 illustrates the assignment of kernels to four different stimulus conditions in the context of electrically evoked neuronal responses. The stimulus conditions reflect four stimulation current amplitudes.

The kernel estimation procedure is performed for all kernels simultaneously. It is comparable to the approach given by Warland et al. (1997), who constructed multiple reverse kernels by linear regression in the time domain with the aim to reconstruct the visual stimulus based on multiple binned spike trains. Here, the analytical derivation is given in the frequency domain to avoid large equation systems. We follow a derivation similar to that presented by Rieke et al. (1998) but reversed and extended their method to multiple kernels (see also Bode and Shannon, 1950; Wiener, 1950).[9] Note, that the kernels are intentionally stimulus specific and are not suitable to generally describe the behavior of the system to an arbitrary stimulus (Palm and Pöpel, 1985).

Let $s_z(t)$ denote the stimulus conditions $z \in \{1,\ldots,m\}$ which each have been assigned a corresponding kernel $h_z(t)$ that we want to estimate. Then the linear response estimate is a superposition of the stimulus conditions convolved with the kernels

$$\rho(t) = \sum_{z=1}^{m} s_z(t) * h_z(t) \; . \tag{2.2}$$

[9] The recent publication by Theunissen et al. (2001) that appeared shortly after submission of our manuscript takes a similar approach. The authors model the neuronal response to natural or artificial spatio-temporal stimuli by linear kernels and implicitly address the problem of decorrelation.

We want to minimize the mean squared error between the response and its deterministic estimate

$$\chi^2[\{h_z(t)\}] = \left\langle \int_{-\infty}^{+\infty} \left| r(t) - \sum_{z=1}^{m} s_z(t) * h_z(t) \right|^2 dt \right\rangle \tag{2.3}$$

by adjusting the kernels $h_z(t)$. The minimization is to be performed in the frequency domain. According to Parseval's theorem we have

$$\begin{aligned} \chi^2[\{h_z(t)\}] &= \int_{-\infty}^{+\infty} \left\langle \left| R(j\omega) - \sum_{z=1}^{m} S_z(j\omega) H_z(j\omega) \right|^2 \right\rangle \frac{d\omega}{2\pi} \\ &= \int_{-\infty}^{+\infty} \left\langle |R|^2 \right\rangle \frac{d\omega}{2\pi} \\ &\quad - \sum_{z=1}^{m} \int_{-\infty}^{+\infty} H_z^* \left\langle R S_z^* \right\rangle \frac{d\omega}{2\pi} \\ &\quad - \sum_{z=1}^{m} \int_{-\infty}^{+\infty} H_z \left\langle R^* S_z \right\rangle \frac{d\omega}{2\pi} \\ &\quad + \sum_{x=1}^{m} \sum_{y=1}^{m} \int_{-\infty}^{+\infty} H_x H_y^* \left\langle S_x S_y^* \right\rangle \frac{d\omega}{2\pi} \,. \end{aligned} \tag{2.4}$$

For clarity reasons we have dropped the argument $j\omega$. After taking the derivatives separately for each frequency with respect to the real and imaginary parts of the kernels H_z and setting them to zero,[10] one finally gets an array of m equations, each of the form

$$\left\langle R S_\zeta^* \right\rangle = \sum_{z=1}^{m} H_z \left\langle S_z S_\zeta^* \right\rangle \,. \tag{2.5}$$

The left side of the equations resembles the conditional mean response following the stimulus ζ, the right side the superposition of the frequency kernels weighted with the respective stimulus cross spectra. With the abbreviations

$$\begin{aligned} F_\zeta &= \left\langle R S_\zeta^* \right\rangle \\ N_{z,\zeta} &= \left\langle S_z S_\zeta^* \right\rangle \end{aligned} \tag{2.6}$$

(2.5) can be rewritten as

$$F_\zeta = \sum_{z=1}^{m} N_{z,\zeta} H_z \tag{2.7}$$

or alternatively, in matrix style

$$\vec{F} = \mathcal{N} \vec{H} \,. \tag{2.8}$$

[10] The Hesse matrix containing the second derivatives is positive definite which is sufficient to indicate a minimum of χ^2.

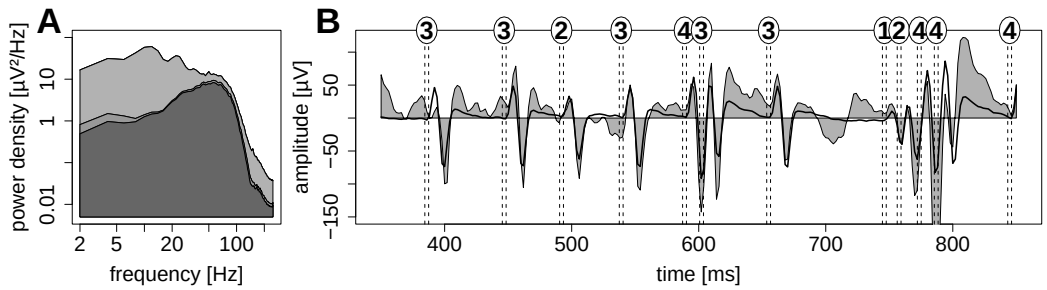

Figure 2.3: Comparison of the single and multi-kernel response estimates corresponding to Fig. 2.2. **A**: Power spectral densities of the original response (light grey), the deterministic estimate using multiple kernels (medium grey), and using a single kernel (dark grey). Note that the power of the estimates is generally less than that of the original response. The better the prediction the more response variance is explained by it. Here, the multi kernel prediction explains slightly more variance than the single kernel estimate. **B**: Representative epoch over 500 ms of the original response (light grey), and the multi kernel estimate (solid line). For clarity, the single kernel signal has been omitted. Dashed vertical lines mark the time and duration of the stimuli, the encircled numbers indicate the corresponding stimulus conditions (Fig. 2.2).

The solution of the system of equations can be carried out separately for each frequency component by means of any standard equation solver that handles the complex data type. The dimension of the square matrix $\mathscr{N}$ is identical with the number of kernels and is very small compared to the corresponding regression matrix proposed by Warland et al. (1997). The algorithm was implemented in R language (Ihaka and Gentleman, 1996).

An example for the deterministic response estimation is shown in Fig. 2.3. The estimates capture the linear deterministic fraction of the stimulus information. There is no linear correlation between response and estimate other than the amount of correlation inherent in the response itself.[11]

2.2.2 Step 2: Coordinate transformation

If the response of the investigated system is a white noise signal, i.e., the samples are statistically independent, T can be calculated as the sum of its partial contributions PT. If the response signal is not white, the sum of PT will exceed the real value of T (Cover and Thomas, 1991). In general, independence of samples at the output of a system cannot be assumed. Serial correlation may already be present in the input or is introduced by the signal transmission properties, e.g., lowpass characteristics, of the information channel. For this reason, the stimulus reconstruction technique is performed in the frequency domain: in spite of serial correlation in the time domain

[11] The "kernel" between the original response and the deterministic estimate is formally a delta impulse.

the frequency components of Gaussian random functions are independent (Rieke et al., 1998). However, the Fourier coefficients of epochs of stationary processes are only approximately uncorrelated (Unser, 1984) and non-stationary processes usually carry correlation both in time and frequency domain.

To solve this problem we have to perform a transformation into a coordinate system where the representation of the data has independent coordinates. Shannon (1948) has shown that, although the entropy may be affected by a change of coordinates, the difference of entropies such as T remains unchanged. It should be noted that the transformation has to be bijective or in the linear case, invertible (Deco and Obradovic, 1997). The above requirements are met by the *principal component analysis* (PCA). It decorrelates the data by finding an orthonormal coordinate system whose axes are the principal axes of a Gaussian ellipsoid fitted to the data.[12] Thus, it is very useful for redundancy reduction. Moreover, the expansion is optimal in the sense that a reconstruction of the original data based on a reduced set of coordinates yields a least squared error fit.

In practice, we proceed as follows. The data is divided into N subsequent epochs of n samples each. Then we construct the principal components (the basic waveforms) from the set of deterministic response epochs. As long as the number of epochs exceeds the number of samples per epoch, the space is large enough to also embed the original response. However, its covariance matrix may not be perfectly diagonalized (see Discussion on p. 30). When PCA is applied to the data, each of the N epochs $\vec{R}_{\text{time}}$ is represented by a new vector of coordinates in the principal component space

$$\vec{R}_{\text{pca}} = \frac{1}{n}\mathscr{F}\vec{R}_{\text{time}} \tag{2.9}$$

with $\mathscr{F}$ being a matrix whose rows are the principal components. There are two alternative ways to calculate the principal components $\mathscr{F}$, the regular approach and the Karhunen-Loeve expansion (Glaser and Ruchkin, 1976). While the regular PCA diagonalizes an $N \times N$ covariance matrix where each element is the covariance between two epochs of $\vec{R}_{\text{time}}$,[13] the Karhunen-Loeve transform uses the $n \times n$ covariance between time points. Depending on the dimensions N and n, the method with the smaller covariance matrix is computationally more efficient. In our analyses we normally have $N > n$. Therefore, we applied the Karhunen-Loeve transform, but determined the eigenvalues with singular value decomposition for reasons of numerical precision (see Documentation of `prcomp()`, Ihaka and Gentleman (1996)).

As stated above, the transformation into principal component coordinates leads to uncorrelated coefficients. This property may be utilized for the application of the

[12] There are other techniques such as *independent component analysis* (Comon, 1994; Karhunen et al., 1995) which aim at statistical independence. However, they require a larger amount of data and are computationally more complex.

[13] Alternatively, the correlation matrix with normalized coefficients may be used (see Discussion in Glaser and Ruchkin, 1976).

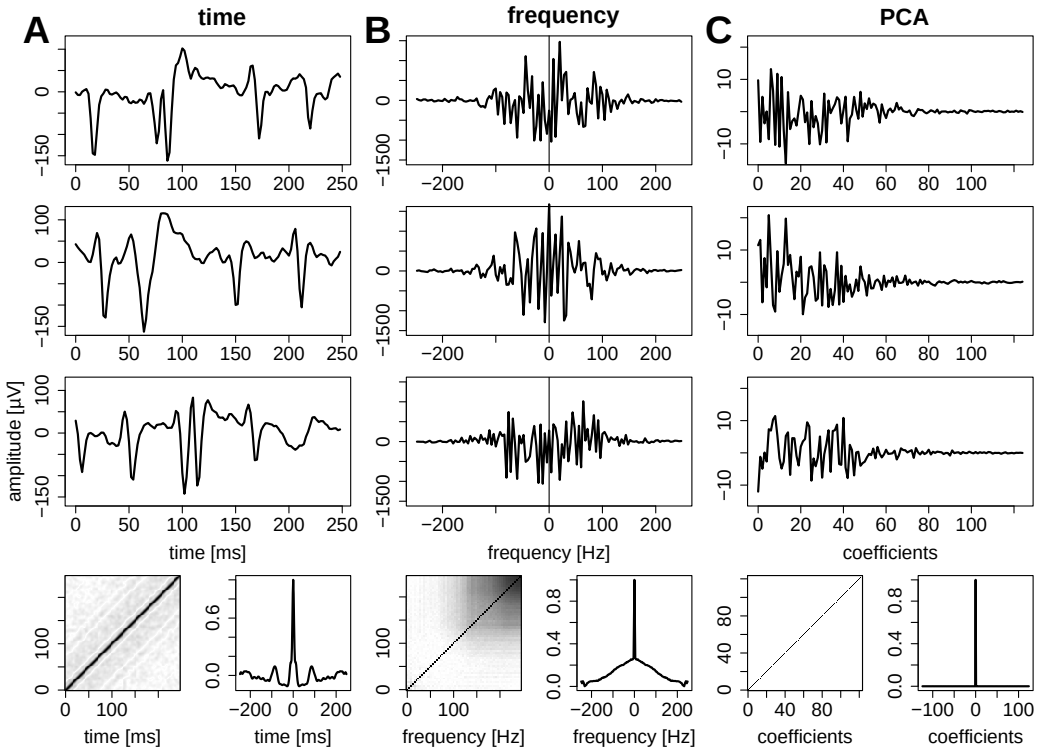

Figure 2.4: Response signal in three different domains. **A**: time, **B**: frequency, and **C**: PCA domain. Upper panels, A: Three representative 250 ms epochs of the same cortical LFP shown in Fig. 2.2. In B the imaginary part of the Fourier coefficients is shown for negative, the real part for positive frequencies. Bottom, left panels: Respective correlation matrices. Each pixel corresponds to the complex correlation coefficient between a pair of bins. Absolute correlation is coded with linear grey scale (white: 0; black: 1). Bottom, right panels: Diagonal mean of the correlation matrix (normalized auto-correlation function). In contrast to the PCA domain, in the time and frequency domains some degree of linear correlation is present. This is indicated by non-zero off-diagonal elements of the correlation matrices and deviation from delta shape of the auto-correlation functions.

three step approach to multi-output systems. In practice this may be achieved as follows. According to the previous section, one deterministic response model for each of the k output channels has to be constructed. Then, epochs of corresponding times have to be serially aligned separately for the original and model response signals. This leads to new combined epochs that are k times longer than the original ones.[14] Now the covariance matrix has to be orthogonalized as described above, resulting in the basic vectors of the combined vector space and the new coordinates: the principal component representation of the combined epochs.

[14]It should be considered to normalize the k output channels to obtain equal variance before aligning the epochs if each channel is equally important. The fact that the alignment of epochs of different channels may cause unsteady transitions does not cause problems.

Fig. 2.4 gives an example of subsequent epochs of data recorded in cat visual cortex. The data is represented in three different coordinate systems: time (A), frequency (B), and principal component domain (C).[15] The depicted correlation matrices correspond to those used in the Karhunen-Loeve transform. Other than in the time and frequency representations, the principal component coefficients are uncorrelated and even statistically independent if deviation from normality and nonlinear correlation in the data is negligible (see Discussion on p. 31). Then the sum of their partial contributions PT does not exceed T.

2.2.3 Step 3: Transinformation based on signal-to-noise ratios

After the continuous stimulus and response signals have been sufficiently finely discretized in time and amplitude, the transinformation can be calculated directly (in bit per sample)

$$T(S;R) = \sum_{i=1}^{m} p(s_i) \sum_{j=1}^{n} p(r_j|s_i) \log_2 \frac{p(r_j|s_i)}{p(r_j)} \; . \tag{2.10}$$

However, the estimation of the conditional probabilities $p(r_j|s_i)$ typically requires a huge amount of stationary data that often cannot be provided by neurophysiological experiments, in particular, when the symbols are serially correlated. The calculation of T is simplified if the stimulus is a Gaussian white noise signal (GWN) with variance σ_s^2. On the assumption of additive channel noise with variance σ_n^2 a lower bound of T (in bit per sample) is given by Shannon (1948)

$$T(S;R) \geq \frac{1}{2} \log_2 [1 + \mathrm{SNR}] \; , \tag{2.11}$$

with the signal-to-noise ratio $\mathrm{SNR} = \sigma_s^2/\sigma_n^2$. Equality is reached if the channel noise is GWN. For band limited signals with Gaussian distributed Fourier coefficients, Shannon (1949, p. 169) has shown that the transinformation (in bit per second) yields

$$T(S;R) = \frac{1}{2} \int_{-W}^{W} \log_2 \left[1 + \frac{|S(f)|^2}{|N(f)|^2} \right] df \; . \tag{2.12}$$

However, we do not want to be constrained to stimuli with independent Fourier coefficients. Only in this case the PT may be added in the Frequency domain to yield T according to (2.12). We have proposed to solve this problem by implementing the PCA instead of the Fourier transform (see previous Section, p. 16). The entropy formulae are equivalent in both cases. Concerning the nomenclature, we will use the character f that – depending on the context – may either denote frequency or principal component coordinates. Further, we use the deterministic model $P(f)$

[15] We applied the discrete Fourier transform (DFT) without zero-padding in order not to introduce correlation between Fourier coefficients.

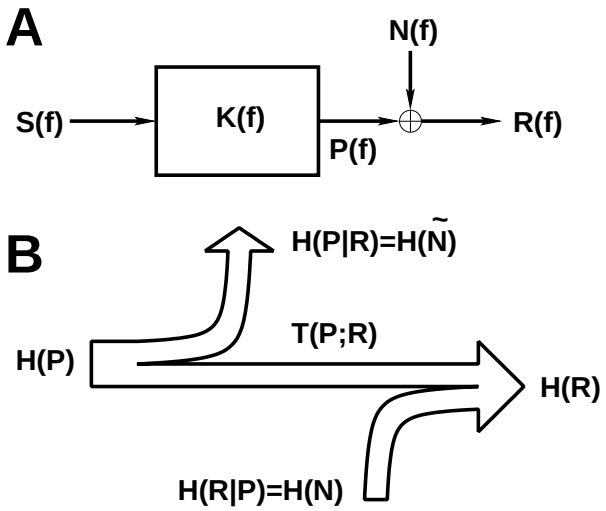

Figure 2.5: Schematic diagrams of the signal (**A**) and entropy flows (**B**) for the calculation of T.

instead of the stimulus $S(f)$ for the calculation of T (see Fig. 2.5). This is necessary because stimulus and neuronal response signals are different signal types and usually are represented in different spaces. Since the model can never outperform the stimulus we have

$$\begin{aligned} T(S;R) &\geq T(P;R) \\ &= H(R) - H(N)\,. \end{aligned} \tag{2.13}$$

Equality holds, when the model perfectly captures the deterministic properties of the stimulus and the noise is additive and independent. We now require that the model signal $P(f)$ is multivariate Gaussian distributed and decorrelated. The latter condition is fulfilled after coordinate transformation based on PCA. Then the differential entropy of $P(f)$ is

$$H(P) = \frac{1}{2}\int_W \log_2 2\pi e|P(f)|^2 df\,. \tag{2.14}$$

The response $R(f)$ and the noise $N(f) = R(f) - P(f)$ are represented in the same coordinate system and may neither be Gaussian nor completely decorrelated. Their entropies are bounded by the Gaussian case, e.g., for the noise entropy we have

$$H(N) \leq \frac{1}{2}\int_W \log_2 2\pi e|N(f)|^2 df\,. \tag{2.15}$$

Since the entropy power of arbitrary noise is less than or equal to its variance and noise correlations decrease the entropy, a lower bound of T results (in bit per second, Shannon, 1948)

$$T(S;R) \geq T(P;R) \geq \frac{1}{2}\int_W \log_2 \left[1 + \frac{|P(f)|^2}{|N(f)|^2}\right] df\,. \tag{2.16}$$

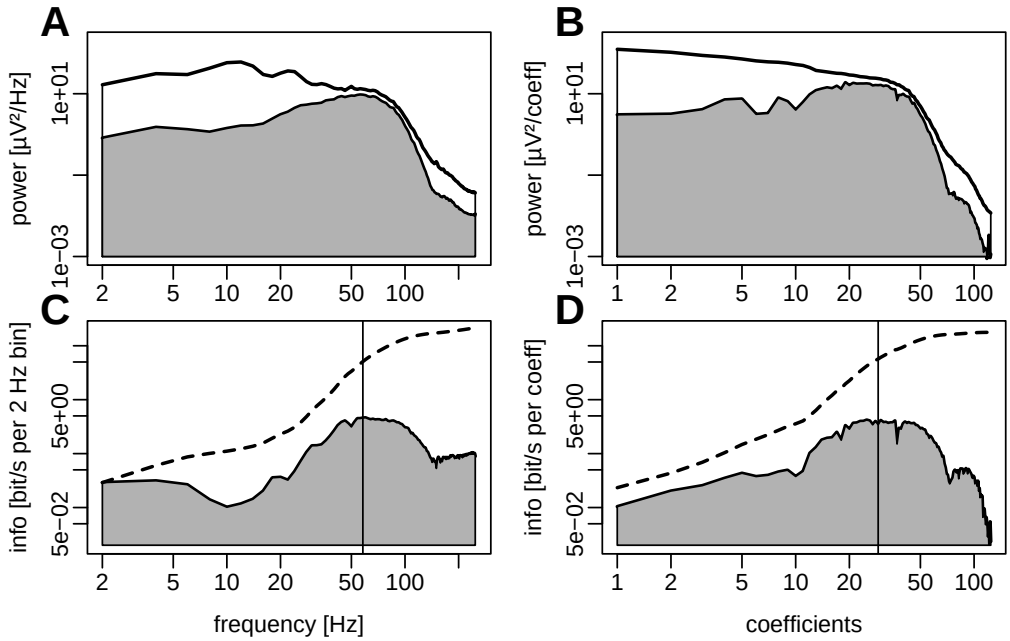

Figure 2.6: Example for the calculation of T in the spectral (left) and PCA domain (right) on the basis of the data shown in Figures 2.2 and 2.3. **A**, **B**: Power densities of original response (unfilled) and multi-kernel estimate (shaded). **C**, **D**: Rates of PT (shaded) and their cumulative sum (dashed) based on the signal-to-noise ratios for each frequency component (**C**) and principal component coefficient (**D**). Only those contributions have been considered that were significantly different from zero (jackknife error estimation, $p \leq 0.05$). The cumulative transinformation in the frequency domain is higher compared to that in the principal component domain ($T_{\text{freq}} = 215.9$ versus $T_{\text{pca}} = 177.3$ bit/s) due to correlation in the spectral representation. Fig. 2.7 exemplifies how the PT are computationally obtained.

The response variance amounts to $|R(f)|^2 = |P(f)|^2 + |N(f)|^2$ since $P(f)$ and $N(f)$ are uncorrelated. The latter may easily be proved in analogy to the derivation by Theunissen (1993) (see Appendix 2.5.9). Thus, T may also be given with respect to the *effective* noise $N_{\text{eff}}(f)$

$$T(S;R) \geq \frac{1}{2}\int_W \log_2\left[1 + \frac{|R(f)|^2}{|N_{\text{eff}}(f)|^2}\right] df \tag{2.17}$$

(Bialek et al., 1991; Theunissen, 1993). Contrary to the noise source in Fig. 2.5 A, the fictitious effective noise may be thought of being additive with respect to the response $R(f)$ (see Appendix 2.5.8).

The above approach has been applied to the data example shown in Figures 2.2 and 2.3, both in the frequency and PCA domains. Fig. 2.6 depicts the power spectra of the original response and its deterministic estimate as well as PT. We generally applied the jackknife error estimation and bias correction method for the calculation

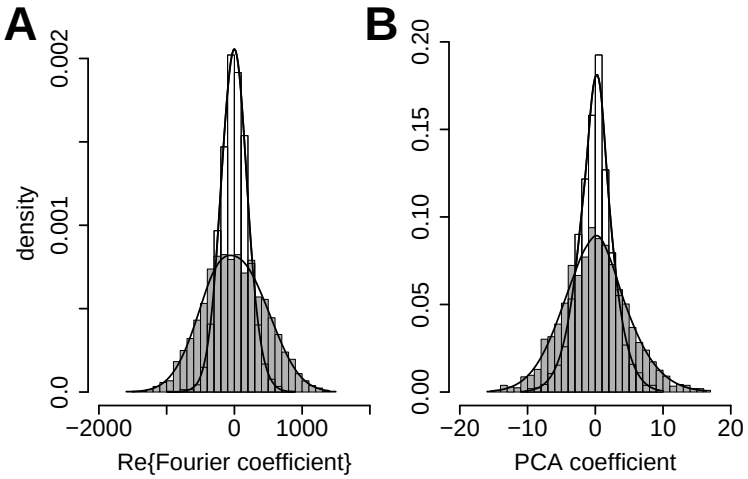

Figure 2.7: Histograms and density estimates in frequency and PCA domain of signal (shaded) and noise components (unfilled) for one selected bin indicated by corresponding vertical lines in Fig. 2.6. **A**: Histograms of the real parts of the Fourier coefficients. The imaginary part is similarly distributed (not shown). **B**: Histograms of the principal component coefficients. In the frequency domain *PT* yields 4.581 bit/s per 2 Hz at $f = 58$ Hz using density estimation vs. 4.585 bit/s per 2 Hz by means of (2.17) (applied to a single bin) based on the signal-to-noise ratio. Similarly, in the PCA domain *PT*s of the 29^{th} coefficient are 3.462 vs. 3.446 bit/s per coefficient, respectively. The calculation based on the signal-to-noise ratio is exact if noise and response are Gaussian distributed and independent.

of *PT* (Efron and Tibshirani, 1991; Efron, 1994). Fig. 2.7 illustrates how a single *PT* value is calculated. As a point of reference we calculated *PT* based on a computationally expensive adaptive density estimation algorithm (see documentation of `locfit()`, Loader (1997, 1999) and Appendix 2.5.5).

2.3 Results

We demonstrate the accuracy and reliability as well as the limitations of the method on the basis of artificial data that have controllable statistical properties and allow the direct calculation of *T*. It is shown to what degree assumptions implicitly made and structural properties of the data affect *T*. As displayed in Fig. 2.8 we selected different types of artificial signals resembling stationary (A, B, C), cyclo-stationary (D, E) and non-stationary processes (F). They cover a broad range of possible input signal and channel properties. The tests are independent of the performances of the response model in order not to influence the results by arbitrary model properties. The generated artificial data represent both, a simulated stimulus evoked response and a corresponding response model with 1 ms resolution. In practice, the deterministic response model data was generated first. In a second step the model was

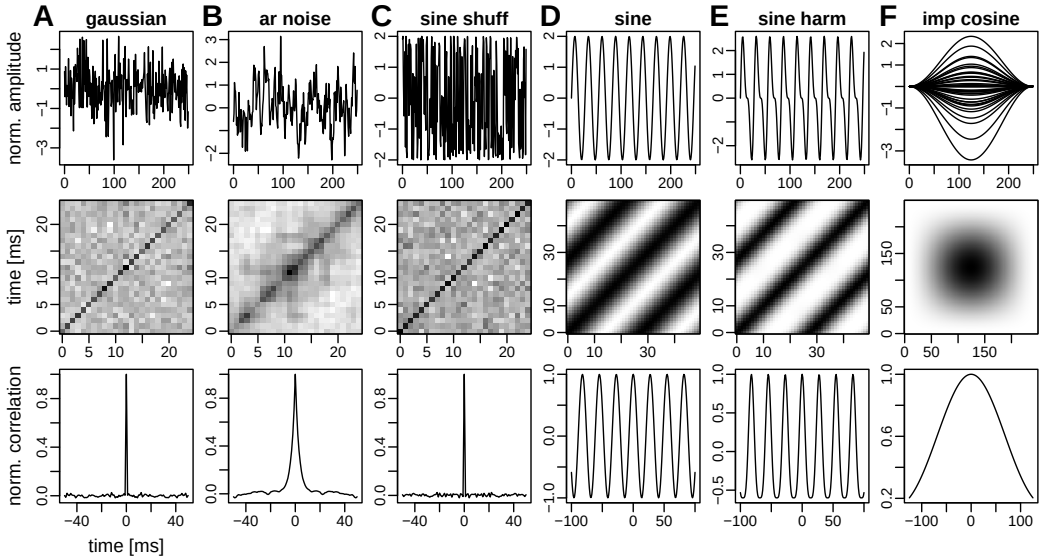

Figure 2.8: Example epochs of the six model signals (upper row), sections of their covariance matrices (middle), and auto-correlograms (bottom). **A**: Gaussian white noise, (GWN), **B**: AR process, **C**: white noise with amplitude distribution of a sinusoid, **D**: sinusoidal epochs at random phase, **E**: sinusoidal epochs with random phase and first harmonic, and **F**: epoch aligned impulse responses to delta impulses with Gaussian distributed amplitude.

superimposed by independent unit variant GWN to yield the simulated response signal. The following model signals were tested:

A *Gaussian white noise.* The model consists of a sequence of N epochs of length n. The samples are independent and Gaussian distributed. Noise is adjusted to yield $S/N = 1$ which leads to $T = 0.5 \log_2 [1 + S/N]\,\text{bit} \cdot 1000\ \text{s}^{-1} = 500$ bit/s.

B *Autoregressive (AR) process.* Different from A the model contains serial correlation while the average signal power is the same. Serial correlation is introduced according to the lowpass properties of the first order AR process. The theoretical value is $T = 385.2$ bit/s (Appendix 2.5.1).

C *Noise with amplitude distribution of a sinusoid.* The signal differs from A only in its bimodal and range limited amplitude distribution with double variance. We get a theoretical value of $T = 764.2$ bit/s (Appendix 2.5.2).

D *Phase jittered epochs of sinusoids.* The amplitude distribution is identical with that in C but the signal is strongly correlated. The amplitude $A = \sigma^2 = 2$ and frequency ($\approx$9 periods per epoch) is fixed. The only variable is the uniformly distributed phase. T does not depend on the frequency but on the

epoch length (Appendix 2.5.3). For a length of 250 ms a value of $T = 20.4$ bit/s is transmitted.[16]

E *Phase jittered epochs of sinusoids with first harmonic.* The signal differs from D in the additional first harmonic of half amplitude. The entropy of the model is the same as in D, since the phase locked 1st harmonic does not carry additional phase information. However, the signal-to-noise ratio is increased. For an epoch length of 250 ms we get $T = 36.9$ bit/s, a value larger than in D (Appendix 2.5.3).

F *Epoch aligned random amplitude cosine waveform.* The test signal mimics the response of a filter with impulse response $g(k) = 0.5\,(1 - \cos\,(2\pi k/N))$ to a random amplitude delta impulse at epoch begin. The entropy is identical to that of the delta impulse. The filter introduces serial correlation which leads to an improved signal-to-noise ratio. A value of $T = 13.1$ bit/s results for 250 ms epoch length (Appendix 2.5.4).

Dependence of transinformation on coordinate system. We compare the resulting values of partial (*PT*) and cumulative transinformation (*CPT*) after linearly transforming the signals into the frequency and alternatively, PCA domain. Fig. 2.9 depicts the respective profiles. The dashed lines indicate the theoretical values of *T*. They should coincide with the corresponding asymptotic values of *CPT*. The calculation in frequency coordinates is precise for signals whose Fourier representation yields independent coefficients. This is the case for the stationary signals A, B, and C. For the non-stationary signal F and, surprisingly, for the cyclo-stationary signals D, E the calculation fails and the sum of *PT* overestimates *T*. In contrast, the calculation in the PCA domain yields relatively precise results for all six signal types. It is nearly precise for signals C and F and gives a slight underestimation for A and B. The relative deviations from the theoretical values in case of signals D and E amount to about 50%. Even though the absolute deviations are small compared to those resulting from the calculation in the frequency domain, they are not negligible and may indicate a systematic influence.

Table 2.3 summarizes the results. The values of *PT* were calculated based on individual signal-to-noise ratios using Shannon's formula for information capacity before summing them up, except for the two columns on the right. Here, normality was not assumed. Instead, we performed density estimation for the signal and noise (Loader, 1997) and computed *PT* as the difference of the partial response and noise entropies (Appendix 2.5.5).

Dependence of transinformation on data size. The number of epochs of the model signals was varied in steps from 5 to 1000. In Fig. 2.10 *T* is plotted over

[16] For computational reasons, we generated a continuous signal with odd ratio of periods per epoch. Thus, the phase was not really random. Comparing the results to those of sinusoid epochs with uniform random phase we did not find any difference.

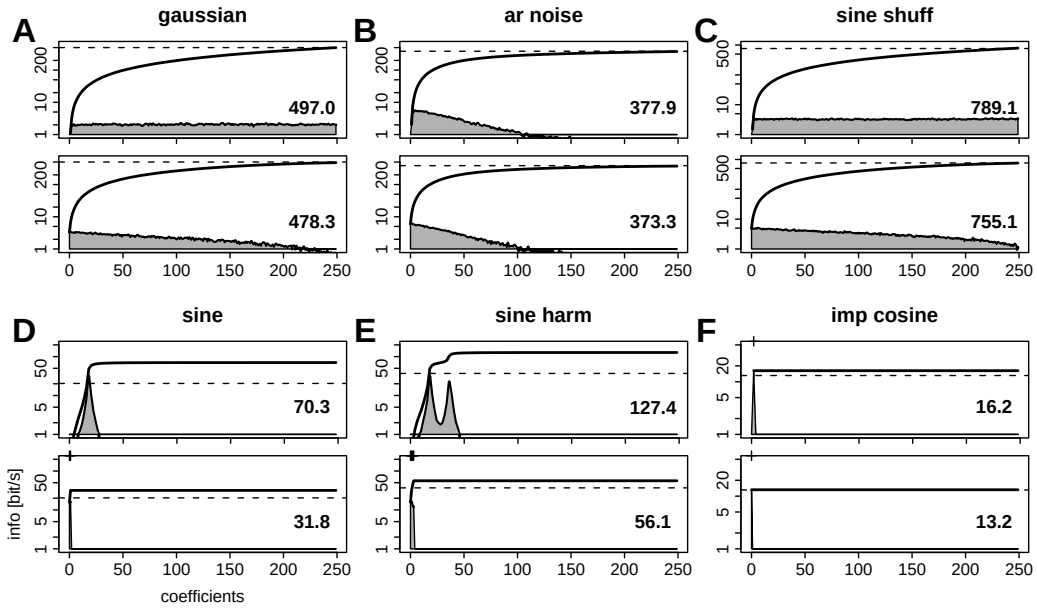

Figure 2.9: *PT* (shaded) and *CPT* (unfilled) profiles for the artificial signals A–F according to Fig. 2.8 in the spectral (upper panels) and PCA (lower panels) domain. The data consisted of 1000 epochs of 250 ms duration and 1 ms resolution. The x-axes denote the coefficients either in the frequency (in Hz) or the PCA domain. *PT* is based on individual signal-to-noise ratios. Corresponding values on the basis of density estimation are visually not distinguishable. The theoretical values of T are indicated by dashed horizontal lines. To calculate *CPT* we performed a jackknife error estimation (Efron and Tibshirani, 1991; Efron, 1994) to obtain an estimate of the standard error, and rejected any *PT* not significantly different from zero. Significant bins ($p \leq 0.05$) are marked at the top line of PCA panels in D, E and both panels in F. In the other panels most bins are significant and therefore not marked. The total values of T (in bit/s) for the frequency and PCA domain are in the respective diagrams. See also Table 2.3.

the epoch number for the frequency and PCA domain and for three different epoch durations (50, 100, 250 ms). T was normalized to its theoretical value in order to better compare the results for the different parameter settings. Note that for signals D, E, and F the number of signal periods per epoch was kept constant while altering the epoch size.

For signals A, B, C the transinformation converges to the theoretical value at different slopes (except for a small asymptotic deviation in case of signal C). For small amounts of data T is underestimated remarkably. The calculation in the frequency domain converges faster, i.e., a smaller number of epochs is needed compared to the calculation in the PCA domain. In the frequency domain the different epoch sizes hardly affect the convergence of T. This is contrary to the PCA domain where smaller epoch sizes are to be preferred. Comparably, fewer epochs are necessary

Table 2.1: Theoretical and numerical values of T (in bit/s) for the six model signals of Figures 2.8 and 2.9.

bit/s	theor	frequency		PCA S/N		PCA density	
model signal		uncorr	corr	uncorr	corr	uncorr	corr
A gaussian	500.0	497.8	497.0	479.0	478.3	479.2	479.2
B ar noise	385.2	378.6	377.9	374.0	373.3	373.9	358.7
C sine shuff	764.2	789.8	789.1	755.8	755.1	755.4	755.4
D sine	20.4	71.9	70.3	31.8	31.8	33.5	29.2
E sine harm	36.9	129.0	127.4	56.1	56.1	54.0	50.8
F imp cosine	13.1	16.9	16.2	13.9	13.2	14.1	13.2

The corrected values differ from the uncorrected ones by the rejection of those PT not significantly different from zero ($p \leq 0.05$). Note that in most cases the discrepancy between corrected and uncorrected T is small. There are exceptions for the calculation in the PCA domain using density estimation.

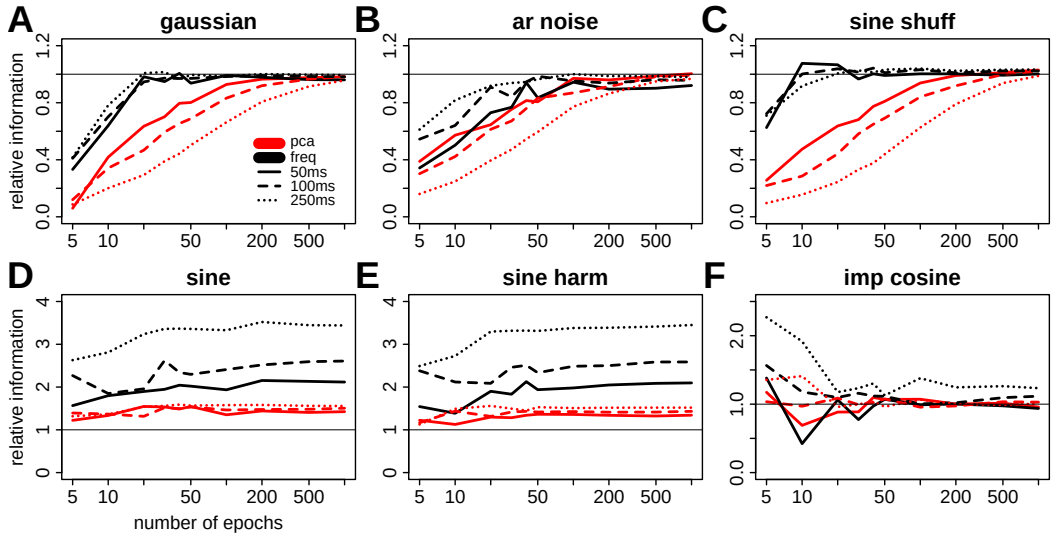

Figure 2.10: Dependence of corrected values of T on the amount of available data for the six test signals (Fig. 2.8). The transinformation was normalized to the theoretical values. Graphs are shown for three different epoch sizes (50, 100, 250 ms) calculated in the frequency and PCA domains (inset A).

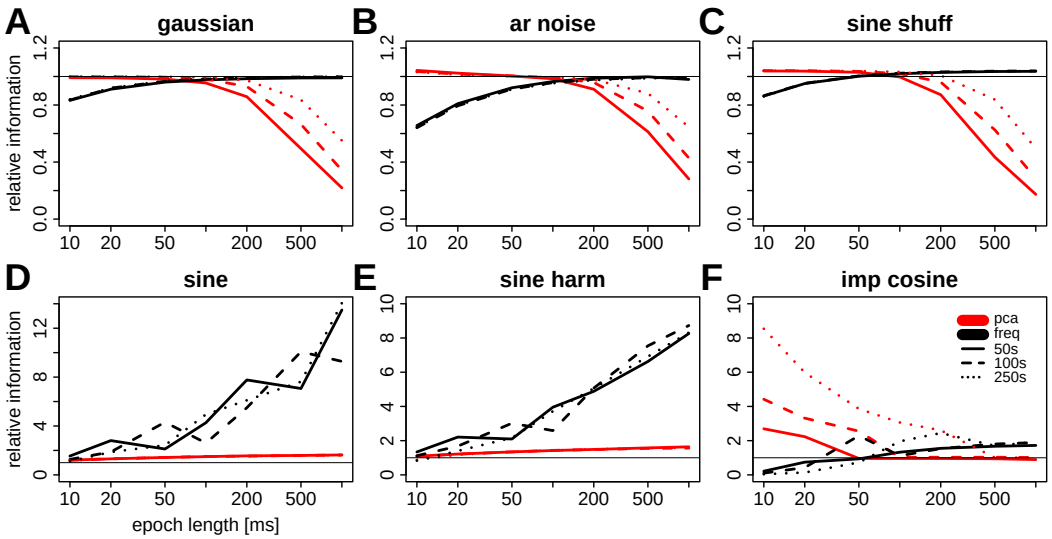

Figure 2.11: Corrected transinformation values normalized to the theoretical values dependent on the epoch duration, while the total data size was kept fixed (test signals in Fig. 2.8). We generated three data sets of 50, 100, and 250 seconds length. They are identical with the largest data sets in Fig. 2.10 (number of epochs N=1000). Note that the basic frequencies for the signals D, E, and F depend on the data size (see legend). In practice, the three data sets per signal were divided into epochs of 10 to 1000 ms and for each epoch duration T was calculated in the frequency and PCA domains.

to yield the same proximity to the theoretical value, even though the total amount of data proportionally decreases with reduced epoch sizes. For signals D, E, and F[17] the calculation in the frequency domain overestimates T considerably. In contrast, PCA is relatively precise even for small epoch numbers and for all three epoch sizes. We will discuss the minor but consistent deviation from the theoretical value in case of signals D and E in Discussion on p. 31.

Dependence of transinformation on epoch size. We examined the dependency of T on the epoch length, i.e., the number of samples per epoch, while keeping the total data size constant. This is relevant with respect to the practical question of how to divide data already available in order to yield values of T that are closest to the theoretical value. We used the data sets examined in Fig. 2.9, lined up the epochs and divided them anew into epochs of varied length (10 to 1000 ms). The periodicity of signals D, E, F is differently affected depending on the new epoch size. For signals D and E the number of periods per epoch hardly influences T. The theoretical values only depend on the number of epoch samples (Appendix 2.5.3 and Rife and Boorstyn (1974)). Signal F is heavily affected if the new epoch size is not an integer multiple of the waveform length. In this case the entropy of the signal increases due to the added uncertainty about the phase. This effect is evident

[17] largest epoch size

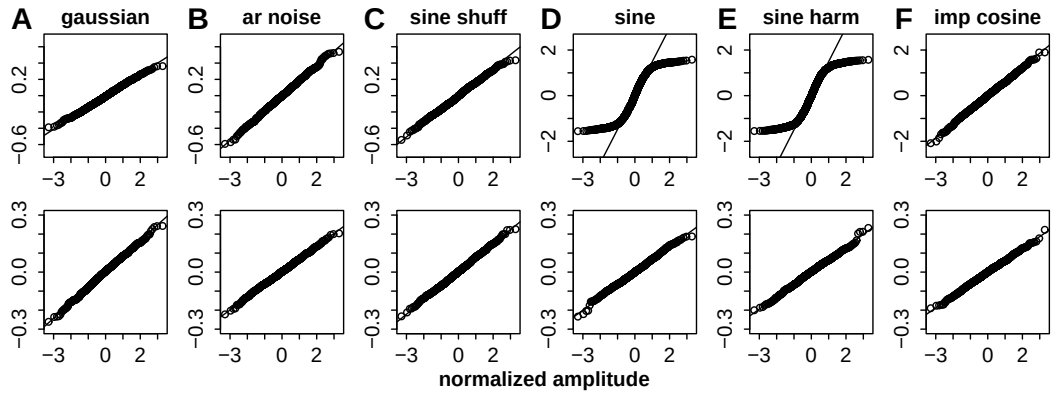

Figure 2.12: Normal probability plots for the first bin (PCA domain) of the signal (top) and noise (bottom) for the six test signals. The axes denote their normalized amplitudes. The shapes are representative for the distributions of the other bins. For signals D and E only the first 2, respectively, 4 bins are significant (jackknife error estimation, $p \leq 0.05$).

in Fig. 2.11 F where the red graphs originate from a large bias yet join the horizontal line (the theoretical value) at an epoch size equal or larger than the waveform length. The reason for this large deviation is not a fault of the algorithm. The theoretical value was calculated on the (violated) assumption of a constant phase and hence, is too small, here.

The graphs for signals A, B, C in Fig. 2.11 look similar, and for both coordinate systems T converges to the theoretical value if one neglects the small deviation for signal C. While the calculation in the frequency domain gives best results for longer epochs, small epoch sizes are preferred for the calculation in the PCA domain. Another difference is that for the PCA domain the results depend strongly on the total data size (note the different position of the bend of the graphs). For the largest data size (250 s) a maximum epoch length of 200 ms seems suitable. For smaller total data sizes shorter epoch lengths should be chosen to prevent a downward bias of T. Signals D and E fail with respect to the frequency method. Only for very small epoch sizes the values of T get relatively close to the theoretical value. In comparison, the calculation in the PCA domain yields values that are much closer to the real value, though there exists a consistent deviation. Interestingly, the slope of the curves is positive contrary to the graphs A, B, C. T of signal F is estimated very precisely in the PCA domain. As mentioned above, the large deviations for short epochs are due to the neglected increase of entropy in the calculation of the theoretical value.

Assumption of Gaussian distribution. Implicitly, normality is assumed when the signal-to-noise ratio is used to calculate PT. Entropy $H(P)$ and consequently, T will be overestimated if the assumption of Gaussian distributed principal component coefficients of the model is violated because a Gaussian distribution maximizes the entropy given the variance. On the other hand, if the noise deviates from normality

the irrelevance is overestimated which leads to a downward bias for T. To test the validity of the assumption of normality, we plotted the empirical quantile function against the normal quantile function for the first coefficients (normal probability plot, Fig. 2.12). Except for signals D and E the graphs resemble a straight line for signal and noise indicating normality (Snedecor and Cochran, 1989). In the other cases the noise is still normal though the normal probability plots of the signal have sigmoid shapes, which reveals bounded distributions.

The results given in Table 2.3 (columns PCA) are based on two different methods for the calculation of the PT. The first method assumes normality because it uses the signal-to-noise ratio. The second method is based on a density estimation procedure (Appendix 2.5.5 and Loader (1997)). In case of approximately Gaussian distributed variables (signals A, B, C, F) T nearly yields the same values with both methods. An exception is the corrected value in signal B which most probably is due to a threshold effect of the significance test. However, when the assumption of normality is violated (D, E) we find a small deviation between both calculations. T is slightly overestimated using the information capacity formula if a significance test has been performed (corrected columns).

The small asymptotic bias for signal C both with the PCA and frequency method (Fig. 2.10, 2.11 C) may have its origin in the non-Gaussian properties of the model although the principal component coefficients are normal (Fig. 2.12 C). This being our conjecture we calculated T additionally in the time domain (Figure not shown) without assuming normality using the density estimation method. Interestingly, the result is $T = 764$ bit/s which is nearly identical to the theoretical value (Table 2.3).

2.4 Discussion

Summary of methods. Having chosen an indirect forward approach for the quantification of T, we adopted the perspective of the sensory information flow and modeled the channel output conditional on its input for several reasons. First, in our experiments we are in control of the stimulation site which allows us to easily inspect the conditional mean responses of the system. Second, we wish to evaluate stimulation parameters, attributes and conditions with respect to their contribution to T. This is equivalent to analyzing the degree to what these stimulus properties are encoded in the evoked responses.

To determine T in case of continuous stimulation we carry out a three steps strategy. After designing a response model that incorporates the encoding of selected stimulus properties we obtain a deterministic estimate of the original response. In a second step we perform a coordinate transformation by means of PCA and yield linear independent coordinates. The last step consists in the calculation of PT on a per coefficient basis by means of either a density estimation algorithm or simply Shannon's information capacity formula.

Coordinate transformation. On the basis of artificial data we demonstrated that our approach is suited to yield a lower bound of T provided that nonlinear correlations between samples of the deterministic response estimate and deviation from normality in the PCA domain are negligible. The calculation of T as the sum of partial values fails in any coordinate system if the coordinates are dependent (Cover and Thomas, 1991). For the same reasons Rieke et al. (1995, 1998) – within the context of the stimulus reconstruction technique – emphasize the use of Gaussian white or colored noise stimuli which do not contain correlated frequency components. Correlated harmonics introduced by potential nonlinearity of the examined channel, are still interpreted as noise and do not violate their lower bound demand, as long as the model does not capture them. The reason is the introduced disorder of frequency bins between original and model signal which is not taken into account by a linear algorithm.

A data structure (epoch length and number) dependent side effect of the same origin appears when the coordinate transformation is performed by means of PCA. Equivalently to the frequency method that provides a diagonal covariance matrix of the stimulus (Rieke et al., 1998), here, PCA transforms the deterministic model data and diagonalizes its covariance matrix (Glaser and Ruchkin, 1976). However, the same transformation applied to the original response signal may not completely diagonalize its covariance since the correlation structure of the inherent noise may differ from that of the deterministic model. If now T is calculated solely based on the diagonal elements of the model and noise covariance matrices,[18] the noise entropy is overestimated which assists a lower bound of T (see Hadamard's inequality, Cover and Thomas (1991)).

We find this effect in Fig. 2.10 and 2.11 (stationary signals A, B, C) where T converges to the theoretical value for both the calculation in the frequency and PCA domain. However, the calculation in the PCA domain underestimates the theoretical value to a larger degree and converges more slowly. A crucial difference between Fourier and principal component analysis is that for the latter the basic waveforms are generated from the analyzed data themselves.[19] In order to reflect relevant generalized features of the data instead of casual deflections it is necessary to provide a sufficient number of epochs for the construction of the principal components. The ratio of the epoch number N vs. epoch length n is to be chosen between two extremes: $N = n$ is just sufficient to provide a vector space large enough to cover any possible signal. Only $N/n \to \infty$ guarantees a perfect generalization.

Since $N/n = 2$ showed a downward bias of T of less than 10% (Fig. 2.10 A, B, C), we applied the rule of thumb $N/n > 2$ for the parameter settings in our analyses of real data, i.e., we adjusted the number n of samples per epoch to at most half of the epoch number N. An advantage of smaller epochs at fixed total data sizes

[18] equivalent to (2.16)

[19] The eigenvectors of the covariance matrix of a stationary signal are not identical but for large n close to the Fourier basis (Unser, 1984; Gray, 2001).

is that the precision of the variance or density estimates increases with growing number of epochs. On the other hand, the epoch size should not be chosen too small. This is equivalent to disregarding data inherent correlation, which extends further than one epoch and would lead to an upward bias of T. This effect can be seen in Fig. 2.11 B (red curves at 10 or 20 ms) where the epoch size was decreased to a degree that it reaches the proximity of the time constant of the autoregressive process (Appendix 2.5.1).

We have shown that by means of transformation into the PCA domain we may relatively well deal with cyclo-stationary or non-stationary model signals that comparably fail if T is calculated in the frequency domain (Figures 2.10, 2.11 D, E, F). The precise result for signal F is astonishing though logical when one realizes that the PCA achieves complete linear decorrelation. The Fourier coefficients of cyclo-stationary signals (D, E) are correlated depending on the epoch length n. This is due to spectral leakage caused by discontinuities of the periodically continued signal when an integer multiple of the fundamental period does not exactly fit into the epoch. The signals D and E are examples for which the PCA performs much better than the frequency method. However, a 50% overestimation of the theoretical value is not satisfactory. An explanation is that PCA does not achieve statistically independent coefficients if nonlinear correlations are present in the non-transformed signals. Moreover, deviation from normality of the resulting principal component coefficients indicates that PCA probably did not work optimally (Deco and Obradovic, 1997). As shown in Appendix 2.5.6, the correlation between subsequent temporal samples are not linear but rather circular due to systematic phase shifts of the sinusoids (Fig. 2.15 E). This kind of correlation cannot be completely removed by means of PCA. Note that in Fig. 2.9 D, E only two or four principal component coefficients are needed, respectively, to describe the epochs (note markers at the top). Looking more closely into the data reveals that the principal components are simply a sine and cosine waveform for signal D and two pairs of them for signal E. This illustrates the proximity to the Fourier transformation (Makhoul, 1981; Unser, 1984; Gray, 2001). Clearly, any shifted sinusoid may be described by a linear combination of a sine and cosine waveform demanding for two coefficients. However, even though there is no residual linear correlation between these coefficients, they are affected by a strong nonlinear correlation and a scatter plot resembles a circle (Fig. 2.15 F). This may explain why the values of T for signals D and E are overestimated.

A lower bound of transinformation. What can be done to not overestimate T using the PCA approach? As pointed out above, nonlinear correlations between samples cannot be resolved and result in an upward bias. One example for temporal nonlinear correlations are oscillatory components. Other causes may be nonlinear amplitude characteristics of the system such as threshold effects (cf. Priestley, 1980). Nonlinear correlations may be detected visually by means of scatter diagrams of the examined variables (Fig. 2.13 A). However, the pairwise consideration

of variables is not sufficient to exclude nonlinearity. Concepts for testing linearity based on higher order spectra are available (Rao and Gabr, 1980; Hinich, 1982). If nonlinearities cannot be neglected, nonlinear decorrelation concepts such as independent component analysis (ICA, cf. Comon (1994); Karhunen et al. (1995); Deco and Obradovic (1997)) should be considered.

A further cause for an upward bias is an unjustified assumption of normality. A deviation from normality results in an overestimation of the respective entropy. Normality may be visually tested on the basis of quantile-quantile plots (Fig. 2.12 and Snedecor and Cochran (1989); Venables and Ripley (1999)). This argument may serve for the explanation of both an upward and a downward bias of T since it is the difference of the signal and noise entropy. For example, if normality of the noise is erroneously assumed, its entropy will be overestimated leading to a decreased value of T.

An interesting observation is the small asymptotic bias for signal C both with the PCA and frequency method (Figures 2.10, 2.11 C) although the coefficients are normal (Fig. 2.12 C). A recalculation in the time domain without assuming normality using the density estimation method yields the nearly exact value of 764 bit/s. We explain this effect as follows: The application of PCA yields orthogonal coefficients, but it cannot be guaranteed that the result is statistically independent. Especially, if the original variables already are statistically independent it is most probable that the result is only linearly uncorrelated, i.e., nonlinear correlations may be introduced by the PCA. At the same time the linear transformation affects the distributions of the variables. The originally bimodal distributions of signal C are changed towards approximate normality (central limit theorem, see Snedecor and Cochran (1989)). If now PT is summed up to estimate T, the nonlinear correlations implicitly introduced by the PCA are not taken into account. The entropies of the individual variables are increased due to the change in their distribution, which results in an upward bias for T.

PCA has been applied to obtain a dimensional reduction of recorded data (Richmond et al., 1987). Linsker (1988) showed that if a network performs a PCA of its Gaussian distributed inputs, it maximizes T across the net. This illustrates that there is no other transformation that better preserves information in the Gaussian case. Accordingly, Plumbley (1991) pointed out that PCA leads to the tightest upper bound for information loss introduced by the net. The bound being an upper one results from the input's potential deviation from normality. These findings encourage the use of PCA in case of normally distributed variables but one has to be careful if this cannot be ensured. For a separation of a nonlinear mixture of independent variables with arbitrary distributions other approaches utilizing higher order statistics have to be considered (Deco and Obradovic, 1997; Yang et al., 1998).

The choice of a response model. The transinformation yielded by an indirect method depends on the utilized model. The less exact it predicts the deterministic response of the system, the lower is T because response features not explained

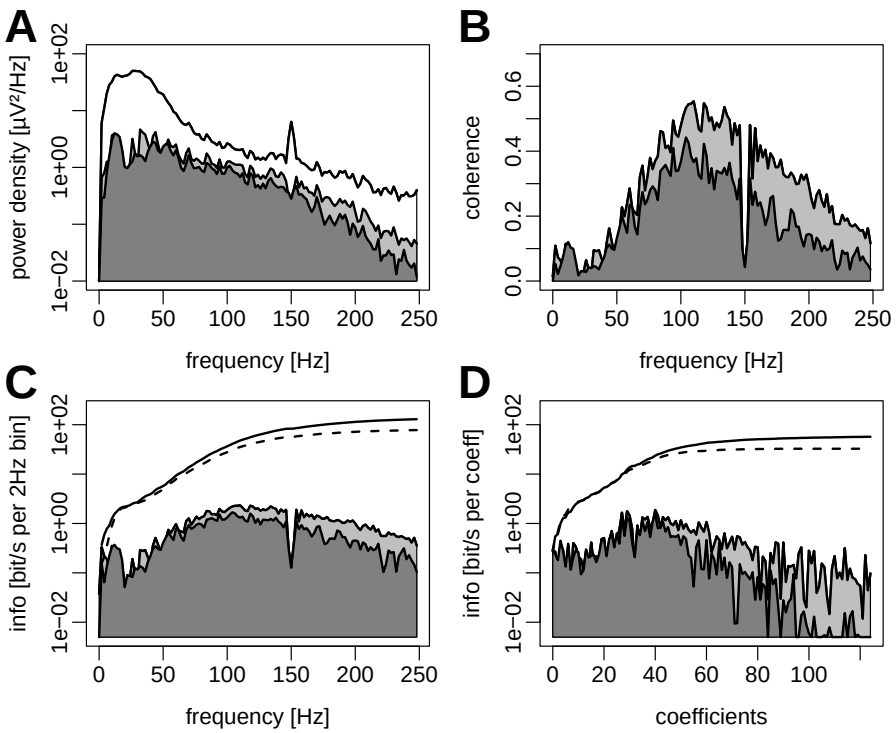

Figure 2.13: Example of cortical local field potentials evoked by a four channel electrical stimulation of the retina (Gamma distributed inter-stimulus intervals, mean rate 20 s^{-1} per channel, 100 µA, 400 µs charge balanced current impulses with positive leading phase). **A**: Response power density (unfilled) and power densities for the linear response model (dark grey) and the optimal model according to Roddey et al. (2000) (light grey). **B**: Coherence measures $\gamma(R,P)$ (dark grey) and $\sqrt{\gamma(R,\tilde{R})}$ (light grey). **C** and **D**: *PT* and *CPT* for the spectral respectively PCA domain (linear model: dark grey, dashed line; optimal model: light grey, solid line). The peak in A and the corresponding cut in B, C at 150 Hz are caused by the 2nd harmonic of 50 Hz mains interference.

by the model are automatically interpreted as extra noise. In their recent approach, Roddey et al. (2000) demonstrate that the performance of a response model may be assessed in relation to an optimal model. Optimality is defined with respect to the least squared error between the original response and the model signal. To get an impression about the performance of a utilized linear model, and to illustratively discuss the consequences of the model choice, we applied their approach to data recorded from the cat visual cortex (Fig. 2.13). Diagram A shows the spectral power densities for the recorded response R, for a deterministic linear response model P, and the optimal model P_{opt}, which is identical to the conditional mean response. The model power spectra are relatively similar, though not very close to the original response spectrum. This indicates that the original response contains a large amount of variability not related to the stimulus. Diagram B displays the coherence measures $\gamma(R,P)$ and $\sqrt{\gamma(R,\tilde{R})}$. R and $\tilde{R}$ denote two long duration response sections

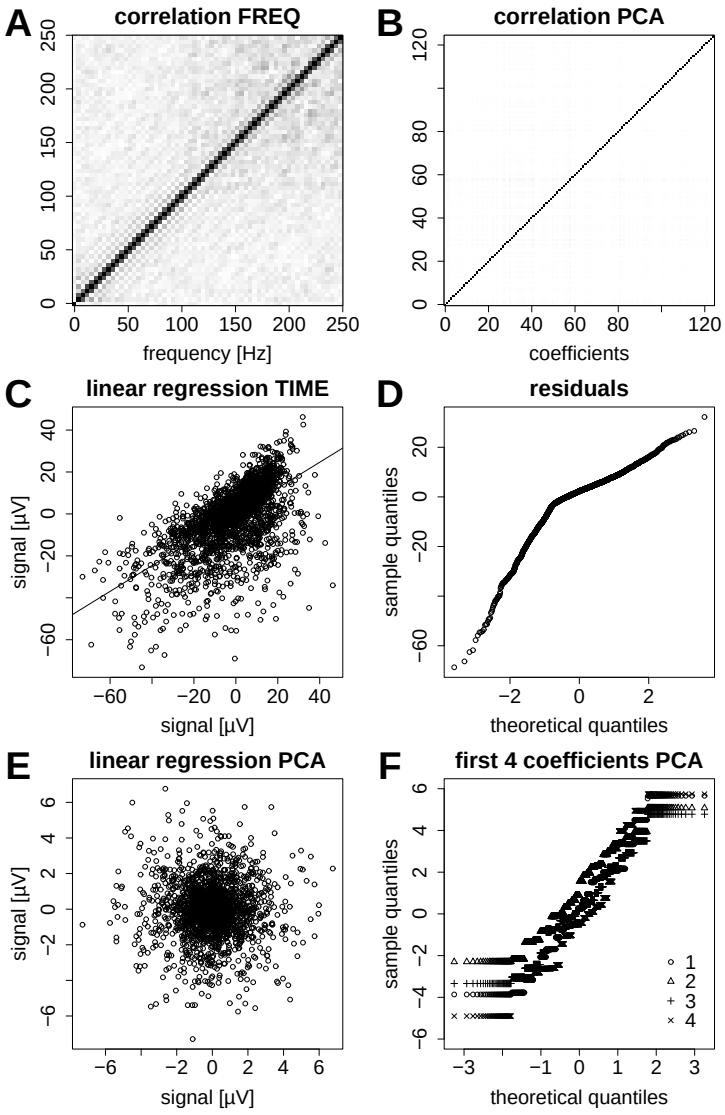

Figure 2.14: Exemplary analysis for the applicability of PCA. **A**, **B**: Correlation matrices for frequency and PCA domain, respectively. Absolute correlation is coded with linear grey scale (white: 0; black: 1). **C**: Scatter plot of consecutive samples of P_{opt} in time domain reveals linear correlation. **D**: Normal probability plot of residuals indicates that the data may not be described sufficiently by a linear model. **E**: Scatter plot in the PCA domain, does not suggest any correlation. **F**: Normal probability plots of the first 4 principal component coefficients roughly indicate normality. The curves are structured and range-limited due to the deterministic model properties and stimulus repetitions.

evoked by an identical stimulus. Roddey et al. (2000) have shown that the performance of the model P is reflected in the proximity of both functions. One would conclude that for the present example a linear model might be sufficient because the optimal model does not explain much more variance compared to the linear model (A). The graphs Fig. 2.13 C, D show PT and CPT for the linear and optimal model calculated in the frequency and PCA domain. The total rate of $T_{\text{freq}} = 78.4$ (linear) respective 129.3 bit/s (optimal) in C exceeds the values $T_{\text{pca}} = 33.4$ (linear) respective 57.0 bit/s (optimal) in D by more than 50%. The reason is revealed by the correlation matrix of the Fourier coefficients (Fig. 2.14 A) revealing correlation between frequency components. In contrast, the correlation matrix of the principal component coefficients (Fig. 2.14 B) is diagonal indicating an absence of linear correlation. The disregarded correlation between frequency components leads to an unnecessary overestimation of T. It is possible that nonlinear correlations between samples in the time domain not assessed by the PCA are not negligible or the principal component coefficients are not Gaussian distributed. As exemplified with artificial data, then also the PCA approach would overestimate the real value of T. Still, the prerequisites for its applicability can be tested.

The scatter plot of consecutive samples of the model signal in the temporal domain (Fig. 2.14 C) may expose a nonlinear relation. Yet, even an evident linear relation does not exclude nonlinearity (cf. Hinich, 1982; Rao and Gabr, 1980). In our case, an F-test indicates a very significant linear correlation ($p \leq 1^{-15}$) but at the same time the regression yields a distribution of residuals that clearly deviates from normality. The quantile-quantile plot (Fig. 2.14 D) is not straight indicating that a linear statistical model is not sufficient. Hence, PCA does not capture all relevant correlation. It should be noted that it is important to test the properties of the optimal response model P_{opt} instead of the original response because a potential nonlinearity might be obscured by the response variability. The scatter plot for the PCA domain (E) does not reveal any correlation. The normality of the principal component coefficients is visually tested, again with a quantile-quantile plot (Fig. 2.14 F).

To summarize, any weakness of the response model will always lead to a decrease of T because a filter introduces entropy loss (Shannon, 1948). The inclusion of nonlinear properties into the model is mandatory for the analysis of chiefly nonlinear systems when a larger downward bias is to be avoided. Especially the analysis of stimulus evoked action potential responses may require a nonlinear model. Thus far, we have neglected the estimation of higher order kernels in order to reduce the computational intensity. A practical approach for the design of nonlinear models based on higher order correlation kernels is given by Lee and Schetzen (1965). They used GWN as input signal because it is optimal for determining the transfer properties of nonlinear time invariant systems (Wiener, 1958). The design and assessment of a two-input nonlinear model of an insect nervous system by means of nonlinear forward and cross kernels has been presented by Marmarelis and McCann

(1973). When the encoding of certain stimulus aspects is to be investigated, signal types other than GWN might be preferred (Krausz, 1974). Finally, Rieke et al. (1998) note that a nonlinear encoding of stimulus properties might not exclude linear decodability.

Applicability of different methods. The presently known indirect methods for the quantification of T have certain restrictions. The *stimulus reconstruction technique* suffers from the limitation to single channel Gaussian noise stimuli (Rieke et al., 1998) although it has been extended to multiple outputs resembling spike trains of several neurons (Warland et al., 1997). Theoretically the method could also apply to time series data such as graded potentials. However, we have not found a hint in the literature. The stimulus reconstruction technique always yields a lower bound of T independent of the model, as long as the frequency components of the stimulus are Gaussian distributed and independent.

The *upper bound method* (de Ruyter van Steveninck, 1996; Borst and Theunissen, 1999) is not restricted to Gaussian single channel stimuli and applies to any kind of response data. However, a requirement is the normality of the noise, in order not to overestimate the irrelevance. When the stimulus contains correlated frequency components or the investigated system bears strong nonlinear properties, the upper bound may be unnecessarily far from the real value. Here, we suggest to replace the Fourier transformation by the PCA (see example, Fig. 2.13). The epoch size should be chosen small enough to obey the upper bound demand. In the proposed formula for an upper bound of the conditional entropy (p. 110, D.1 in Roddey et al. (2000)) in the Fourier space the variance could be replaced by the determinant of the covariance matrix to yield a closer upper bound (compare Shannon, 1948, p. 37). An approach related to the upper bound method has been recently presented (de Ruyter van Steveninck et al., 1997; Strong et al., 1998) and is widely used (Brenner et al., 2000a; Buracas and Albright, 1999; Reinagel and Reid, 2000). It is restricted to point process responses (trains of action potentials) elicited by repeated stimulation with an arbitrary time dependent stimulus and provides either an upper or lower bound of T depending on the chosen entropy estimates.

The novel approach by Dimitrov and Miller (2001) published shortly after our submission is theoretically not constrained to certain stimulus or response configurations. The main idea is related to step 2 of our approach: by introducing equivalence classes using an optimal quantization of stimulus and response the dimensionality of the data is reduced and redundancy is removed without losing relevant information. An advantage is that it yields the largest lower bound of T given a limited data set. In practice, e.g., when analyzing multi-channel time series data, the minimization of the distortion measure bears some problems, especially for complex input stimuli. As a consequence, they restrict their maximum entropy model to 2^{nd} order statistics (Dimitrov et al., 2002) – which is equivalent to PCA. Moreover, for continuous time series responses the quantization does not provide a clear criterion for limiting the number of classes (Dimitrov and Miller, 2001).

The *three step method* we have presented bears advantages in several aspects. It is (1) not restricted to stationary single channel stimuli, (2) it is easily applicable to multi-output systems due to the use of PCA, (3) the calculation of multiple kernels in the frequency domain is computationally less costly than in the time domain, (4) it provides a lower bound of T at least if the signal inherent correlation is linear and the samples are Gaussian distributed, and finally (5), due to its forward perspective and adjustable model properties it facilitates the investigation of the encoding of stimulus aspects.

Conclusion. Our motivation for the development of the presented method was threefold. First, for the analysis of graded potentials evoked by multi-channel non-Gaussian continuous stimulation, a feasible method for the quantification of T did not exist especially with regard to the constraints of relatively small amounts of stationary data. Since for the analysis of action potential responses well known and approved methods are available, a comparison of T based on different response signal types is now possible and might be helpful to investigate their coding efficiency. Second, the three step method allows to analyze potential redundancy of cortical activity and calculate T between different recording positions. This is possible because it is not restricted to signals with uncorrelated frequency components. Third, the separation of the procedure into three steps is advisable for indirect methods because the estimation of T is affected by three separate factors: the model performance, the coordinate system, and the precision of the entropy estimation. Finally, we want to encourage the use of test data as objective benchmark for implemented algorithms.

2.5 Appendix

2.5.1 Transinformation for AR noise signal and additive GWN

Let the channel input be described as an AR process

$$s_k = \mu\xi_k + \lambda s_{k-1} \tag{2.18}$$

comprised of an independent GWN component ξ and its own history. The output of the additive noise channel be

$$r_k = s_k + n_k = \mu\xi_k + \lambda s_{k-1} + n_k \tag{2.19}$$

with GWN n being independent of ξ and having the same variance. Let $S(j\omega)$ and $X(j\omega)$ denote the Fourier spectra of the sequences s_k and ξ_k, respectively.[20]

[20]The discrete time Fourier transform is defined as $F(j\omega) = \sum_{k=-\infty}^{\infty} f_k\, e^{-j\omega k}$ (Oppenheim and Schafer, 1989).

Table 2.2: Output entropy and transinformation dependent on epoch size n for phase shifted sinusoids and additive GWN in bit per epoch. Upper rows: Values for single sinusoid. Lower rows: Values for sinusoid with phase locked first order harmonic.

n	10	20	50	100	200	250	500	1000
$H(R)$	2.744	2.717	2.693	2.681	2.670	2.672	2.663	2.659
T	2.858	3.331	3.968	4.456	4.947	5.108	5.598	6.095
$H(R_1)+H(R_2)$	5.581	5.500	5.427	5.391	5.362	5.361	5.340	5.328
T	4.809	5.727	6.977	7.941	8.912	9.232	10.212	11.200

Then $S(j\omega)e^{-j\omega k}$ corresponds to the shifted sequence s_{k-1}. We may now formally replace the sequences by their respective spectra in (2.18) and find

$$F(j\omega) = \frac{S(j\omega)}{X(j\omega)} = \frac{\mu}{1-\lambda e^{-j\omega}} \,. \tag{2.20}$$

The transinformation between input s and output r may be calculated as

$$\begin{aligned} T &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_2\left[1+\frac{|S|^2}{|N|^2}\right] d\omega \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_2\left[1+|F(j\omega)|^2\right] d\omega \end{aligned} \tag{2.21}$$

since n and ξ are equivariant. Numerical integration with $\mu = \lambda = 1/\sqrt{2}$ yields the transinformation $T = 0.386$ bit per sample. The time constant of the process is revealed when the impulse response of $F(j\omega)$ is considered. Therefore, we rewrite (2.20) as geometric sequence

$$F(j\omega) = \mu\left(1+\lambda e^{-j\omega}+\lambda^2 e^{-j2\omega}+\lambda^3 e^{-j3\omega}+\cdots\right) \tag{2.22}$$

and immediately find

$$f_k = \mu\lambda^k = \mu e^{k\ln\lambda} = \mu e^{-\frac{k}{-1/\ln\lambda}} \tag{2.23}$$

for $k \geq 0$. The time constant amounts to $\tau = -1/\ln\lambda = 2.89$ samples.

2.5.2 Transinformation for noise signal with distribution of a sinusoid and additive GWN

The density of the signal $s(t)$ is the derivative of its cumulative density which is given by

$$P_s(s_o) = \lim_{T\to\infty;\ \Delta t\to 0} \frac{1}{T}\sum_{T} \Delta t_{s(t)<s_o} \tag{2.24}$$

with duration T and time intervals Δt. We drop index 'o' and for $s(t) = A\sin(t)$ we have

$$P_s(s) = \begin{cases} 0 & \text{if } s < -A \\ \frac{1}{2} + \frac{1}{\pi}\arcsin(\frac{s}{A}) & \text{if } -A \le s \le A \\ 1 & \text{if } s > A \end{cases} \tag{2.25}$$

and the density $p_s(s) = dP_s(s)/ds$

$$p_s(s) = \begin{cases} 0 & \text{if } s < -A \\ \frac{1}{A\pi\sqrt{1-(s/A)^2}} & \text{if } -A < s < A \\ 0 & \text{if } s > A \end{cases} \tag{2.26}$$

The density $p_r(r)$ of the output $r(t) = s(t) + \nu(t)$ results from the convolution of $p_s(s)$ with the Gaussian noise density $p_\nu(\nu)$

$$p_r(r) = \int_{-A}^{A} \frac{1}{A\pi\sqrt{1-(s/A)^2}} \frac{1}{\sqrt{2\pi}\sigma_\nu^2} e^{-\frac{(r-s)^2}{2\sigma_\nu^2}} ds\,. \tag{2.27}$$

We may numerically calculate the differential output entropy with

$$H(R) = -\int_{-\infty}^{\infty} p_r(r)\log_2 p_r(r) dr\,. \tag{2.28}$$

For univariant GWN and signal amplitude $A = 2$ we finally get

$$\begin{aligned} T(S;R) &= H(R) - H(N) = \\ &= 2.811 - 2.047 = 0.764 \end{aligned} \tag{2.29}$$

bit per sample.

2.5.3 Transinformation for random phase sinusoid and additive GWN

For the estimation error $\sigma^2_{\nu_{\text{err}}}$ of the unknown phase of a sinusoid with amplitude A disrupted by additive Gaussian noise ν of variance σ_ν^2 the following Cramer-Rao bound is valid (Rife and Boorstyn, 1974)

$$\sigma^2_{\nu_{\text{err}}} \ge \frac{2\sigma_\nu^2}{A^2 n}\,. \tag{2.30}$$

n is the number of samples per epoch. The phase error does not depend on the number of periods within one epoch. Equality is achieved if the signal-to-noise ratio is sufficiently large, i.e., SNR > 3 dB, which is true for the analyses shown in Fig. 2.11 except for the smallest epoch size n=10. The phase error is very close to

being normally distributed which is revealed by quantile-quantile plots (Figure not shown). Hence, its entropy yields

$$H(N_{\text{err}}) = \frac{1}{2}\log_2(2\pi\sigma^2_{\nu_{\text{err}}}) = \frac{1}{2}\log_2\frac{4\pi\sigma_\nu^2}{A^2 n} \; . \tag{2.31}$$

The phase s is uniformly distributed and has the entropy

$$H(S) = -\int_{-\pi}^{\pi} p_s(s)\log_2 p_s(s)\; ds = \log_2 2\pi \approx 2.651 \text{ bit} \; . \tag{2.32}$$

The transinformation may be calculated as

$$T(S;R) = H(R) - H(N_{\text{err}}) \approx H(S) - H(N_{\text{err}}) \; . \tag{2.33}$$

For $\sigma_{\nu_{\text{err}}} \ll \sigma_s = \pi^2/3$ the approximation may be justified. Otherwise, the entropy $H(R)$ of the output signal has to be determined. Assuming an additive noise model $r = s + \nu_{\text{err}}$ we may estimate the density $p_r(r)$ from the convolution

$$p_r(r) = \int_{-\infty}^{\infty} p_s(s)p_\nu(r-s)ds \; . \tag{2.34}$$

Now, the expression

$$H(R) = \int_{-\infty}^{\infty} p_r(r)\log_2 p_r(r)dr \tag{2.35}$$

can be solved numerically. For the settings $A^2 = 4\sigma_\nu^2$ we yielded the values shown in the upper rows of Table 2.2 depending on different epoch sizes n. The output entropy $H(R)$ is smaller than T because the irrelevance is negative. In case of two sinusoids with the second a phase-locked harmonic of the first, the signal entropy is not increased because no information has been added. However, the signal-to-noise ratio is improved. Assuming that the detection of the phases is independent for both sinusoids due to their orthogonality, we may add the separate transinformation values

$$\begin{aligned} T &= T_1 + T_2 \\ &= H(R_1) - H(N_{\text{err}_1}) + H(R_2) - H(N_{\text{err}_2}) \\ &\approx 2H(S) - H(N_{\text{err}_1}) - H(N_{\text{err}_2}) \; . \end{aligned} \tag{2.36}$$

Using the exact approach we get the values given in the lower rows of Table 2.2 if we set $A_1^2 = 4\sigma_\nu^2$ and $A_2^2 = \sigma_\nu^2$.

2.5.4 Transinformation for epoch aligned impulse response and additive GWN

The channel input may be seen as an epoch aligned impulse response evoked by a delta impulse with Gaussian distributed amplitude ξ of unit variance. The additive

Table 2.3: Transinformation T for epoch aligned impulse response of length n and additive GWN given in bit per epoch.

n	10	20	50	100	200	250	500	1000
T	1.065	1.511	2.138	2.626	3.120	3.280	3.778	4.277

unit variant GWN ν is independent for each of the n samples within the epoch. Thus, we face the problem of parallel Gaussian channels with correlated input and additive independent noise (cf. multiple-access channel with correlated sources in Cover and Thomas, 1991). However, one may simplify the task as follows. The input is identical for each channel despite of a factor due to the weighting introduced by the shape of the impulse response. Let $f(k)$ be the value of the impulse response at sample k. In the trivial case of only one sample per epoch the transinformation is given by

$$T = \frac{1}{2}\log_2\left[1+\frac{f^2(1)\sigma_\xi^2}{\sigma_\nu^2}\right] . \tag{2.37}$$

If the epochs are longer and the impulse response is extended to epoch length, the transinformation per sample may not be added since the input is correlated and the channels chiefly transmit the same information. However, the signal-to-noise ratio is improved by connecting the channels in parallel. Dependent on the weighting $f(k)$ those channels with small weight mainly transmit noise, others with larger weight contribute more significantly to the transinformation. But how do the noisy channels have to be weighted to maximize the information about the input ξ? Intuitively one would expect to weight them according to the factors $f(k)$. This shall be proved as follows. Let $g(k)$ be the unknown weighting factors and $\nu(k)$ independent GWN sources for each channel. The channel input is

$$s(k) = \xi\, f(k) \tag{2.38}$$

and thus, the channel output is

$$r(k) = \xi\, f(k) + \nu(k) . \tag{2.39}$$

Let ρ be the result of the superposition of the weighted channel output

$$\rho = \sum_k r(k)g(k) = \sum_k \left[\xi\, f(k)g(k) + \nu(k)g(k)\right] . \tag{2.40}$$

The noise term may be seen as one new random variable $\tilde{\nu} = \sum_k \nu(k)g(k)$ with variance

$$\sigma_{\tilde{\nu}}^2 = \sigma_\nu^2 \sum_k g^2(k) \tag{2.41}$$

and it can be written

$$\rho = \tilde{\nu} + \xi \sum_k f(k)g(k) \; . \tag{2.42}$$

We may now demand unbiasedness for ρ to uniquely determine $g(k)$. Using (2.40) and knowing that the noise $\nu(k)$ has zero mean we have

$$\langle \xi - \rho \rangle = \left\langle \xi \left[1 - \sum_k f(k)g(k) \right] \right\rangle \stackrel{!}{=} 0 \; . \tag{2.43}$$

The condition is fulfilled if

$$\sum_k f(k)g(k) = 1 \; . \tag{2.44}$$

Now the weights $g(k)$ shall be adjusted to maximize T between ξ and ρ with

$$T = \frac{1}{2} \log_2 \left[1 + \frac{\sigma_\xi^2 \left[\sum_k f(k)g(k) \right]^2}{\sigma_\nu^2 \sum_k g^2(k)} \right] \; . \tag{2.45}$$

It is sufficient to maximize

$$\frac{\left[\sum_k f(k)g(k) \right]^2}{\sum_k g^2(k)} \; . \tag{2.46}$$

Since the numerator is constant (see (2.44)) we merely need to minimize the denominator. From Schwarz's inequality

$$\left[\sum_k f(k)g(k) \right]^2 \leq \sum_k f^2(k) \sum_k g^2(k) \tag{2.47}$$

we get

$$\sum_k g^2(k) \geq \frac{\left[\sum_k f(k)g(k) \right]^2}{\sum_k f^2(k)} \; . \tag{2.48}$$

Equality and thus, the minimum of $\sum_k g^2(k)$ is achieved if $g(i) = \lambda f(i)$. From (2.44) follows

$$\lambda = \frac{1}{\sum_k f^2(k)} \; . \tag{2.49}$$

We finally get the weights

$$g(i) = \frac{f(i)}{\sum_k f^2(k)} \; . \tag{2.50}$$

For a cosine impulse with unit amplitude of the form shown in Fig. 2.8 F, the transinformation depends on epoch length n as shown in Table 2.3.

2.5.5 Entropy estimation and transinformation of scalar variables by means of density estimation

If enough samples of a random variable x are available, its differential entropy can be estimated by means of density estimation. The simplest way of density estimation consists in the construction of a histogram, though the discretization may introduce a bias (Moddemeijer, 1989). Examples for different ways of entropy estimation are given by Bercher and Vignat (2000). Throughout the paper we have applied a local polynomial fit density estimation procedure (Loader, 1997, 1999). The smoothing parameter was chosen in a way to minimize the crossvalidation error. Having obtained an estimate $p(x)$ the entropy is given by

$$H(X) = \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx \,. \tag{2.51}$$

If two variables x and y are related according to an additive noise model $y = x + n$, with the noise n being independent, the transinformation between the variables may be calculated as

$$T(X;Y) = H(Y) - H(Y|X) = H(Y) - H(N) \tag{2.52}$$

(Shannon, 1948). The entropies may be estimated as indicated above.

2.5.6 Decorrelation by means of PCA - an example

Let us consider signal epochs of length n that are to be analyzed with respect to correlation between different sample positions. As long as each sample $\vec{x}_j$[21] can be described as the sum of an independent signal part $\vec{\xi}_j$ and a redundant part comprised of a linear combination of the other samples, i.e.,

$$\vec{x}_j = \vec{\xi}_j + \sum_{i=1;\ i\neq j}^{n} k_{j,i} \vec{x}_i \tag{2.53}$$

then PCA yields a representation of the epochs with uncorrelated samples. This is exemplified in the scatter plots C and D of Fig. 2.15. While linear correlation is present in the original epochs (C) after linear transformation using PCA this correlation vanishes (D). It should be noted that PCA does not achieve statistical independence, i.e., it is not able to restore the original independent variables $\vec{\xi}_j$. It merely yields variables that span the same linear space. If the dependency is strongly nonlinear (noisy sinusoid epochs in B), the linear decorrelation by means of PCA does not yield a satisfying result as illustrated by scatter plots E and F (cf. Deco and Obradovic, 1997).

[21] The components of the sample vector $\vec{x}_j$ are the values of the j^{th} sample for different epochs.

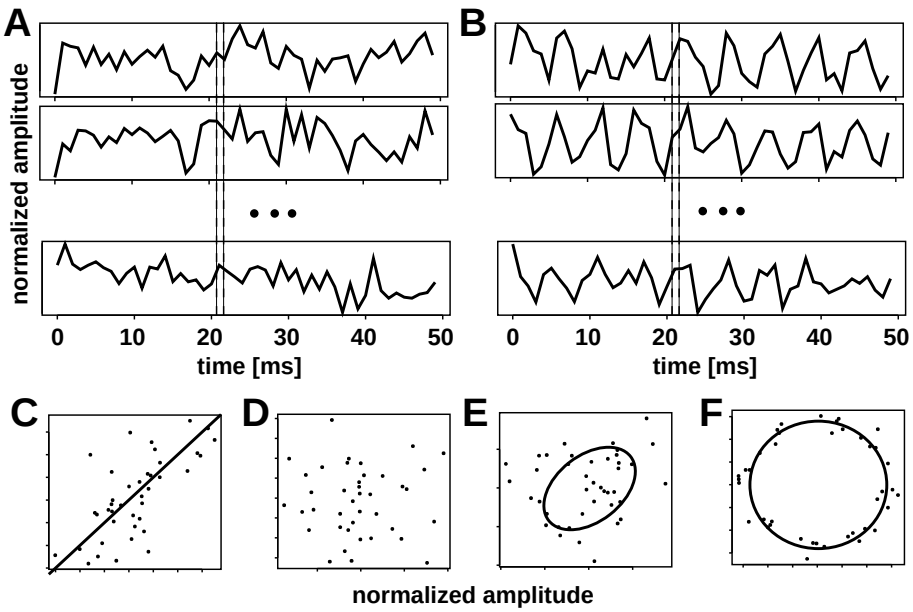

Figure 2.15: The example epochs of colored noise (**A**) and randomly phase shifted sinusoids (**B**) with superimposed GWN. The vertical lines indicate those variables that are related to each other by means of the scatter plots (**C**, **E**) at the bottom. The units are normalized amplitudes and have been omitted for clarity reasons. The lines or circles in the scatter plots are interpolations in the noiseless case and reveal the relation between the variables. The scatter plots **D**, **F** are the result after performing PCA. D does not seem to contain any residual structure, the linear correlation has been removed. However, F resembles a circle indicating the presence of nonlinear dependencies which cannot be removed by means of PCA. It should be noted, that after applying PCA to the epochs B, merely two significant coefficients remain (see Fig. 2.9 D).

2.5.7 Relation between irrelevance and equivocation

Let a continuous channel be perturbed by additive Gaussian white noise (GWN) N. If the transmitter signal P is GWN, the power of the received signal R is $\mathbf{R} = \mathbf{P} + \mathbf{N}$. With the entropies

$$H(R|P) = H(N) = \frac{1}{2}\log_2 2\pi e\mathbf{N} \text{ and} \tag{2.54}$$

$$H(R) = \frac{1}{2}\log_2 2\pi e\mathbf{R} \tag{2.55}$$

we obtain the transinformation

$$T(P;R) = \frac{1}{2}\log_2 \frac{\mathbf{R}}{\mathbf{N}} \tag{2.56}$$

and consequently, the equivocation

$$H(P|R) = H(\tilde{N})$$

$$\begin{aligned} &= H(P) - H(P;R) \\ &= \frac{1}{2}\log_2 2\pi e\mathbf{N}\frac{\mathbf{P}}{\mathbf{R}} \end{aligned} \tag{2.57}$$

with $\tilde{\mathbf{N}} = \mathbf{NP}/\mathbf{R}$. The constant

$$g = \frac{\mathbf{P}}{\mathbf{R}} \tag{2.58}$$

relates the fictitious noise power $\tilde{\mathbf{N}}$ of the equivocation and the noise $\mathbf{N}$. Only if $\mathbf{P} = \mathbf{R}$ the irrelevance is equal to the equivocation.

2.5.8 Noise from two perspectives

In the simple case of GWN signals and additive noise, the noise power can be indirectly defined in two ways (cf. Fig. 2.5): $\mathbf{P}+\mathbf{N} = \mathbf{R}$ and $\mathbf{R}+\mathbf{N}_{\text{eff}} = \mathbf{P}/\mathbf{G}$. Now $\mathbf{G}$ shall be determined. From the two equations follows

$$\frac{\mathbf{N}}{\mathbf{N}_{\text{eff}}} = \frac{1-\frac{\mathbf{P}}{\mathbf{R}}}{\frac{\mathbf{P}}{\mathbf{R}}\frac{\mathbf{1}}{\mathbf{G}}-1}\,. \tag{2.59}$$

On the other hand T has to be identical in both cases

$$T = \frac{1}{2}\log_2\left[1+\frac{\mathbf{P}}{\mathbf{N}}\right] = \frac{1}{2}\log_2\left[1+\frac{\mathbf{R}}{\mathbf{N}_{\text{eff}}}\right] \tag{2.60}$$

which leads to the ratio

$$\frac{\mathbf{N}}{\mathbf{N}_{\text{eff}}} = \frac{\mathbf{P}}{\mathbf{R}}\,. \tag{2.61}$$

Replacing $\mathbf{N}/\mathbf{N}_{\text{eff}}$ in (2.59) with (2.61) we finally get

$$\mathbf{G} = \left[\frac{\mathbf{P}}{\mathbf{R}}\right]^2\,. \tag{2.62}$$

2.5.9 Derivation of coherence and gain with regard to the effective noise

The derivation shall be performed in analogy to the calculation in the frequency domain indicated in (Theunissen, 1993), however the simplifications allowed for single kernel estimation do not apply here. Nevertheless, the result turns out very similar (cf. Roddey et al., 2000). Here, the parameter f denotes the coordinates of the used coordinate system, e.g., "frequency" in the frequency domain or principal component coefficients. The effective noise relative to the response shall be defined as the deviation between the original and the predicted response according to (Bialek et al., 1991)

$$N_{\text{eff}}(f) = P(f)/g(f) - R(f)\,. \tag{2.63}$$

$g(f)$ is a parameter dependent gain constant to be adjusted by means of linear regression between R and P using a least squares formulation on several signal sections

$$\chi^2[g(f)] = \left\langle \int_{-\infty}^{+\infty} |P(f) - g(f)R(f)|^2 \, df \right\rangle . \tag{2.64}$$

The brackets indicate that the mean over these signal sections is being taken. Minimization with respect to the real and imaginary components of $g(f)$ leads to

$$g(f) = \frac{\langle P(f)R^*(f)\rangle}{\langle R(f)R^*(f)\rangle} \tag{2.65}$$

which means that any residual correlation between the model error $P(f)/g(f) - R(f)$ and the original response is eliminated since (2.65) can be transformed to yield

$$\left\langle \left(\frac{P(f)}{g(f)} - R(f) \right) R^*(f) \right\rangle = 0 . \tag{2.66}$$

By means of (2.63) and (2.65)

$$\begin{aligned} \left\langle |N_{\text{eff}}(f)|^2 \right\rangle &= \left\langle \left| \frac{P(f)}{g(f)} - R(f) \right|^2 \right\rangle \\ &= \frac{\left\langle |P(f)|^2 \right\rangle}{|g(f)|^2} - \left\langle |R(f)|^2 \right\rangle \end{aligned} \tag{2.67}$$

we finally get the signal-to-noise ratio

$$\begin{aligned} \text{SNR}(f) &= \frac{\left\langle |R(f)|^2 \right\rangle}{\left\langle |N_{\text{eff}}(f)|^2 \right\rangle} \\ &= \frac{|\langle P(f)R^*(f)\rangle|^2}{\left\langle |R(f)|^2 \right\rangle \left\langle |P(f)|^2 \right\rangle - |\langle P(f)R^*(f)\rangle|^2} . \end{aligned} \tag{2.68}$$

Obviously, it can be written as a function of the coherence between the response and its estimate (cf. Stein and French, 1970)

$$\text{SNR}(f) = \frac{\gamma^2(f)}{1 - \gamma^2(f)} \tag{2.69}$$

with

$$\gamma^2(f) = \frac{\langle P(f)R^*(f)\rangle \langle P^*(f)R(f)\rangle}{\left\langle |R(f)|^2 \right\rangle \left\langle |P(f)|^2 \right\rangle} . \tag{2.70}$$

The calculation of T based on this signal-to-noise ratio yields a lower bound if the samples of R are Gaussian distributed and independent (Rieke et al., 1998). When

stimuli are identically repeated, (2.68) is simplified since the response estimate does not vary over the repetitions. With

$$P(f) = \langle P(f) \rangle \tag{2.71}$$

the signal-to-noise ratio is

$$\begin{aligned} \mathrm{SNR}(f) &= \frac{|\langle R(f)\rangle|^2}{\left\langle |R(f)|^2 \right\rangle - |\langle R(f)\rangle|^2} \\ &= \frac{|\langle R(f)\rangle|^2}{\left\langle |R(f) - \langle R(f)\rangle|^2 \right\rangle} \\ &= \frac{|\langle R(f)\rangle|^2}{\left\langle |N(f)|^2 \right\rangle} \,. \end{aligned} \tag{2.72}$$

The transinformation calculated on the grounds of this signal-to-noise ratio resembles an *upper* bound of T (Haag and Borst, 1997; Borst and Theunissen, 1999). This illustrates how different the results might be depending on the exact stimulus paradigm. To calculate a *lower* bound of information, the examined signal sections need to reflect different stimulus conditions.

3 Assessing the Encoding of Stimulus Attributes with Rapid Sequences of Stimulus Events

In a preceding paper (cf. Chapter 2, Eger and Eckhorn (2002b)) we have published a *three step method* for the quantification of transinformation in multi-input and -output neuronal systems. Here we present an extension that applies to rapid series of transient stimuli and thus, fills the gap between the discrete and continuous stimulation paradigm. While the *three step method* potentially captures all stimulus aspects, the present approach quantifies the discriminability of selected attributes of discrete stimuli and thus, assesses their encoding. Based on simulated and recorded data we investigate the performance of the implemented algorithm. Our approach is illustrated by analyses of neuronal population activity from the visual cortex of the cat, evoked by electrical stimuli of the retina.

3.1 Introduction

Since the pioneering work of Shannon (1948) information theory has become popular in many fields including neuroscience (e.g., MacKay and McCulloch, 1952; Werner and Mountcastle, 1965; Eckhorn and Pöpel, 1974; Optican and Richmond, 1987; Bialek et al., 1991; Rieke et al., 1998; Borst and Theunissen, 1999; Buracas and Albright, 1999; DeWeese and Meister, 1999; Victor, 1999; deCharms and Zador, 2000; Reich et al., 2000). A central concept is the *transinformation*[1] T which describes the amount of transmitted information between the input and output of an information channel. T is advantageous compared to other correlation measures because it is independent of the types of the related signals, their units and gains and takes the noise into account. Yet, depending on the information channel, the complexity of the signals, and the degree of signal correlation, a precise estimation of T is difficult. In order to directly calculate T, the a priori and conditional probability densities are needed. In practice, the probabilities are often not available or too demanding to estimate with sufficient precision. A precise estimation requires

[1] We use the equivalent term “transinformation” instead of “mutual information” because the systems we investigate have a directed information flow.

long sequences of stationary data. For this reason, models of the noiseless information channel are being used (eg. Rieke et al., 1998; Gershon et al., 1998; Wiener and Richmond, 1998). Instead of quantifying T between the input and output of the system one may alternatively calculate T between the original signal and its deterministic model (Eckhorn and Pöpel, 1981; Eckhorn and Querfurth, 1985). In this context, two different stimulation paradigms are common in the field of neuroscience and at present are being heavily discussed with respect to controversial amounts of transinformation values (Optican and Richmond, 1987; Buracas et al., 1998; deCharms and Zador, 2000; Reich et al., 2000).

The *continuous stimulation* paradigm constitutes temporally continuous stimulation signals as input to the investigated neuronal information channel. The channel output, i.e., the neuronal response is likewise continuous. Due to the non-zero memory of the causal system, a response sampling value at t_0 depends on previous sampling values of the stimulation signal at times $t < t_0$. In order to avoid taking into account numerous joint probabilities associated with stimulus and response samples at *different* times, a channel model may be used that incorporates an estimate of the channel's dynamics. Thus, one may constrain the analysis to *equal* times of the stimulus-based deterministic response model and the original response. It is not necessary to consider dependencies between different times. In a preceding paper we presented a *three step method* for the calculation of T by using a linear stimulus–response model (cf. Chapter 2, Eger and Eckhorn (2002b)). To some degree similar approaches are the *upper bound method* (Haag and Borst, 1997; Borst and Theunissen, 1999) and the *stimulus reconstruction technique* (Bialek et al., 1991; Warland et al., 1997; Rieke et al., 1998). Each of these methods have their advantages and restrictions (cf. Discussion in Chapter 2 on p. 29, Eger and Eckhorn (2002b)).

In this paper we focus on a stimulation paradigm which is based on temporally *discrete stimulation* events. This stimulation paradigm has been widely applied in neuroscience, especially with respect to analyses of sensory systems (e.g., Werner and Mountcastle, 1965; Stein, 1967; Richmond et al., 1987). While the investigated neuronal system is being repeatedly stimulated with different types of stimuli constituting different discrete stimulus conditions (e.g., varied stimulus intensities or spatial visual patterns) the conditional neuronal or behavioral responses are averaged to yield the deterministic response model and so to define the reference for the response variability. Normally, the stimulus intervals are adjusted to not fall below the system's memory time constant, i.e., a new stimulus event is not started before the responses to previous ones have sufficiently subsided (cf. Theunissen and Miller, 1995). Then, responses to individual stimulus events may be analyzed independently of each other (e.g., Optican and Richmond, 1987; Richmond and Optican, 1987).[2] In our case, we want to assess the capability of the investigated

[2] When the response epochs are described by a single parameter (e.g., the spike count in the case of action potential data) then for each of m stimulus conditions, conditional spike count histograms are necessary for the calculation of T (3.8). Note that the assignment of a single parameter such as

neuronal system to distinguish between different stimulus conditions also at higher stimulation rates. Therefore, we need an approach that is not constrained to long inter-stimulus intervals. For this reason, we have developed a new method that is similar to the approach mentioned above but not constrained to independent responses, i.e., the inter-stimulus interval may be shorter than the system's memory. With this method we can address relevant questions, such as how well selected stimulus attributes are neuronally encoded at high stimulation rates.

In combination with our *three step method* (cf. Chapter 2, Eger and Eckhorn (2002b)) it is possible to use identical stimulation signals with two methods attaining comparable values of T, though under different assumptions concerning the underlying code. Whereas the *three step method* potentially captures information about all stimulus aspects, the present method focuses on the discriminability of selected stimulus attributes encoded in the response – while each response to a discrete stimulus event is handled independently. A comparison reveals the contribution to the transinformation that is selectively due to the temporal properties of the stimulus events.

In the following section we will derive the method which is exemplified by data recorded from the cat primary visual cortex (V1) in response to electrical retina stimuli (Hesse et al., 2000). The method has been applied to simulated and experimental data to investigate the accuracy and possible limitations of the algorithm.

3.2 Methods

A response to a stochastic sequence of stimulus events with 2 different intensities may at most yield 1 bit per stimulus if one focuses on the discrimination of the stimulus conditions, i.e., the encoding of the intensity attribute. In contrast, when also the timing of the stimulus events is taken into account much higher values of T may result. Here, we are interested in the encoding of selected stimulus attributes in the case of sequences of stimulus events with relatively high mean stimulus event rate. As a consequence, response deflections evoked by subsequent stimulus events may not be well separated in time.

This temporal interference requires a new approach to assign a suitable response measure to the corresponding stimulus event. As long as there is no interference, a dimensional reduction of the response space may be performed by applying principal component analysis (PCA) on the response epochs directly (Glaser and Ruchkin, 1976; Richmond and Optican, 1987). In our case of mutually interfering responses we proceed as follows.

First, we calculate conditional linear stimulus–response kernels as outlined in Chapter 2 (Eger and Eckhorn, 2002b). These waveforms characterize the deter-

the spike count as response measure is arbitrary and here, it corresponds to the assumption of a rate code.

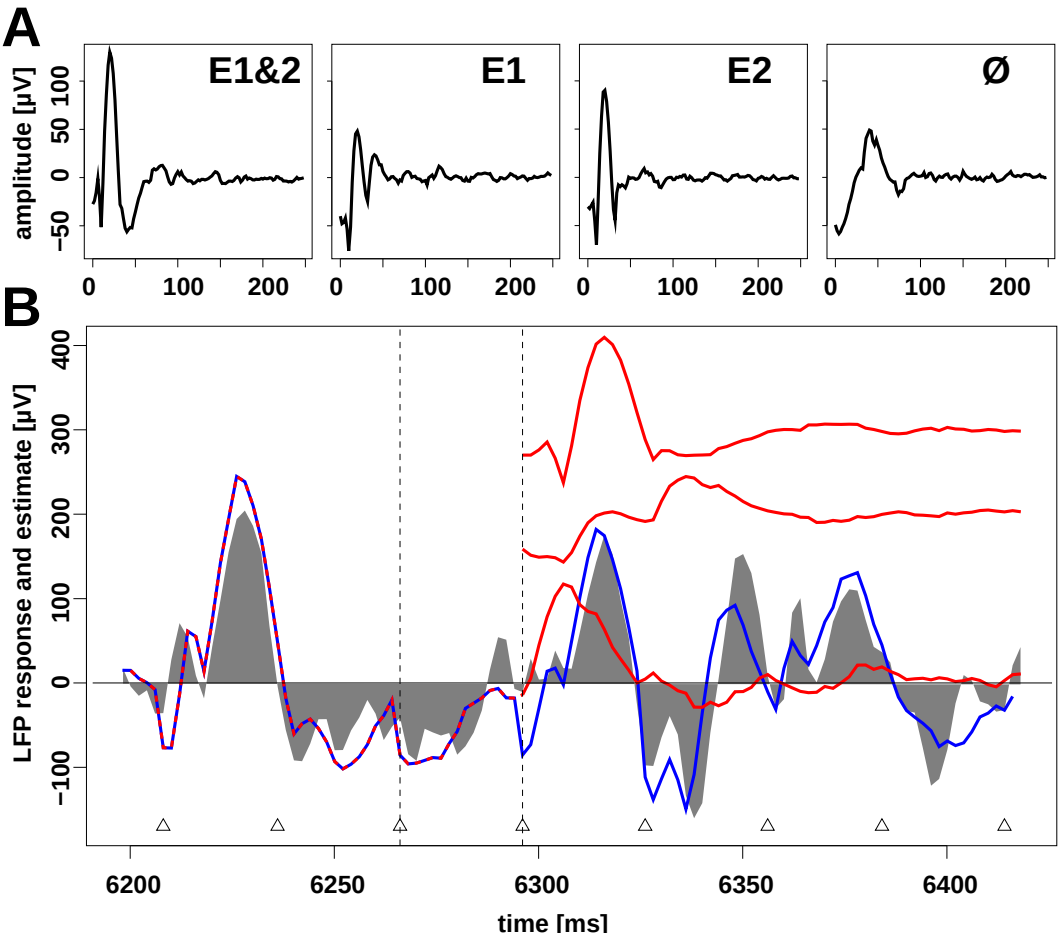

Figure 3.1: Example of electrically evoked local field potentials from cat striate cortex and the construction of a response fit using shifted and weighted principal components. The retina was stimulated rhythmically with biphasic current impulses (100 µA amplitude, 200 µs phase width, negative leading phase, charge balanced, 29 imp/s) randomly with one or two electrodes 300 µm apart. **A**: Linear kernels corresponding to the four possible stimulus conditions with two stimulation electrodes (both (E1&2), either (E1, E2) or none (Ø) electrode(s) active). The deflection in panel Ø may be interpreted as off-response. The procedure for the calculation of the kernels is outlined in Chapter 2 (Eger and Eckhorn, 2002b). **B**: Original response (shaded) and fit (blue) based on all stimuli. The triangles at the bottom mark the stimulus events. The red curve shows the fit based on all events up to the first dashed vertical line. The two signals (top right, red) are the first two principal components of the kernels shown in A, scaled to fit best to the response. They resemble the effect of the stimulus (at second dashed vertical line) on the response. The sum of the three dashed curves yields the blue curve until the next stimulus.

ministic fraction of evoked response deflections for each different stimulus condition. However, they are not mutually orthogonal and may even be very similar for different stimulus conditions. To reduce the redundancy PCA is applied on the conditional kernels yielding a set of orthogonal principal components. Now, an optionally reduced set of these waveforms is used to construct a least squares fit of the evoked response. The principle of this approach is depicted in Fig. 3.1. Superimposed principal components are shifted and weighted to fit the original response signal minimizing the squared error. The weighting factors describe the contribution of the respective waveform to each stimulus evoked response epoch and hence, provide the desired response measures. In case of non-interfering responses the response measure vectors are equivalent to the principal component coefficients. In the more general case the response measures are determined as follows:

Let $r(t)$ be the response signal and $\vec{p}(t)$ a time-dependent vector. Its elements are waveforms, i.e., the principal components of the m kernels computed according to Chapter 2. To characterize a short epoch of the response signal $r(t)$ following the i^{th} of n_z repetitions of stimulus z at time t_{z_i}, we can write

$$r(t) = \vec{c}_{z_i}^{\,T} \left[\, \vec{p}(t) * \delta(t - t_{z_i}) \right] \, , \tag{3.1}$$

or in short, omitting the index t

$$r = \vec{c}_{z_i}^{\,T} \vec{p}_{z_i} \tag{3.2}$$

where "$*$" indicates convolution and $\vec{c}_{z_i}^{\,T}$ the vector of yet unknown weighting coefficients. The subscript of $\vec{p}_{z_i}$ denotes that the components $\vec{p}(t)$ have to be temporally shifted to their respective stimulus event z_i. In the case of non-interfering responses the principal component coefficients are determined by the scalar product of the evoked response with the shifted principal components (Glaser and Ruchkin, 1976)

$$\vec{c}_{z_i} = \left\langle r \vec{p}_{z_i} \right\rangle_t \, . \tag{3.3}$$

However, when the intervals between stimulus events become shorter than the system's memory, the interference of individual responses needs to be taken into account. In simplified form the stimulus signal may be described as a sequence of delta impulses each symbolically weighted by the identifier a_{z_i} of the stimulus event

$$s = \left\langle \sum_{i=1}^{n_z} \delta(t - t_{z_i}) a_{z_i} \right\rangle_z \, . \tag{3.4}$$

Accordingly, the response fit

$$r_{\text{fit}} = \left\langle \sum_{i=1}^{n_z} \vec{c}_{z_i}^{\,T} \vec{p}_{z_i} \right\rangle_z \tag{3.5}$$

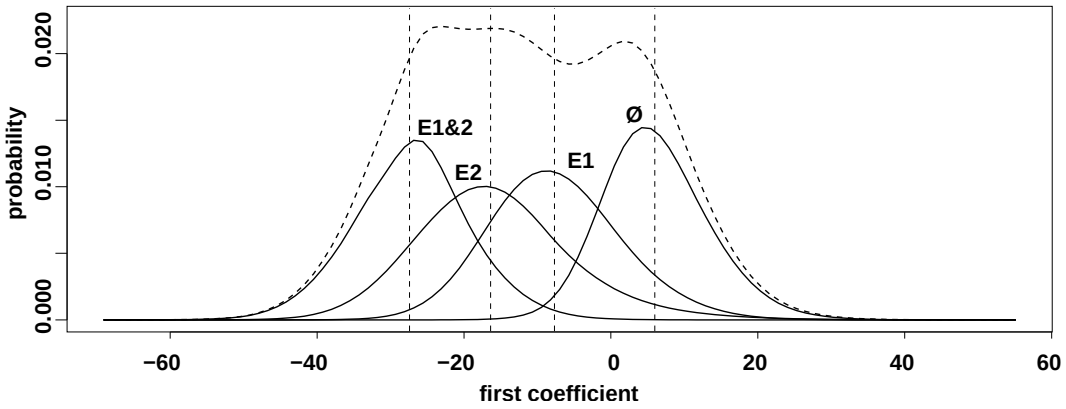

Figure 3.2: Example of probability densities of the first weighting coefficient corresponding to the four stimulus conditions given in Fig. 3.1 (solid) and the overall response (dashed). In order to illustrate how the conditional densities form the overall density the conditional densities were scaled with the probability of the respective stimulus condition. Note that the four distributions can well be distinguished. The rate of transinformation calculated using (3.8) is $T = 22.3 \pm 0.7$ bit/s or 0.77 ± 0.02 bit/symbol at stimulus event rate of 29 Hz. The standard deviation was determined using the jackknife algorithm (Efron and Tibshirani, 1991; Efron, 1994). In comparison, a backpropagation neural net (Kjaer et al., 1994) yielded $T = 0.70 \pm 0.004$ bit/symbol for the first coefficient and 0.79 ± 0.004 bit/symbol for two coefficients. The very low standard deviations in the latter case are estimates based on 4 measurements and not directly comparable with those yielded by the density approach.

is a sum of weighted and shifted principal components. To determine the vectors $\vec{c}_{z_i}$ the squared error between the response r and the fit r_{fit} shall be minimized

$$\chi^2 = \left\langle \left| r - \left\langle \sum_{i=1}^{n_z} \vec{c}_{z_i}^T \vec{p}_{z_i} \right\rangle_z \right|^2 \right\rangle_t . \tag{3.6}$$

The derivation with respect to $\vec{c}_{z_i}$ at $z = \zeta$ and $i = \iota$ with $\zeta \in \{1, \ldots, m\}$ and $\iota \in \{1, \ldots, n_z\}$ leads to a system of $m \times n_z$ equations of the form

$$\left\langle r \vec{p}_{\zeta_\iota} \right\rangle_t = \left\langle \sum_{i=1}^{n_z} \left\langle \vec{p}_{z_i} \vec{p}_{\zeta_\iota}^T \right\rangle_t \vec{c}_{z_i} \right\rangle_z . \tag{3.7}$$

The solution of these equations provides the response measures for each single stimulus event z_i. This allows to directly calculate T between the stimulus identifiers a_{z_i} and corresponding response measure vectors $\vec{c}_{z_i}$ using standard methods. A comparison of several algorithms regarding their precision under the restriction of a limited amount of data is presented in (Panzeri and Treves, 1996; Golomb et al., 1997). For the example shown in Fig. 3.2 we used an adaptive density estimation algorithm (Loader, 1997, 1999). The smoothing parameter was chosen in a way to

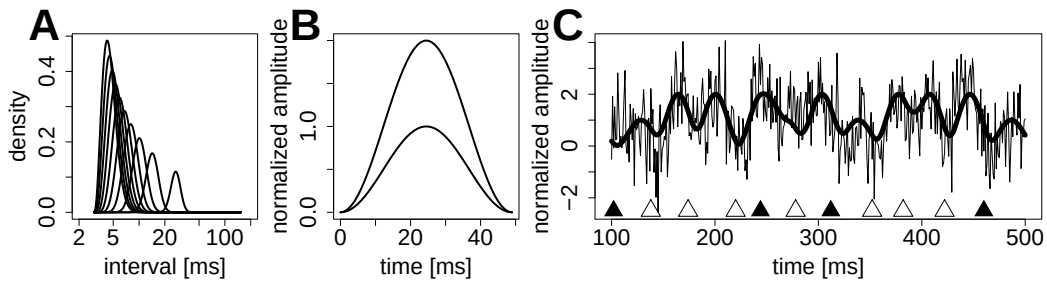

Figure 3.3: Distribution of simulated stimulation intervals (**A**), the convolution kernel (**B**), and an example response section (**C**). The kernel is weighted with factor 2 when the corresponding stimulus belongs to class 2. The stimulus events are indicated in C by filled (1) and non-filled (2) triangles at the bottom line.

minimize the crossvalidation error. We calculated T (in bit per symbol) using the formula

$$T(S;R) = \sum_{z=1}^{m} p_{s_z} \int_{-\infty}^{\infty} p_{r|s_z}(r) \log_2 \frac{p_{r|s_z}(r)}{p_r(r)} \, dr \; . \tag{3.8}$$

Alternatively, we applied the backpropagation neural net (Kjaer et al., 1994) and yielded very similar results. Concerning the quantification of T it should be noted that the temporal interference of responses is implicitly classified as noise which leads to a decrease of T (cf. Discussion on page 59).

3.3 Results

In the following, the accuracy and reliability as well as the limitations of our method will be demonstrated on the basis of artificial data that have controllable statistical properties.

Troy and Robson (1992) have shown that action potential trains of retinal ganglion cells can well be described by a renewal process with Gamma distributed intervals. In our experiments we electrically stimulated the retina in a way to mimic retinal activity evoked by visual stimuli. For this reason, test signals were simulated sequences of stimulus events with either constant or Gamma distributed intervals (Fig. 3.3 A). Half of the randomly selected events had unit amplitude and were assigned to stimulus class 1, the other half of double amplitude was sorted into class 2. The mean rate was varied within 10 and 100 impulses per second. These stimulus events were linearly convolved with a 50 ms cosine waveform of unit amplitude and 1 ms resolution resembling the impulse response of a fictitious linear channel (Fig. 3.3 B). To coarsely mimic the variability of a neuronal system we added unit variant GWN. An example section of the simulated noisy response signal is shown in Fig. 3.3 C.

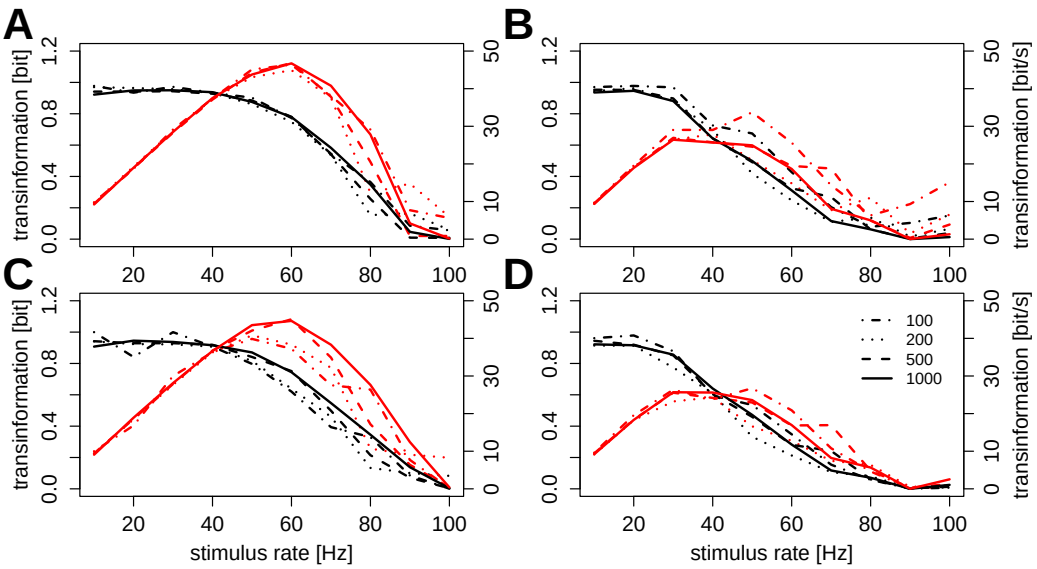

Figure 3.4: Transinformation for the binary distributed stimulus variable dependent on the average stimulus event rate. The absolute number of events was varied between 100 and 1000 (legend in D). The black curves depict the transinformation per symbol (left axes), the red curves the transinformation rate (in bit/s, right axes). A and C are based on constant, B and D on Gamma distributed stimulus intervals according to Fig. 3.3 A. Diagrams A and B result from density estimation, C and D from the application of the backpropagation neuronal net.

According to Methods (p. 50) we performed a least squared error fit of a shifted and weighted waveform (identical to the unscaled kernel) yielding one coefficient per stimulus event. In a last step we applied a density estimation algorithm to calculate T between the binary stimulus sequence composed of 1s and 2s and the continuously distributed response coefficients (Fig. 3.4 A, B) according to (3.8). As a benchmark for the information calculation we applied the *Tbackprop* neuronal net software (Kjaer et al., 1994; Richmond, 1998)[3] (Fig. 3.4 C, D) and compared the results using a scatter plot (Fig. 3.5).

The graphs in Fig. 3.4 have similar shapes for the same stimulus interval distributions but differ with respect to different distributions. The information per symbol yields nearly 1 bit for the lowest rate but flattens at higher rates in the case of Gamma distributed (Fig. 3.4 B, D) compared to constant intervals (Fig. 3.4 A, C). The two algorithms react slightly differently when the number of stimulus events is decreased from 1000 to 100. At smaller data sizes and equally spaced events the density estimation seems more robust, except for the highest stimulus rates (Fig. 3.4 A, C). However, both algorithms mostly underestimate T. In case of Gamma dis-

[3]We expressively thank B. Richmond for supplying us with the *Tbackprop* neuronal net software.

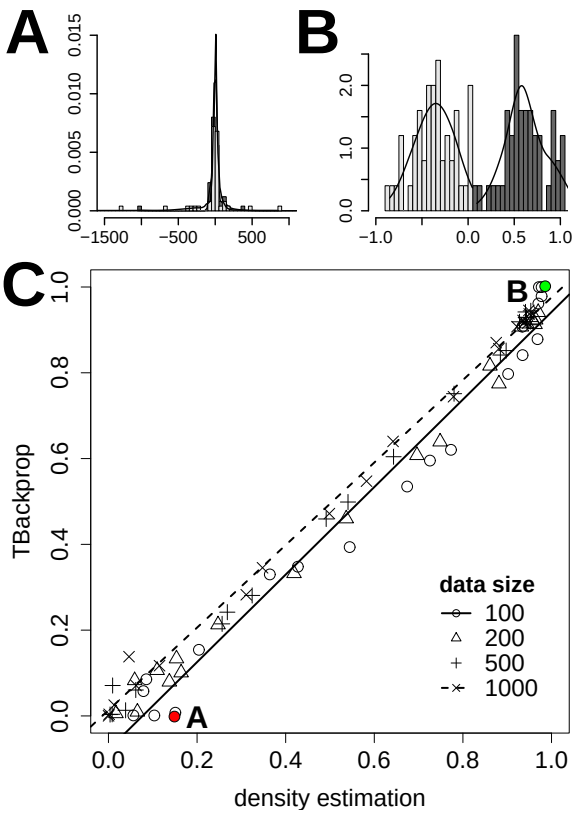

Figure 3.5: Scatter plot of transinformation (in bit per stimulus event) for the density estimation algorithm and the backpropagation neuronal net. **C**: The different dot types indicate the size of the used simulated data sets characterized by the number of stimulus events. The more data available, the better the expected performances of the density estimation and the neuronal net. Especially for the larger data sizes (dot type "x") the deviations between both methods are very small (diagonal dashed regression line). However, there seem to be slightly different slopes and intercepts of the linear regression lines dependent on the data size. **A**, **B**: Histograms and superimposed density estimates of two outliers.

tributed events (Fig. 3.4 B, D), the neuronal net outperforms the density estimation procedure for small amounts of data.

The scatter plot Fig. 3.5 directly compares the performance of both routines based on the data presented in Fig. 3.4. Identical results would lead to a straight line through the origin (0,0) and (1,1). The deviation between both approaches is relatively small, especially for larger data sizes (see inset). However, the two regression lines for the smallest and largest data sample differ slightly in their slope and intercept. This may indicate a small systematic size-dependent deviation between the approaches.

To further demonstrate the suitability of our method we applied it to physiological data from two experiments with different recording lengths, response signal

Table 3.1: Results of information theoretical analysis for 24 physiological measurements.

	size [s]	kno	rate [s^{-1}]	dens1	tback1	dens2	tback2
				[bit/s]			
1	10	4	40.9	0.68	0.65	0.71	0.69
2	10	4	40.9	0.64	0.61	0.69	0.65
3	12	4	33.6	0.10	0.10	0.16	0.15
4	12	4	33.6	0.39	0.34	0.40	0.35
5	14	4	29.1	0.70	0.70	0.77	0.79
6	14	4	29.1	0.56	0.55	0.70	0.65
7	16	4	25.2	0.08	0.06	0.11	0.10
8	16	4	25.2	0.32	0.30	0.38	0.32
9	18	4	22.5	0.44	0.44	0.49	0.49
10	18	4	22.5	0.59	0.57	0.65	0.61
11	26	4	15.5	0.85	0.83	0.91	0.89
12	26	4	15.5	0.65	0.63	0.70	0.66
13	46	4	8.7	0.19	0.18	0.22	0.20
14	46	4	8.7	0.46	0.42	0.51	0.49
15	77	4	5.3	0.19	0.19	0.25	0.23
16	77	4	5.3	0.51	0.44	0.54	0.48
17	87	4	4.7	0.31	0.29	0.36	0.32
18	87	4	4.7	0.51	0.47	0.57	0.52
19	148	4	2.8	0.14	0.13	0.19	0.14
20	148	4	2.8	0.24	0.23	0.32	0.29
21	261	8	15.7	0.30	0.29	0.37	0.37
22	261	8	15.7	0.43	0.43	0.53	0.52
23	261	8	15.7	0.18	0.17	0.27	0.27
24	261	8	15.7	0.10	0.11	0.31	0.22

Data from two experiments with different recording lengths (in seconds, column *size*), signal types (odd row numbers: local field potential; even numbers: multi unit activity, of the same recording) and stimulus conditions (column *kno*, 4 or 8 kernels). We used a single principal component coefficient (column *dens1* vs. *tback1*) and for comparison, the first two coefficients (column *dens2* vs. *tback2*) as response measures.

types, and stimulus conditions (Table 3.1, Fig. 3.6). We extracted a scalar and additionally, for comparison a two-dimensional response measure. The first two principal component coefficients capture usually more than 95% of the response variance. However, in the case of noisy responses, i.e., inefficient stimulation, more than two coefficients may be needed to explain the – admittedly unimportant – response variance. Therefore, we did not consider more than two coefficients. The transinformation between the (4 or 8) stimulus conditions and the extracted response measures was quantified with both, the density estimation approach and the back-

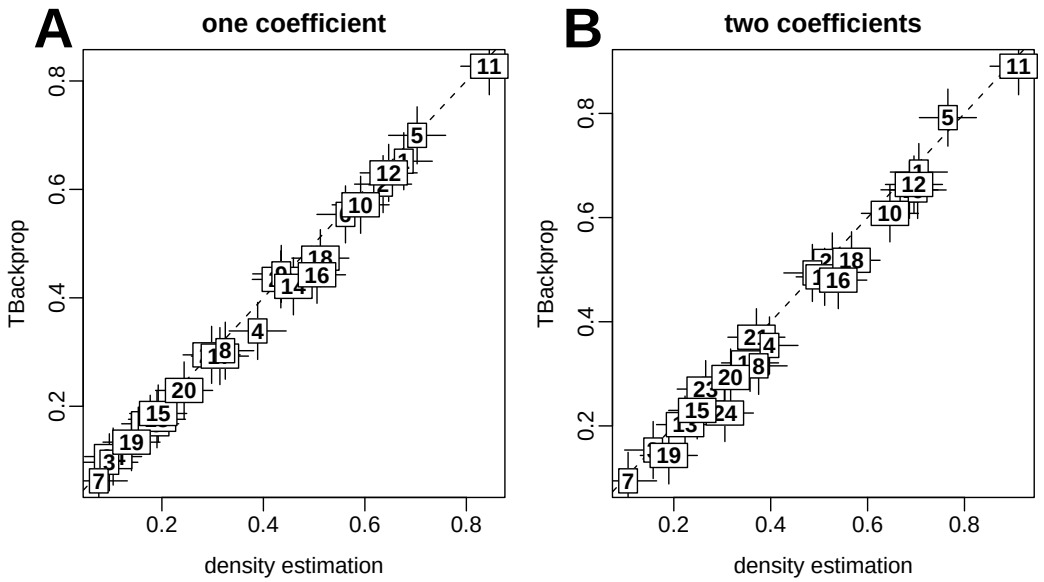

Figure 3.6: Scatter plots of the experimental data shown in Table 3.1 reveal the results of the density estimation approach vs. the backpropagation neuronal net. **A**: Analysis with scalar response measure (first principal component coefficient). **B**: Two-dimensional response measure (first two principal component coefficients). The boxed numbers refer to the measurements in Table 3.1. The diagonal reference line represents equal results for both algorithms.

propagation neuronal net. Table 3.1 presents the quantitative results. As would be expected, the transinformation based on the two-dimensional response measure exceeds the value based on a scalar measure; though often, the deviation is small. Both algorithms yield relatively close results. Fig. 3.6 compares them by means of two scatter plots: Fig. 3.6 A is based on the scalar, Fig. 3.6 B on the two-dimensional response measure. While the scatter along the diagonal reference line is small in A, it is slightly larger in B.

3.4 Discussion

Summary of methods. We have shown that our algorithm may well extract information about the stimulus intensity from simulated noisy response signals. The signals are constructed by linear superposition of temporally shifted and weighted waveforms. The algorithm yields the best linear fit with regard to the least squared error between the simulated response and a response fit based on these to-be-scaled waveforms. For the analysis of experimental data the waveforms' shapes are identical to the principal components of the kernels yielded by step 1 of the three step method (cf. Chapter 2, Eger and Eckhorn (2002b)) and are calculated the same

way. Thus, they span a space that has been previously marked out by the three step method.

As long as the memory of the channel is shorter than the inter-stimulus delay, the response signal sections following subsequent stimulus events do not temporally interfere. In that case, our algorithm corresponds closely to that presented in (Richmond et al., 1987). According to their multiplex filter hypothesis (Optican and Richmond, 1987) these researchers interpreted the principal components as independent characteristic deflections of the responding neuron and used them together with fitted model parameters to predict the neuronal responses to new visual stimuli (Gawne et al., 1991). Although we also calculate a fit, our aim is to yield response measures that correspond to the sequence of stimulus events. The scaling factors resulting from the quadratic fit resemble the desired response measures and are to be information theoretically related to the sequence of stimulus events. This is achieved by means of a density estimation procedure (cf. Chapter 2, Eger and Eckhorn (2002b), Appendix 2.5.5) and alternatively, using a backpropagation neuronal net (Kjaer et al., 1994) as a benchmark. The high quality performance of the net has been assessed in several publications (Panzeri and Treves, 1996; Golomb et al., 1997).

A lower bound of transinformation. The representation of the response signal sections by means of a reduced coefficient vector is not complete. As long as the sections do not temporally interfere, the coefficients are identical to the principal component coefficients yielded by the method presented by Richmond et al. (1987). Due to computational reasons, only the first few coefficients are used omitting the small contributions of the others to the response variance. However, this neglect can only lead to a decrease of T. Additionally, any nonlinearity of the investigated channel results in more or less inaccurate kernels when the stimulus intervals fall below the memory time constant of the system. A nonlinear model that takes into account the temporal properties of the stimulus events might capture more information. In our case, the interference of responses to consecutive stimuli is automatically interpreted as noise because serial correlation of the stimulus is not taken into account. We rather take this as a feature of this method than a flaw. By excluding temporal properties of the stimulus events it is distinct from other methods such as our *three step method* that implicitly captures them.

Performance of the algorithms. A potential source for overestimating T may be hidden in the algorithm that quantifies the mutual information between the stimulus sequence and the response coefficients. However, we did not assume normality and applied a density estimation procedure to handle any type of distribution. A comparison of the results with those of the backpropagation neuronal net (Figures 3.4, 3.5 and Kjaer et al. (1994)) suggests that our density estimation procedure works accurately and does not systematically overestimate T. In case of small data sizes, a reason for minor deviations between the results of the two algorithms probably lies in the different abilities to handle distributions that deviate far from normality.

While the density estimation produces a weak fit in case of long tailed densities and relatively few samples (histogram of outlier in Fig. 3.5 A) the neuronal net has to cope with non-overlapping, range-limited densities (histogram of outlier in Fig. 3.5 B). It has been argued that the network tends to underestimate the transinformation in some cases (Golomb et al., 1997; Panzeri and Treves, 1996). These may be reasons for the deviation of slope and intercept of the regression lines depending on the data size in Fig. 3.5 C.

The performance to extract informative response measures does not merely depend on the stimulation rate but also on the interval distribution. In contrast to Gamma distribution, for constant intervals just exceeding the waveform duration there is no temporal overlap between responses to subsequent stimulus events. Any interference increases the variability of the extracted response measures and thus, impairs the performance. This is the reason why for Gamma distributed intervals the transinformation drops already at low mean stimulation rates (Fig. 3.4 B, D).

The analysis of experimental data further demonstrates the usefulness of our approach. By comparing the transinformation captured by one- and two-dimensional response measures with both algorithms we found that the second coefficient mostly contributes much less to the information transmission than the first. The minimum response measure dimension depends on the complexity of the stimulus set and the dimensionality of the response coding space (cf. Richmond and Optican, 1987). The density estimation procedure and the neuronal net yield similar results. The small deviation is increased in the case of two-dimensional response measures. A possible reason is that in general, the regression variance increases with growing number of variables at constant data size. Moreover, under- or over-fitting of the data may occur (e.g., Loader, 1999).

Table 3.1 shows that the transinformation per stimulus may reach large values for high stimulation rates – even exceeding the results for lower rates in comparable experimental conditions (e.g., compare row 1 with 3, Table 3.1). We explain this effect as a consequence of reduced response variability at high stimulation rates. The transinformation depends on the signal-to-noise ratio that quantifies the discriminability of conditional mean responses normalized by the conditional variance. More detailed analyses that compare the transmitted information about different stimulus aspects are described in Chapter 4.

Finally, as the small deviation between the results suggests, we want to note that the actual choice of the algorithm is not decisive once informative response measures have been extracted from the data. An advantage of the backpropagation neuronal net is its applicability to higher dimensional problems, i.e., to response measures with more than three coefficients. A practical advantage of the density estimation approach is that it is straight forward, comparably fast, and the software is well documented and freely available (Loader, 1997, 1999).

Conclusion. We intended to bridge the gap between the continuous and discrete stimulation paradigms that forcibly lead to different results since the discrete stimu-

lation paradigm does not capture transmitted information about the timing of stimulus events. Therefore, we have developed an alternative method that is a modification of the approach of Optican and Richmond (1987) and works for rapid, stochastic sequences of stimulus events. Our *three step method* (cf. Chapter 2, Eger and Eckhorn (2002b)) and the present approach may be applied to the same data. Then, questions can be addressed such as whether the yielded transinformation values reflect the true encoding capabilities of the investigated system or are rather due to the adopted stimulation paradigm. Here, varying the stimulation rate might give some answer. A further possibility is that the values are due to the applied analysis method that faces implicit restrictions, e.g., the method may capture only certain stimulus attributes encoded in the response. By this paper we hope to harmonize recent controversial results.

4 Temporal, Spatial, and Intensity Resolution with a Retina Implant

Abstract. Our research is connected with the retina implant project that aims at an optimization of retinal electrical multi-site stimulation with respect to perception. We assess the efficiency of various stimulation parameters by means of multi-electrode intracortical recordings of neuronal activity in anesthetized cats using information theory. The analysis of extensive data examples yields the following results: (1) *Temporal*: Single events stimulating retinal X cells evoke precise spike responses with 300 μs standard error. The decorrelated fraction of the population response endures for about 40 ms and carries 0.1–2.0 bit/stimulus. Accordingly, rapid sequences of stimulus events convey a maximum of transinformation (T) in the range of 20–140 bit/s at a mean stimulus rate of 20–40 Hz suggesting a temporal resolution of 25–50 ms. Stimulus efficiency is highest 10–20 ms or long ($>$150 ms) after a facilitating stimulus. When using amplitude modulated trains of stimulus events with Gamma statistic (50 ms mean interval) the temporal stimulus aspect accounts for 50–80% of total T. (2) *Intensity*: injected charge is a weak parameter and requires at Most three quantization steps. Multi unit activity (MUA) encodes intensity better than local field potentials (LFP, 3% vs. 15% of total T) yet shows an increasing variance with growing response strength. This recommends scarcely over-threshold stimuli. (3) *Spatial*: The values of T are in the range of 20–100 bit/s per recording electrode for both electrical and visual stimulation. Not all recording positions are equally informative. With electrical stimulation the cortical profiles of T show a pronounced peak of $<$1 mm radius suggesting a resolution of 2°. In case of visual stimulation the profiles are highly structured. T saturates when stimulating the retina with 7 electrodes per mm^2 confirming the resolution of 2°.

4.1 Introduction

It has been shown that electrical stimulation of still intact retinal ganglion cells can elicit phosphenes in patients who are blind due to photoreceptor degeneration (Potts and Buffum, 1968; Kato et al., 1983; Humayun et al., 1996, 1999). Since the mid nineties several research groups have been working towards the goal to restore basic vision by means of multi-focal electrical stimulation using an implantable electronic device (Chow and Chow, 1997; Shimazu et al., 1999; Zrenner et al., 1999; Grumet et al., 2000; Eckhorn et al., 2001). Besides challenges in the areas of implantability, biocompatibility, and tissue safety, the main task within the development of a

retina implant is to provide useful visual percepts. Decisive questions address the achievable temporal, intensity, and spatial resolutions by means of electrical retina stimulation. Final answers can only be given by blind patients in direct dialogue with the researcher adjusting stimulus parameters. However, apart from pioneering key studies with humans (Humayun et al., 1996, 1999), animal experiments are to be preferred at the current state of research for ethical reasons and may contribute considerably to the understanding of the visual system.

In our investigations we used the anesthetized cat as the choice animal model (cf. Dawson and Radtke, 1977; Hesse et al., 2000). Its visual system is well known and relatively similar to that of primates as regards size and structure (Orban, 1984). While electrically stimulating the cat retina with multiple microelectrodes, we recorded evoked neuronal activity mainly from the primary visual cortex (area 17). Perception is based on activity in the early visual areas (Doty, 1973). With respect to perception, analyses of electrically evoked activations by means of a retina implant may serve for estimating achievable temporal, intensity, and spatial resolutions. Concerning precision, this approach at first applies to the utilized animal model, but because of the similarity between the cat and human visual system also to human patients (Eckhorn et al., 2001).

The following four characteristics are necessary requirements for successful retinal stimulation: (1) *efficiency*, i.e., low stimulation currents elicit perceptually relevant cortical activity, (2) *reliability*, i.e., a sufficiently low noise level provides reproducible activity, (3) *selectivity*, i.e., cortical activity patterns reflect specific stimulus properties, and (4) *stability* with respect to stationary selectivity. To some degree the first point has been addressed by Humayun et al. (1999); Zrenner et al. (1999); Hesse et al. (2000); Stett et al. (2000); Eckhorn et al. (2001) but systematic analyses are still needed, especially regarding perception. The stability issue (4) requires long term measurements based on functional implanted devices and is presently out of scope. It may be analyzed together with plasticity, a factor related to cortical reorganization, that might benefit the patients decisively.

In the current investigation we focus on the second and third point: reliability and selectivity of evoked cortical activity. Both are closely related to the concept of resolution. Here, we examine the results of $N = 5$ experiments with anesthetized cats that were aimed at yielding reliable values of temporal, spatial, and intensity resolutions theoretically achievable by a retina implant. We apply information theory to electrical stimulus evoked responses as a useful tool to quantify achievable resolutions. The transinformation T between stimulus and response is a plausible measure for describing the reduced uncertainty about the stimulus after registering the response. Given in bit, it reflects the equivalent minimum number of yes–no questions which – after being answered correctly – suffices to quantify the change in uncertainty.

Since information theory considers both the deterministic properties of the analyzed system and the noise that unifies all non-deterministic properties and under-

mines the information transfer, it is very useful for the estimation of resolutions. For example, we wanted to know how many different stimulation intensity steps are at most discriminable by analyzing the evoked activity. It is clear that this value depends on both, the discriminability of the mean responses and likewise on the variability of the single responses. In case the mean responses cannot be distinguished, even the lowest response variability is not sufficient to facilitate information transmission. On the other hand, if the mean responses differ significantly, but the variability is large, a single response reveals little about the stimulus and hence, not much information can be transmitted, either. As long as the response range is unlimited, one can always separate the stimuli in a way, say, a dozen times of the response parameter's standard deviation to ensure that individual responses are well discriminable. However, in reality the response range is closely bounded. There is no room for many broad response distributions. Here, resolution with respect to a stimulation parameter such as intensity can be defined as the maximum number of quantization steps whose response distributions fit in the limited response range and transmit more information than would be possible with less quantization steps.

For the estimation of temporal resolution we follow different schemes. First, we look at the timing precision of single unit action potentials. Secondly, we vary the stimulation rate in search of the rate that maximizes T. The quantification of spatial resolution is more difficult because the distance between electrodes cannot easily be adjusted during an experiment. However, having several retinal stimulation and cortical recording electrodes available, we analyze the spatial transinformation profile of cortical activity with and without taking redundancy, i.e., correlation between recording channels into account. We hope that the values of achievable resolutions provided in this paper may serve for the improvement of the conceptual design of future retina implants. Part of the results in this chapter has been published as conference abstracts (Eger et al., 1999; Eger and Eckhorn, 2001a).

4.2 Methods

Surgery and preparation. The animal experiments were conducted in accordance with the guidelines of the European Communities Council Directive (86/609/EEC) and the NIH Principles of Laboratory Animal Care (Publication 86-23, revised 1985). A detailed description is found in (Hesse et al., 2000).

Electrodes and electrode positioning. Fiber electrodes (80 µm shaft diameter, see Reitboeck, 1983) were used in different designs for stimulation and recording. For stimulation, coarse tips were fabricated by grinding (metal cone of 20–25 µm diameter and 20–30 µm height; ≈100 kΩ impedance at 1 kHz) allowing safe charge injections up to 20 nC per impulse polarity (Brummer and Turner, 1977). The retinal electrodes were axially tractable by means of a multiple fiber electrode manipulator (Eckhorn and Thomas, 1993) with concentric bundles (diameter $<$1.2 mm) of 3 or 7 electrode guide tubes. After introducing the guide tubes through a small incision

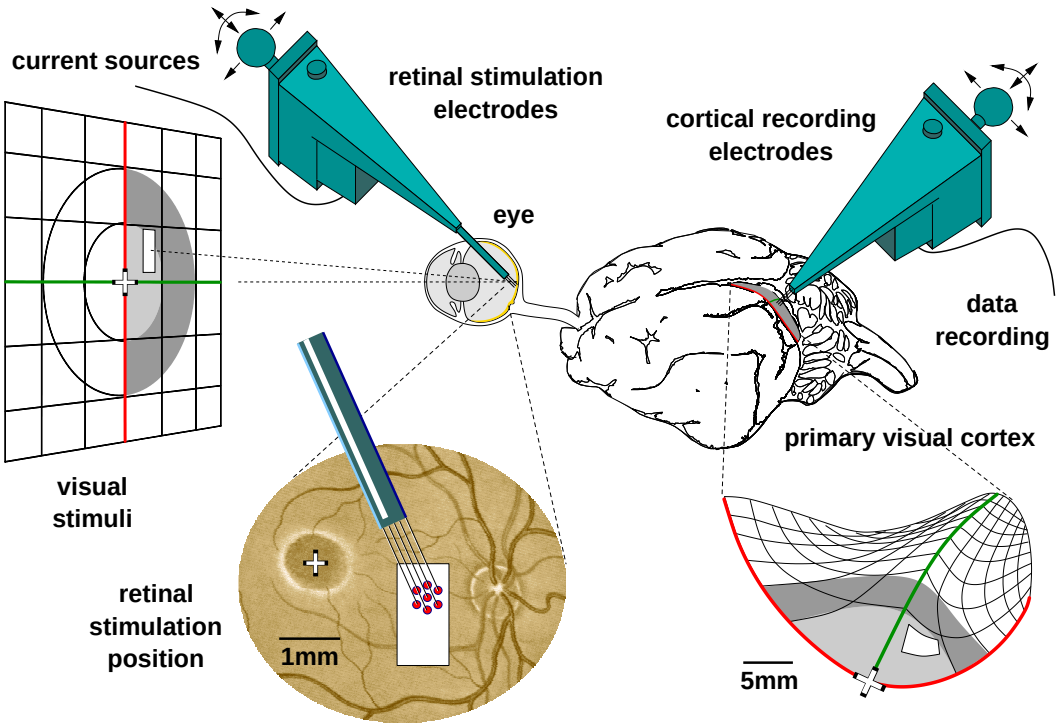

Figure 4.1: Positioning of cortex and retina multi-electrodes at retinotopically corresponding sites. After the adjustment of the cortical electrode positions within area 17, the cortical RFs are plotted. Then, the retinal stimulation electrodes are moved iteratively to retinal locations while back projecting the stimulation electrode tips onto the tangent screen.

of the left or right eye about 4 mm behind the limbus[1], the retinal electrodes were moved to positions corresponding to the receptive fields (RFs) of the recorded cortical neurons by means of a specially designed spherical manipulator (Schanze et al., 1998). Close contact of the retinal electrodes was verified by direct sight through the ocular media and by recording of retinal activity via the stimulation electrodes. The fiber electrodes for intracortical recordings had fine tips (cone diameter and height in the µm range) with impedances of 2–3 MΩ. Individual, computer controlled axial positioning of the cortical electrodes was obtained by a 7 or 16 fiber electrode manipulator (Eckhorn and Thomas, 1993). Fig. 4.1 shows the scheme of the experimental setup.

Recorded signals. Cortical recordings were normally performed in area 17 (Horsley-Clarke A2 to P7, L0.5 to L3) with up to 16 electrodes. Sometimes the linear electrode array also covered parts of area 18 (A5 to P4, around L1). The extracellularly registered neuronal activity usually yielded small receptive fields ($<2^\circ$) with simple cell properties (Orban, 1984). We restricted our analysis to these recording positions (except for experiment [140], see Fig. 4.2). Three types of signals were

[1] i.e., the borderline of the sclera turning into the cornea

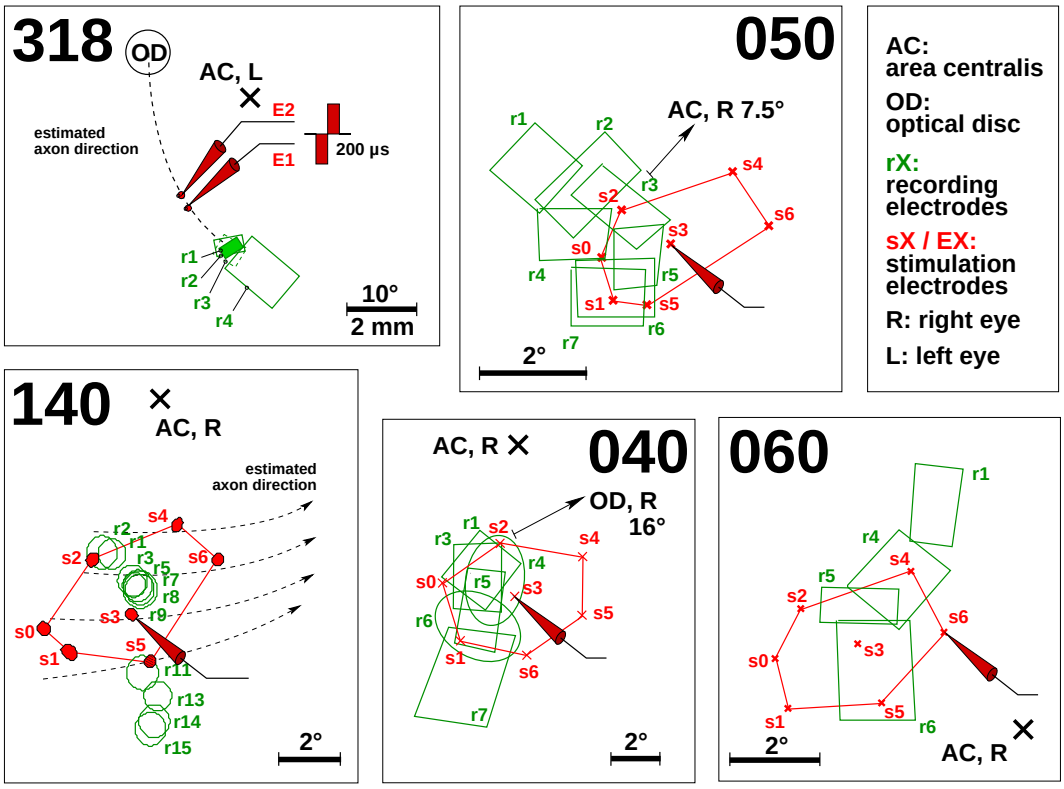

Figure 4.2: Receptive field (RF) plots of the five experiments investigated. The large numbers are the identifying keys for the respective experiments and are referenced throughout the paper. In addition to the cortical RFs marked in green, the positions of the retinal stimulation electrodes projected on the screen are marked in red. The recordings were performed in area 17 and only for [140] in area 18. The small size of the RFs of [140] is due to the threshold criterion implemented in the computer controlled RF calculation (Eckhorn et al., 1993; Wilms, 2001). The RFs of the other experiments were plotted manually be means of a hand-held projecting device. In all cases stimulation currents were charge balanced biphasic impulses with 200 µs phase width and negative leading phase. In experiment [060] bursts of 5 impulses with 200 µs inter-impulse delay were used, otherwise single impulses.

extracted either by on-line filtering or off-line from broad band recordings (BBR) at 20 kHz:

SUA *single unit activity*, after applying a bandpass 0.5–10 kHz, -3 dB at 12 dB/oct. and subsequent window discrimination with 2 ms hold-off,

MUA *multiple unit activity*, after applying the same bandpass as for SUA, the signals were full wave rectified and then lowpass filtered at 140 Hz, -3 dB at 12 dB/oct., and

LFP *local field potentials*, resulting from bandpass filtering 10–140 Hz, -3 dB at 12 dB/oct.

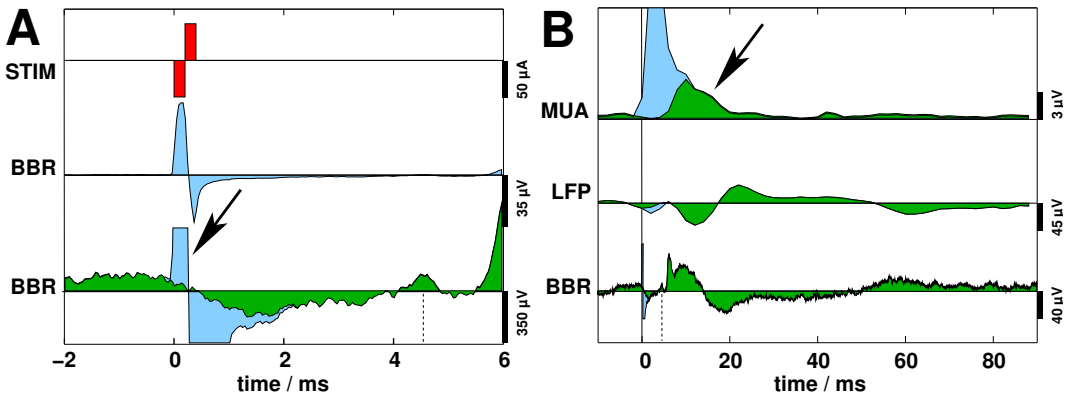

Figure 4.3: Example to illustrate electrical stimulus artifact rejection on two time scales (A, B). The data stem from experiment [318], stimulation with electrode E1 at 100 µA (Fig. 4.2). The blue curves are averages of the unprocessed signals (N=100), the green curves are the result after artifact rejection. Note the strong but temporally localized (clipped) artifact in the BBR signal (A, arrow). In contrast to the LFP signal, MUA is very sensitive to stimulus artifacts due to its bandpass properties (B). Only after artifact cancellation in BBR and subsequent off-line generation of MUA the neuronal activity can be distinguished from the artifact (B, MUA, arrow).

The SUA signal consists of a sequence of event time stamps, whereas MUA and LFP are continuous signals sampled at 500 Hz (CED 1401Plus, Cambridge Electronic Design). All signals including the stimuli were stored on a hard disk for off-line data evaluations.

Stimulation. Visual stimulation was performed using a computer monitor with high frame rate (101 Hz). The receptive fields (RF) were measured quantitatively, either with a hand-held projection device or by means of the computer monitor presenting single spots of light or m-sequences for multi-focal stimulation (Eckhorn et al., 1993; Wilms, 2001). Electrical stimulation was controlled by a computer attached to a second CED 1401Plus that generated zero mean voltage waveforms to feed subsequent fast voltage-current converters. The current sources were equipped with preamplifiers to allow simultaneous stimulation and recording with the same retinal electrode and thus, served for measurements of electrode–tissue impedances in situ. Stimulation currents consisted of sequences of biphasic charge balanced impulses with negative leading phase, 200 µs phase duration and amplitudes ranging from 5 to 100 µA. The positions of the stimulation electrodes for each conducted experiment are given in Fig. 4.2. In all cases the large and distant reference electrode is the ground.

Artifact rejection. Apart from the neuronal activity, the recording electrodes register an electrical stimulus artifact due to the voltage drop caused by the current injection (cf. Freeman, 1971; Roby and Lettich, 1975; Eger et al., 1998; Parsa et al., 1998). As described above, cortical response signals (SUA, MUA, LFP) were

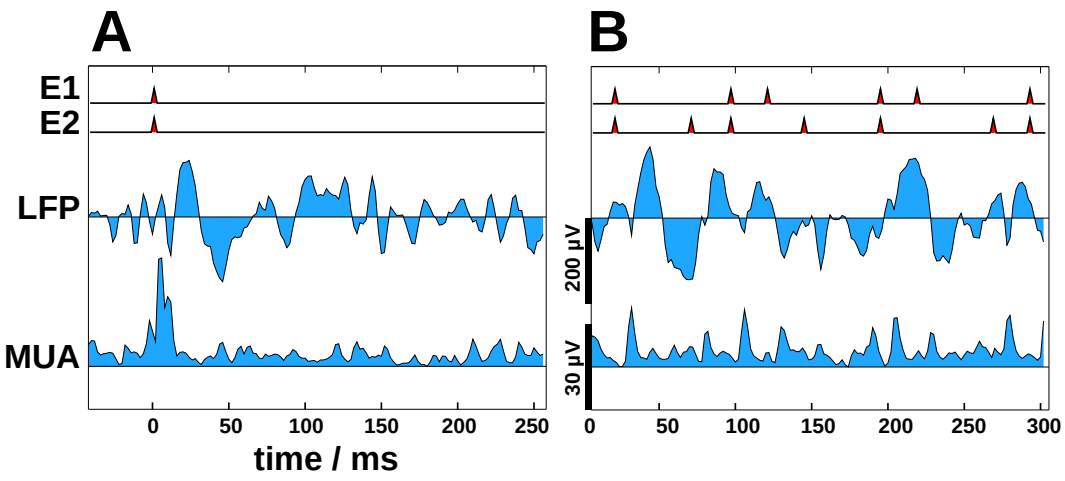

Figure 4.4: Example of non-averaged local field potential (LFP) and multi unit activity (MUA) responses evoked by **A**: transient and **B**: continuous electrical stimulation. In this experiment the cat retina was electrically stimulated at two positions 400 µm apart using biphasic 100 µA current impulses (Fig. 4.2 [318]). The stimulus events are depicted as small triangles at the top.

mostly generated off-line, based on broad band recordings at high sampling rate (20 kHz). Thereby, the stimulus artifacts could easily be identified and canceled in the broad band recorded data because their exact times of occurrence were logged by means of a trigger file. In practice, we assigned a range typically beginning at stimulus onset and ending 2 ms after the stimulus. We replaced the artifact impaired signal course by an artificial signal with nearly identical statistical properties compared to the signal directly before and after the range. After that, we corrected a potential slope caused by capacitive effects (see Fig. 4.3).

Calculation of transinformation. The theoretical background for the methods used has been presented in detail elsewhere (Eger and Eckhorn, 2002a,b,c, cf. Chapter 2). This paper merely gives an overview of our approach. For the calculation of the transmitted information T over the multi-input and -output channel, we follow an indirect forward approach taking the perspective of the information flow. Indirect methods use a model of the undisturbed information channel and have the advantage of being applicable even if only small amounts of stimulus response data are available.

In our analyses we distinguish between two popular stimulation paradigms: series of discrete transient stimulus events (*DS*) and continuous stimulation (*CS*). Fig. 4.4 illustrates the conceptual differences between the stimulation paradigms. While diagram A shows a single stimulus event and its corresponding population (LFP and MUA) response, B displays a rapid sequence of stimulus events whose corresponding responses temporally interfere with each other. Both stimulation paradigms are typically paired with appropriate analysis methods. In the case of

DS, usually an inter-stimulus delay is chosen that is longer than the memory of the system, i.e., the response signals are merely affected by directly preceding stimuli. Hence, serial correlations between different stimulus events do not need to be considered. With regard to the discrimination of responses to discrete independent stimuli, well-known methods are available for the estimation of *T* (Werner and Mountcastle, 1965; Stein, 1967; Optican and Richmond, 1987).

Whereas *DS* merely requires the independent discrimination of relatively few stimulus conditions, e.g., different stimulation intensities or positions, for *CS* serial correlation has to be taken into account, which substantially increases the complexity of the task to quantify *T*. Here, we calculate *T* by means of a newly developed method which is somewhat related to the stimulus reconstruction technique (Bialek et al., 1991; Rieke et al., 1998). This *three step method* (Eger and Eckhorn, 2002b, cf. Chapter 2) first requires a continuous deterministic response estimate, which we typically compute by convolving linear stimulus–response kernels with the stimulus signal. Instead of quantifying *T* between the stimulus and its response directly, we replace the stimulus by the deterministic response estimate, defining its deviation from the original response as noise. As a second step, a coordinate transformation by means of principal component analysis (PCA, Glaser and Ruchkin, 1976) is performed in order to achieve independent coefficients contributing additively to *T*. As long as nonlinear correlations may be neglected and approximate normality of the coefficients is ensured, the summation of partial transinformation values *PT* on the basis of individual signal-to-noise ratios is conservative (Shannon, 1948; Rao and Gabr, 1980; Hinich, 1982; Eger and Eckhorn, 2002b). The method is applicable to multi-input and multi-output systems. The latter is achieved by serially aligning corresponding epochs of the recording channels before PCA is performed. Thus, an analysis of redundancy in the evoked neuronal activity registered by multiple electrodes is made possible.

Apart from the total value we are also interested in *T* based on specific stimulus aspects in case of *CS*. E.g., we want to assess the spatial and intensity resolution at higher stimulation rates under the potential constraint of mutually interfering responses to consecutive stimulus events. For this reason we have developed a method originally based on the one presented by Optican and Richmond (1987); Richmond and Optican (1987). It compactly characterizes response epochs following the stimuli by means of coefficient vectors yielded by a modified principal component analysis (Eger and Eckhorn, 2002a, cf. Chapter 3 on p. 50). The method provides a measure of *T* solely based on the discriminative power of the system with respect to transient stimulus patterns.

4.3 Results

4.3.1 Discrete stimulation

Temporal aspects of spike activity. Electrical stimulation of retinal ganglion cells shortcuts the relatively slow information processing in the deeper retinal layers (Potts and Buffum, 1968). Thus, we may expect to transmit more temporal information and achieve a higher temporal resolution compared to natural visual stimulation. In the following we will investigate the temporal precision of electrically evoked spike activity.

Fig. 4.5 shows a raster plot of action potentials recorded from area 17 (Horsley-Clarke A0.5, L1, simple cells, small RF) and sorted into classes according to the height of the spikes. The plot reveals very precise responses at about 6 ms following the 100 µA current stimulus using two electrodes E1&2 (see Fig. 4.2 [318]). The earlier cluster of events at 1.5 ms is due to the electrical stimulus artifact which is revealed by the broad band recorded signal (see Fig. 4.3 A, same experiment). Directly after the first activation a short period of inhibition follows before a second cluster of spike events occurs at about 8 or 9 ms. In comparison, the zero stimulus Ø does not reveal any structure in the spikes.

In Fig. 4.6 we have quantified the spike precision by means of Gaussian fits dependent on the stimulation electrode configuration and the stimulation current (see also Fig. 4.2 [318]). The mean delay amounts to about 6 ms with a standard deviation even less than 300 µs. There seems to be a tendency for a shorter delay and higher precision with increased currents, though this needs to be tested on the basis of additional data. This kind of precise and structured spike response has been typical in our experiments. A prerequisite is that the positions of stimulation and recording electrodes projected into the visual space are sufficiently close. Alternatively, – this is the case here – the stimulation electrodes excite *axons* of retinal cell bodies retinotopically congruent with the cortical recording positions (cf. Discussion on p. 95).

Additional to the precision, we wanted to know whether there was a dependence of the occurrence of spikes. Fig. 4.7 depicts the occurrence of all spikes within the interval of 3–10 ms following the stimulus. Second and later spikes after 5.3 ms (dashed line) are superimposed (red). An activation within the second cluster always (except of 2 events, left column) was preceded by a spike belonging to the first cluster in case of nonzero stimulus condition. However, this is not astonishing when one realizes that the first cluster yields a spike nearly at each stimulus event except the Ø stimulus. The temporal scatter of the relative delay between an activation in the first and second cluster is nearly identical with the variability of the second cluster (Figure not shown).

Temporal aspects of population activity. The first row in Figures 4.8, 4.9, 4.10 gives an impression about the variability of the evoked LFP, respectively MUA re-

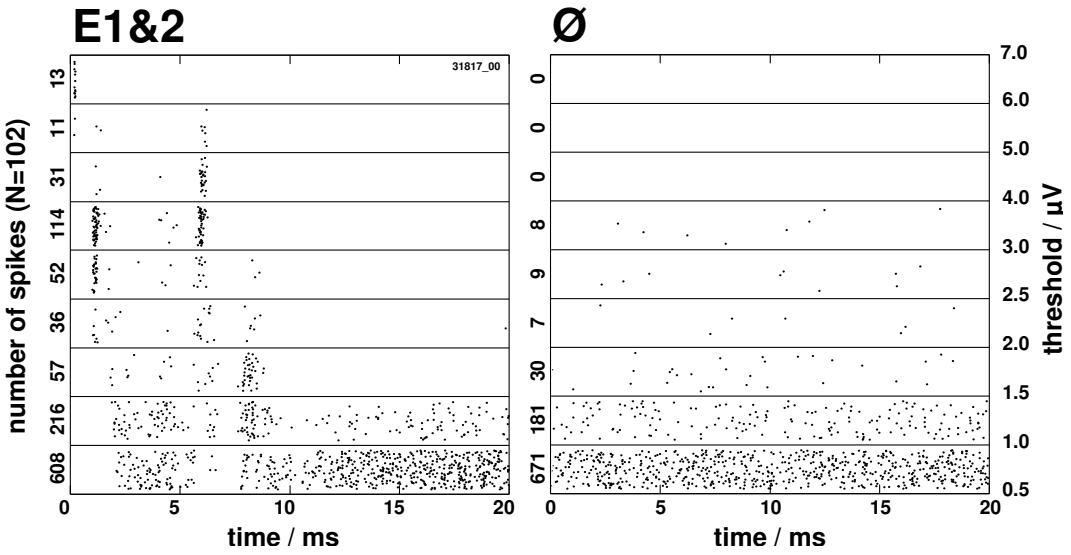

Figure 4.5: Evoked action potential responses for stimulus conditions E1&2 (synchronous stimulation) and Ø (zero stimulus) according to Fig. 4.2 [318] with respect to the size of the potentials. Note the lack of structure in case of stimulus condition Ø.

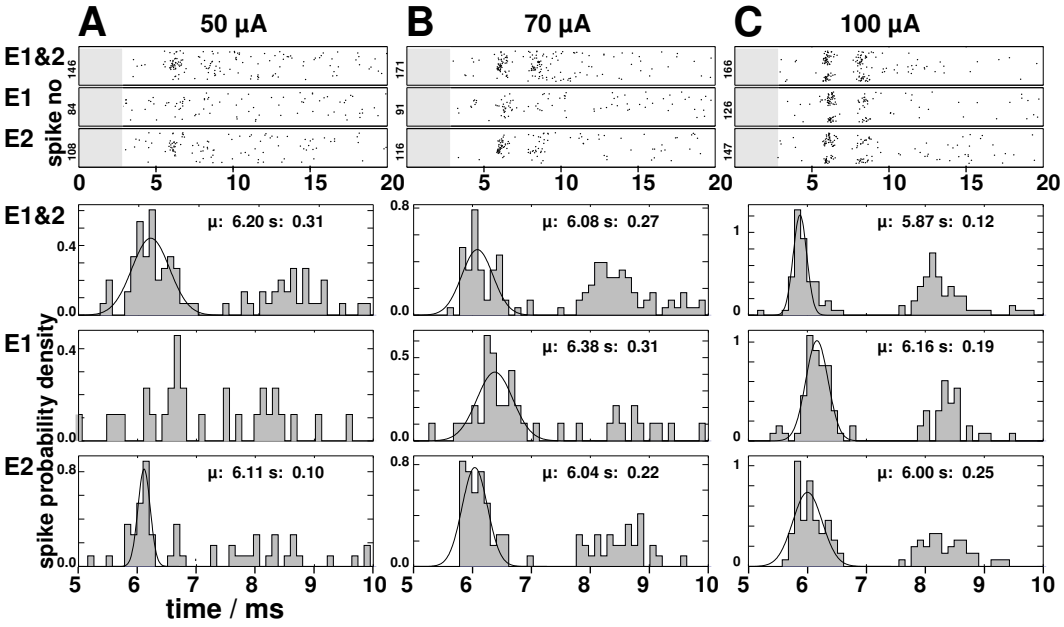

Figure 4.6: Distribution of action potentials dependent on stimulation current (**A**: 50 µA, **B**: 70 µA; **C**: 100 µA) and electrode configuration E1&2 (see Fig. 4.2). Upper panels: Raster plots. The shaded area indicates a range excluded due to the stimulus artifact. Lower panels: Peri-stimulus time histograms and Gaussian fits. The histograms are normalized to unit area within the interval of 3–20 ms. The fitting parameters are printed within the diagrams.

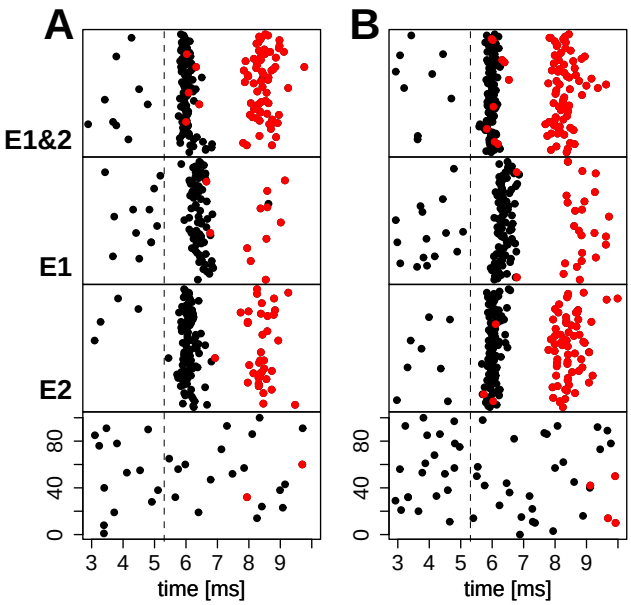

Figure 4.7: Rasters of times of occurrence of all and superimposed second and later spikes (red) after 5.3 ms (dashed line) in two recording sessions of [318]. The inter-stimulus delay was 45 ms and 35 ms for A and B, respectively. The panels correspond to stimulus conditions E1, E2, and E1&2 and Ø (bottom). The rasters show that the second cluster of spikes at 8.5 ms consists only of second and later events (except of two spikes of 270, for left column), i.e., any activation is preceded by a spike belonging to the first cluster at 6 ms.

sponse on the basis of the two experiments [318, 040] and different stimulus conditions or recording electrodes. The diagrams refer to electrical stimulation except B, D of Fig. 4.10. All individual response epochs (black) and their average (white) are plotted superimposed on each other ($N = 100$). Apparently, diagrams A and D in Figures 4.8, 4.9 and A in Fig. 4.10 reveal relatively low variability compared to the others. We quantified the time dependent response strength and variability in terms of partial transinformation values *PT* in the middle row (see Appendix 4.5.2). The *PT* estimate the information capacity per time bin assuming a linear channel and a time dependent signal-to-noise ratio that relates the mean response and the deviation from it. Intuitively the *PT* indicate when the evoked response is most reliable. Thus, their temporal course reveals important additional information compared to a conventional peri-stimulus time histogram. In case of *electrical* stimulation the *PT* rise abruptly at about 10 ms following the stimulus. Especially for LFP they form an attenuated oscillatory temporal profile with a basic frequency of about 50 Hz. However, the later activation peaks convey relatively little additional information. Fig. 4.8, bottom row, depicts the residual *PT* after linearly correlated signal fractions have been removed by means of Schmidt's orthogonalization procedure (see Appendix 4.5.3 and Eger and Eckhorn (2002c)). A response carries between

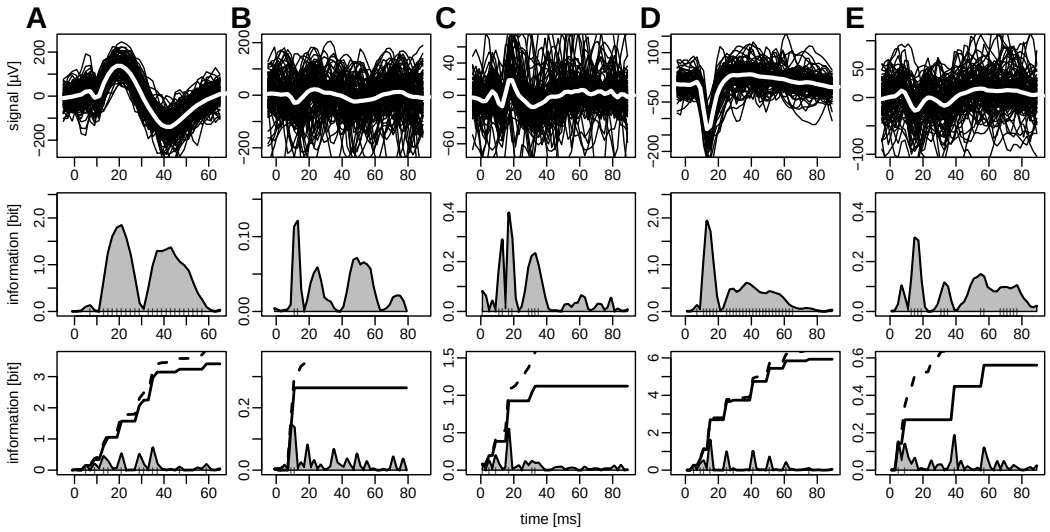

Figure 4.8: Examples of typical LFP responses to electrical stimulus impulses of 100 µA, $N = 100$. **A**: Simultaneous stimulation with two electrodes E1&2, 400 µm apart, 65 ms inter-stimulus delay, [318]. **B**: Stimulation with one electrode E1, 115 ms inter-stimulus delay, [318]. **C**, **D**, **E**: One stimulation electrode, 170–250 ms uniform inter-stimulus delay, [040]. Responses C and D have been evoked from the same stimulation electrode s5, yet were simultaneously recorded from two different cortical electrodes 1 mm apart (r1 and r3). D and E have been registered at the same recording electrode r3, but evoked by different retinal stimulation electrodes s5 and s3. Top row: Single trials and average (white). Middle row: *PT* without taking serial correlation into account. Bottom row: Residual partial and cumulative information after correlated signal fraction has been removed. The small markers indicate significant ($p \leq 0.05$) values. Dashed lines indicate the cumulative profile constrained to the significant contributions.

0.1 and 2.0 bit per stimulus event but most information is transmitted within the first 40 ms following the stimulus (exceptions are Fig. 4.8 D and Fig. 4.10 A). After that, the cortical activity does not reveal considerable information about the stimulus. The *visual* stimulation example (Fig. 4.10 B, D) shows a relative large degree of variability and a late response at 40 ms delay. The *PT* of the decorrelated LFP (B, bottom) reveals three prominent equidistant peaks.

Intensity aspects. For illustration Fig. 4.11 depicts the mean responses of four different signal types (MUA, LFP, BBR, and SUA) to two electrical stimuli that differ with respect to the intensity, i.e., the stimulation current. The mean response characterizes the deterministic behavior of the system. Distinguishable mean responses to different stimuli may indicate that the neurons encode the varied stimulus aspect. In Fig. 4.11 each time bin of the average LFP and MUA signal is provided with an error bar as a measure of the response variability. For information theoretical analyses both the mean response and the variability have to be taken into account. The more variable the neurons react to a specific stimulus, the less reliable is the

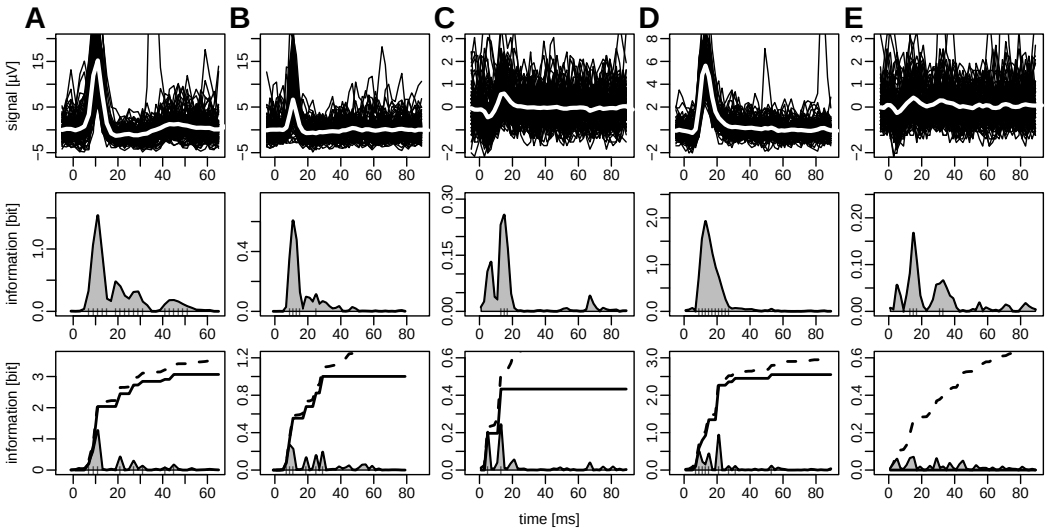

Figure 4.9: Examples of typical MUA responses and transinformation profiles at the same stimulation and recording configuration as presented in Fig. 4.8.

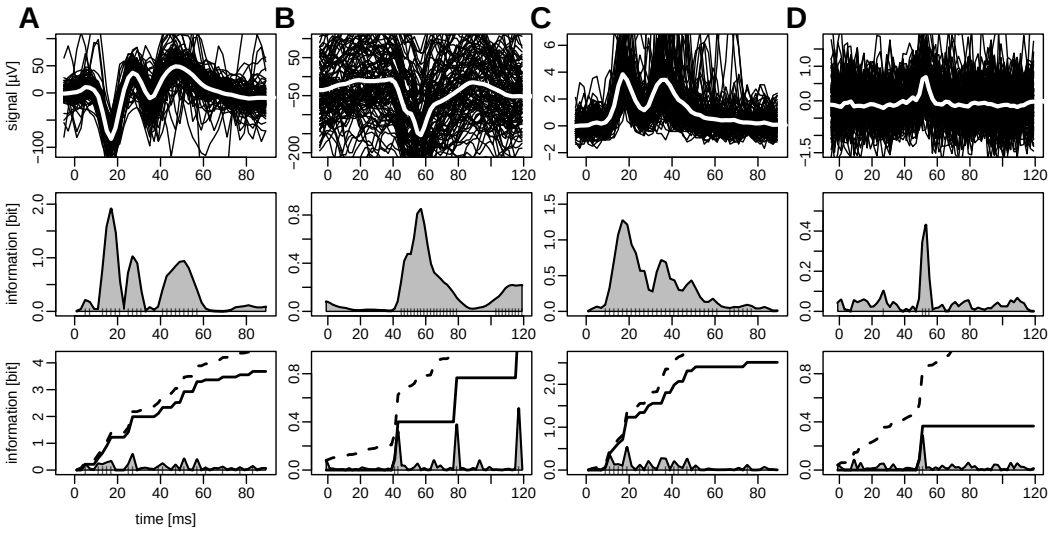

Figure 4.10: Comparison between electrical and visual stimulation using examples of LFP (A, B) and MUA (C, D) responses and transinformation profiles at identical stimulation and recording positions, [040]. Electrical stimulation (A, C) was performed with single 40 µA current impulses ($N = 100$) applied by electrode s3 (Fig. 4.2). The preceding inter-stimulus interval amounted to 300 ms (conditional analysis). The stimulus intervals were Gamma distributed (mean interval 100 ms, see Fig. 4.16). Visual stimulation (B, D) was performed at the same position using a bright spot of light presented by a computer monitor with 100–150 ms inter-stimulus intervals (see detailed description on p. 83). The visual stimulus evoked delayed response deflections (40 ms) compared to the electrical stimulus (<10 ms). The cumulative transinformation in response to the electrical stimulus (LFP, A) saturated late at 60 ms.

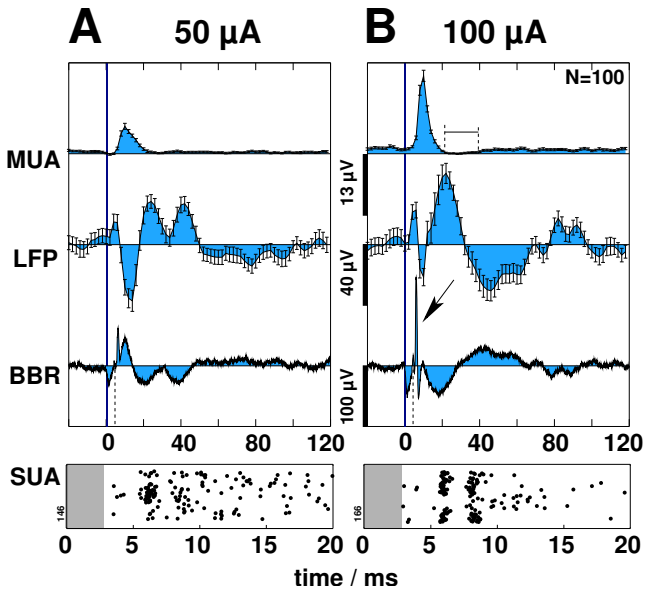

Figure 4.11: Example of averaged response signals evoked by biphasic current impulses with amplitudes 50 µA (A) and 100 µA (B), experiment [318] (see Fig. 4.2, two active electrodes E1&2). Top: After a sharp peak of MUA a time of inhibition follows (reduced activity between 20 and 40 ms, only in B). Middle, upper panel: Stimulus dependent LFP deflections. Error bars in MUA and LFP indicate the standard deviation of the mean. Upper panel, bottom: Broad band recorded signal (BBR). Note the needle like peak at about 6 ms following the stimulus (arrow, B). It reflects the high temporal precision of evoked action potentials. Bottom panel: More or less structured patterns of action potentials. Note the different time scale.

response and hence, the less information can be transmitted. The variability, i.e., the deviation of the single responses from the mean is often defined as noise in the sense that it is uncorrelated with the stimulus. However, it is clear that it reflects the unexplained variance of ongoing neuronal activity (Arieli et al., 1996; Tsodyks et al., 1999).

In the following we systematically investigated the discriminability of cortical activity regarding different stimulus intensities. We varied the current amplitude – most probably in correspondence to a change in size of the stimulation field surrounding the electrode tips (Ranck, 1975). Another possibility for the modulation of stimulation strength consists in the use of bursts with varied length (Brindley and Lewin, 1968). The average and the variance of the responses is critical concerning the discrimination of stimuli since the response range is limited (Stein, 1967; Gershon et al., 1998). As regards analyses of single unit activity, often the mean–variance relation of the spike count is investigated. It usually reveals a linear relation between the logarithms of mean and variance (Dean, 1981; Tolhurst et al., 1981). Here we analyze the mean–variance relation for electrically evoked population activity responses.

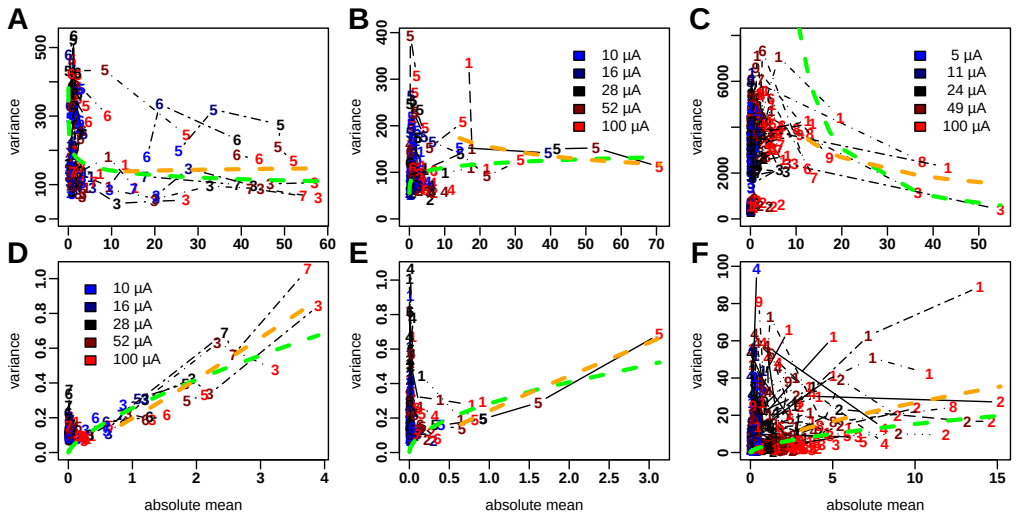

Figure 4.12: Mean–variance relation in electrically evoked LFP (top) and MUA (bottom) responses in case of three different experiments [040] (A, D), [050] (B, E) [140] (C, F). We performed an interleaved electrical stimulation (140–250 ms inter-stimulus delay) with 5 different current amplitudes (see inset) at 7 retinal positions according to receptive field plots in Fig. 4.2. Simultaneously, the cortical activity was recorded with 7 [040, 050] respectively 15 [140] fiber electrodes. For each of the 35 stimulus conditions we applied PCA to the responses and plotted the mean vs. the variance of the first coefficients. The time interval on which PCA was applied was chosen to cover the relevant response deflections and was determined by observing the mean responses. We mainly chose a 60 ms window following the stimulus and were careful to exclude a stimulus artifact if it existed (only in [140]). The numbers indicate the recording electrodes, the colors the current amplitude, and the line styles the stimulation electrodes. The dashed green lines show the result of an indirect nonlinear fit between the mean m and the variance s^2 using the model $s^2 = \frac{1}{a^2} m^k$.

Fig. 4.12 shows the mean vs. the variance of LFP (top row) and MUA (bottom row) multi-channel responses to interleaved electrical stimuli applied by 7 electrodes for three experiments ([040] A, D; [050] B, E; [140] C, F). Each of 35 stimulus conditions (5 current intensities, 7 stimulation electrodes) and cortical recording channels (7 in A, D and B, E; 15 in C, F) was handled separately. Every stimulus condition was repeated $N = 100$ times. We applied PCA to the N repeated responses for each combination of stimulus condition and recording electrode. Finally, we plotted the mean vs. the variance of the first principal component coefficient as shown in Fig. 4.12. The dense clustering at small response strengths indicates either ineffective stimulation or the possibility that the evoked cortical activity occurred outside the range of the recording electrodes. In order not to obscure the mean–variance relation by this dot cluster, we did not present the data on a logarithmic scale as otherwise usual. While the variability of MUA seems to be growing with increased response strength the LFP variability does not grow. A

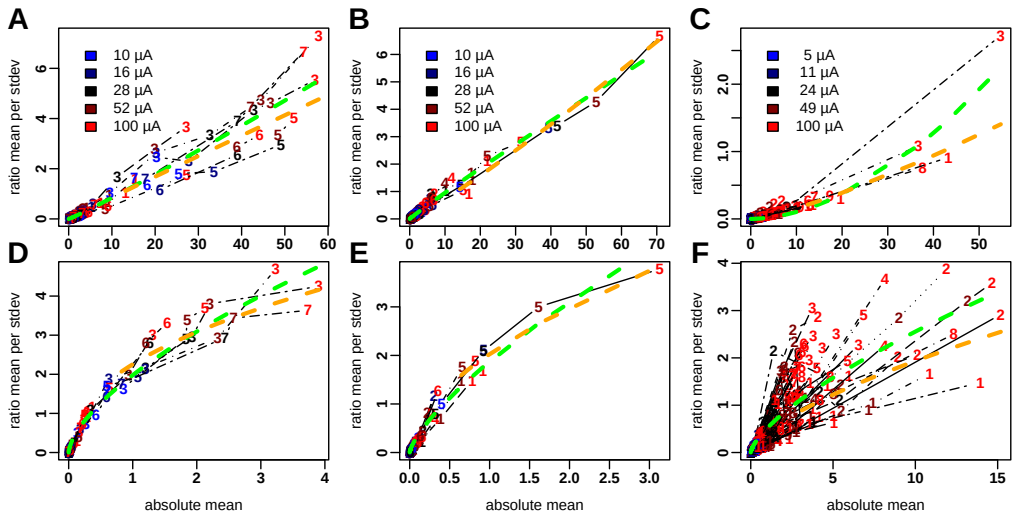

Figure 4.13: Mean vs. mean normalized by the standard deviation of the LFP (top) and MUA (bottom) responses shown in Fig. 4.12. The dashed green line indicates the result of the nonlinear regression. Obviously the relation is linear for LFP and sublinear for MUA indicating an increased variance with growing mean. Experiment [140] (C, F) does not contradict the tendency. However, due to the higher variability the result is not as reliable as that of [040] (A, D) and [050] (B, E).

Table 4.1: Response range R, the best value of the normalized standard deviation s/R, and the exponents of the nonlinear model.

	[040]		[050]		[140]		[060]	
	LFP	MUA	LFP	MUA	LFP	MUA	LFP	MUA
R	57.8	3.90	70.8	3.13	54.7	15.2	50.1	2.51
s/R	0.181	0.210	0.163	0.231	0.446	0.293	0.195	0.254
k	-0.141	0.713	0.095	0.531	-1.614	0.610	-0.761	0.485
$k_{80\%}$	0.042	1.102	-0.246	0.884	-0.562	0.675	-1.378	0.662

Data from 4 experiments and for the two signal types LFP and MUA. The exponent k is the result of an indirect nonlinear regression between the mean m and the mean normalized by the standard deviation m/s. $k_{80\%}$ results from a direct fit of the variance s^2 based on 80% of the data omitting the weakest 20% responses due to inefficient stimulation.

few cases, i.e., combinations of stimulus and recording electrodes, show extremely small variability and systematic structure.

To quantify the relation between mean m and variance s^2 we applied the following nonlinear regression to the data

$$\frac{m}{s} = a \cdot m^b + \varepsilon \tag{4.1}$$

to first determine b. The relation between m and m/s and the corresponding fit is depicted in Fig. 4.13. The fit between m and m/s is advantageous compared to the fit between m and s^2 directly because it is not obscured by inefficient stimulations (clusters in Fig. 4.12). Neglecting the error ε and rewriting of (4.1) yields the model

$$s^2 = \frac{1}{a^2} m^{2(1-b)} = \frac{1}{a^2} m^k \tag{4.2}$$

with $k = 2(1-b)$. The dashed green curve in Fig. 4.12 shows the same fit as in Fig. 4.13 but projected into the mean–variance plain. Table 4.1 summarizes the results of the fit, including the exponent k, the response range R which we define as the largest mean response and the inverse of the fit value m/s at $m = R$. We will refer to this ratio as the best value of s/R and later use it for the estimation of the maximum possible number of discriminable stimulus intensities. Alternatively we also performed a direct fit of the variance omitting 20% of the data with the lowest response strength to avoid the inefficient stimulations. The exponents $k_{80\%}$ are given in Table 4.1. The corresponding fit is indicated by orange regression curves in Figures 4.12 and 4.13.

Table 4.1 also presents the results of an additional experiment [060] with different stimulus configuration (see caption in Fig. 4.14 A, D). In all cases the best value of s/R is for MUA slightly smaller than for LFP except for [140] which is most probably due to the large variability in this experiment. Note the obviously incorrect indirect fit for LFP (Fig. 4.12 C). Omitting experiment [140] the value s/R is approximately 23% for MUA in contrast to 18% for LFP.

A further difference between MUA and LFP is revealed by the exponents of the regression fits. Generally, the exponents k and $k_{80\%}$ are larger for MUA compared to LFP. For MUA the exponent is greater than zero and in one case even exceeds 1 indicating increasing variance with growing response strength. On the other hand, LFP yields smaller exponents mostly below zero. Obviously, the LFP variance is independent of or rather decreases with growing response strength. This systematic difference in the behavior of the response variability between MUA and LFP will be discussed on page 99.

Fig. 4.15 shows the response strength dependent on the stimulation current based on the same data as shown in Fig. 4.12. The activity increases with growing current but reduced slope leading to a half-sigmoid shape. The response range seems to be very sensitive with respect to the combination of stimulus and recording electrode. Note that in B and E the recording electrode r5 yields the strongest responses but is critically dependent on the stimulus electrode (color coded). The response variability is indicated by the size of the squares measuring $1/5^{\text{th}}$ of the standard deviation. We will later discuss how the tuning of the response strength maybe used for the selection of suitable stimulus intensity steps (cf. p. 103).

Fig. 4.14 B, E shows the same relation, though based on an experiment [060] with different stimulus configuration (see caption). Different from Fig. 4.15, here

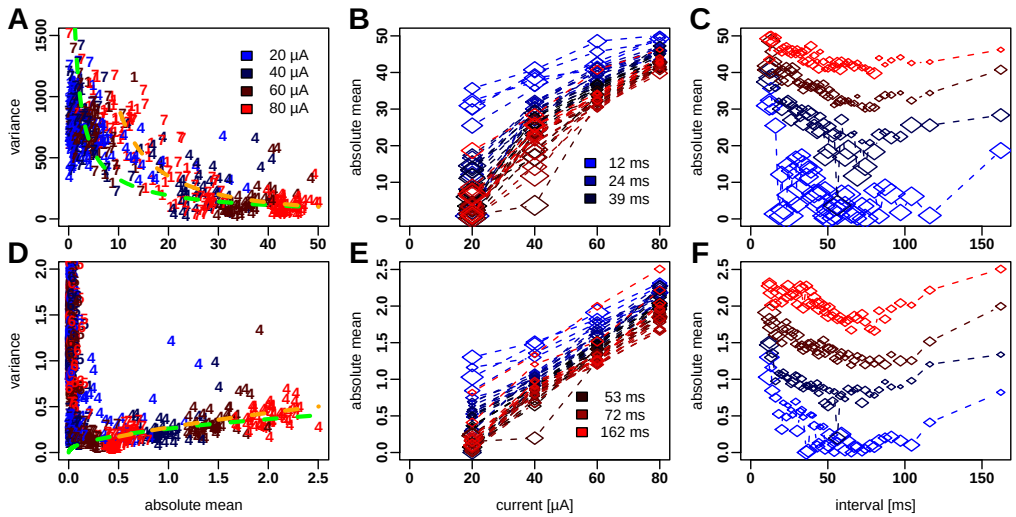

Figure 4.14: Mean–variance relation (A, D), response strength vs. stimulation current (B, E), and vs. temporal distance to previous stimulus event (C, F) of LFP (top) and MUA (bottom), [060]. Numbers indicate recording electrodes. In diagrams B, E and C, F data from only one recording electrode (r4) is depicted in order not to obscure the plot by responses to less efficient stimuli. In A, D and C, F colors code the stimulation current, in B, E they code the temporal distance (in ms). Lines connect the same recording electrodes. The size of the squares measures $1/5^{\text{th}}$ of the standard deviation. Electrical stimulation consisted of current impulses with 4 different amplitudes (inset) and Gamma distributed inter-stimulus intervals with 50 ms mean (see Fig. 4.16 D). Stimuli were supplied by one stimulation electrode (s6). The 30 ms time interval on which PCA was applied is smaller compared to that of the other experiments (cf. Fig. 4.12). According to the analyses in Figures 4.8, 4.9, a 30 ms interval is still long enough and captures most of the relevant response deflections.

the colors do not code the stimulation electrode but the temporal distance to the previous stimulus event. The bluish lines appearing at the top indicate that stimuli are most effective if they are given very shortly after a preceding stimulus. Otherwise the curves are similar to those in Fig. 4.15 revealing an increased response strength with growing stimulus intensity. LFP other than MUA saturates at highest stimulus currents. Diagrams C, F explicitly demonstrate the relation between the temporal distance to the preceding stimulus and the response strength. Here, the colors code the stimulation current ranging from 20 (blue) to 80 µA (red). Interestingly, both for LFP and MUA there is a minimum of the curves roughly at 70 ms. This indicates that a stimulus is most effective if given shortly or very long after a preceding stimulus. However, it is ineffective if the preceding stimulus was given about 70 ms ago. This seems to be an important result for the development of the implant and is going to be discussed on p. 98.

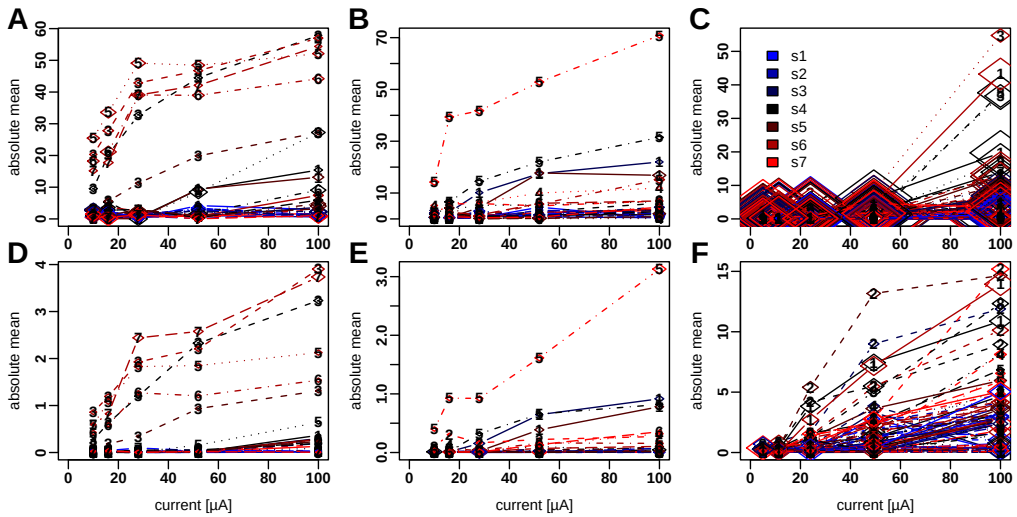

Figure 4.15: LFP (top) and MUA (bottom) response strength vs. stimulation current amplitude of the same data as shown in Fig. 4.12. The stimulation electrodes are indicated by colors, the recording electrodes by numbers as well as line types. The size of the squares measures $1/5^{\text{th}}$ of the response's standard deviation. In all cases the response strength increases with growing stimulation intensity but the variability is relatively low in [040] (A, D) and [050] (B, E) compared to [140] (C, F).

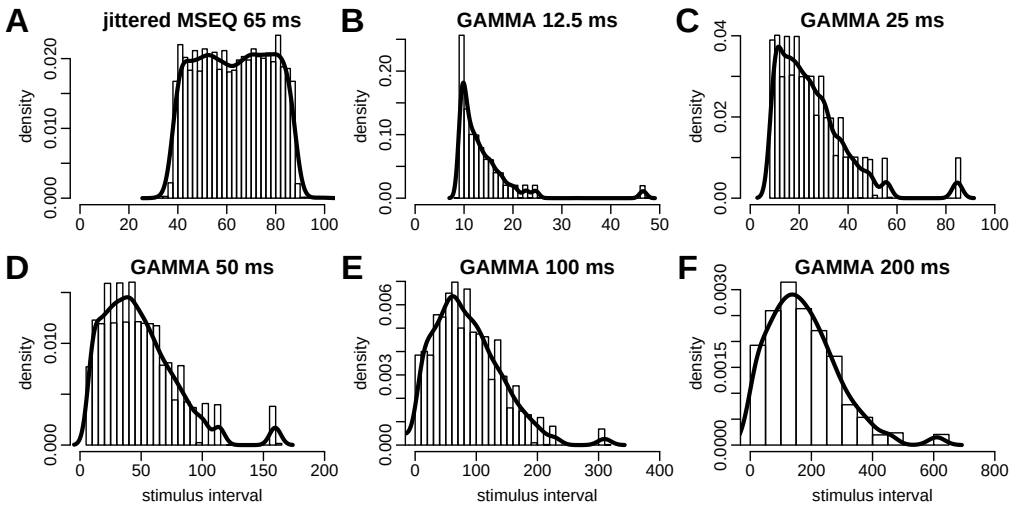

Figure 4.16: Interval distributions of applied electrical stimuli. **A**: Jittered m-sequence [040]. **B**–**F**: Approximately Gamma distributed intervals [040, 060].

Spatial aspects. For the discrete stimulation paradigm we want to refer to (Wilms et al., 2001; Wilms, 2001) and a paper in preparation. Analyses of cortical activity profiles in the continuous case are given on p. 88 and following.

4.3.2 Continuous stimulation

The data presented in this section were recorded in parallel with 5, 7, or 15 cortical electrodes in response to continuous electrical or visual stimulation (*CS*) of the cat retina in contrast to discrete stimulation (*DS*) as described in the previous section. The stimulation signal consisted of rapid single or multi-channel sequences of transient stimulus events such as short biphasic current impulses or light flashes emitted by a computer monitor. The inter-stimulus intervals usually were approximately Gamma distributed. In some cases we utilized jittered 12 bit m-sequences (MacWilliams and Sloane, 1976; Sutter, 2001) and in case of visual stimuli uniform distributed intervals. The interval distributions of the applied electrical stimuli are depicted in Fig. 4.16.

For all Figures quantifying the transinformation, we calculated estimations of the standard error based on the jackknife bias correction and error estimation procedure (Efron and Tibshirani, 1991; Efron, 1994). The resulting errors assess inconsistencies within the data and detect state changes due to instationarity. In our case the individual recording sessions were relatively short and only those recordings were chosen that had not shown a state change within the experiment (e.g., due to anesthesia). For this reason, the errors are mostly very small and often in the dimension of less than few percent. As they are hardly visible in a diagram they are simply left out for clarity reasons. The presented spectra of partial transinformation values are not that precise due to neuronal variability not explained by the utilized deterministic model. Nonetheless, all individual *PT* are depicted in the spectra regardless of their significance. Error bars are omitted in order not to obscure the qualitative shape of the spectral profiles. Yet, only significant values of *PT* are considered when summing them up to calculate total *T*. A different type of error bars that assesses the variability of different stationary conditions would have been useful. However, it would have required many repetitions of the data recordings, which was possible only in some cases because the duration of the experiments was tightly limited. Fig. 4.21 (first and last group of bars in A, C) is an example which reveals quite different results after repeating an identical stimulus about an hour later. While error bars calculated from the individual data sets are hardly visible, the deviation between the results before and after the state change is not negligible.

Temporal aspects. To assess the neuronal encoding of temporal stimulus aspects, the mean stimulus rate was varied while keeping the spatial and intensity configurations constant. We estimated the amount of transmitted information dependent on the stimulus rate according to the method described in (Eger and Eckhorn, 2002b, cf. Chapter 2). For one selected recording electrode the spectral profiles of *PT* for different frequency components are pictured in A and C of Figures 4.17, 4.18 (electrical stimulation in two independent experiments [060, 040] with different animals) and Fig. 4.19 (visual stimulation, [040]). The *PT* quantify the relation between the deterministic response fraction vs. the components uncorrelated with the stimulus.

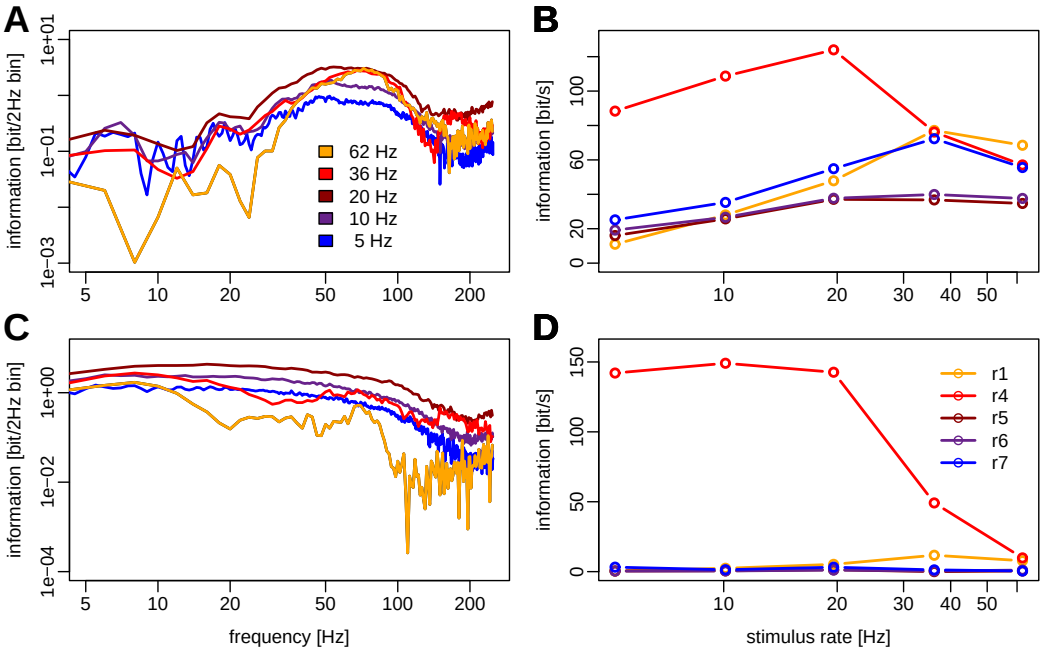

Figure 4.17: Partial and total transinformation (LFP: top; MUA: bottom) dependent on the mean *electrical* stimulus rate, experiment [060]. Stimulation with electrode s6 using Gamma distributed stimulus intervals and 40 µA biphasic current impulses. **A**, **C**: *PT* of recording electrode r4. **B**, **D**: *T* depending on recording electrode and stimulus rate.

The values typically range from 0.01 to 1 bit per second per 2 Hz frequency bin. In Fig. 4.17 C the *PT* even exceed 5 bit per second for many bins. When summing them up we only included values significantly different from zero ($p \leq 0.05$). As mentioned above, for clarity reasons the superimposed spectral profiles display all original *PT* values.

At a first glance, LFP (A) and MUA (C) signals yield similar *PT* profiles, especially for Fig. 4.18 and 4.19. This might be astonishing since the signals should reflect different neuronal processes. Looking more closely reveals that LFP emphasizes the higher frequency bands slightly more compared to MUA. For high stimulus rates LFP and MUA demonstrate a drop with respect to the information transfer at lower frequencies, both in case of electrical and visual stimulation. This is plausible since frequency components below the fundamental frequency can only contribute stimulus information if they exploit the weak serial correlation introduced by the sequence of stimulus events.

Due to correlations between frequency components it is not possible to estimate the total *T* based on the *PT* in the frequency domain. Therefore, we applied PCA as coordinate transformation to yield uncorrelated coefficients. As long as nonlinear correlation is negligible and the coefficients are approximately normally distributed the *PT* may be summed up in the space of principal components (Eger and Eck-

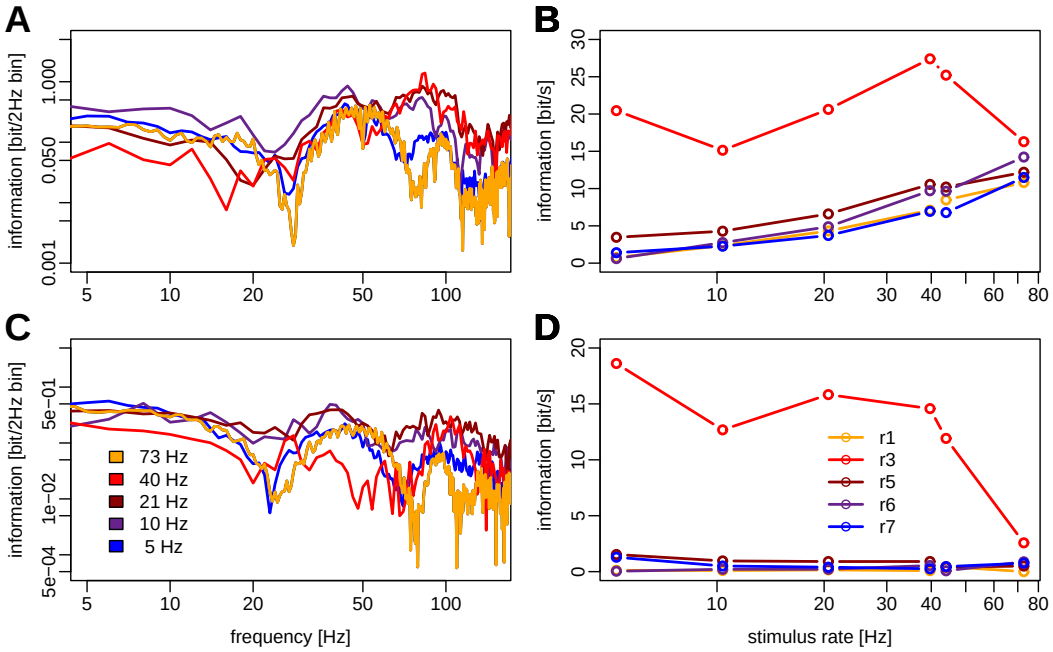

Figure 4.18: Partial and total transinformation dependent on the mean *electrical* stimulus rate, experiment [040]. Configuration identical with Fig. 4.17 except that electrode s3 is used for stimulation. **A**, **C**: *PT* of recording electrode r3. **B**, **D**: *T* depending on recording electrode and stimulus rate.

horn, 2002b, cf. Chapter 2 on p. 16). B and D in Figures 4.17, 4.18, 4.19 show *T* dependent on the mean electrical respective visual stimulus rate. Again, the curves are similar for LFP and MUA signals. In case of electrical stimuli, LFP (B) transmits a maximum of information at a stimulus rate of 20 Hz (125 bit/s, [060]) or 40 Hz (27 bit/s, [040]). The absolute values differ considerably between experiments but the shape of the curves is similar. For MUA we do not find a clear maximum at these stimulus rates. Instead, *T* starts of at high values (140 bit/s for [060], 18 bit/s for [040]) but decreases with growing rate – at least within the range of tested stimulation rates – and drops to very small values at a rate where LFP is still on a medium level. It is to be expected though, that *T* would also drop at stimulus rates below 10 Hz. An important observation is that for both signal types the information per stimulus is maximal for the lowest stimulus rate applied.

Fig. 4.19 shows the corresponding results for visual stimulation. The stimulus consisted of a spot of light emitted by a computer monitor set to 100% contrast and 50% intensity. The spot had 1.5 cm diameter corresponding to 0.8° and was located at the same position as stimulus electrode s3 which was the active stimulation electrode for the measurements given in Fig. 4.18 (see also Fig. 4.2 [040]). The position was not exactly congruent with the RFs of the recording electrodes. A 1° shift to the left would have fitted better to the RF position of r3. However, we judged it

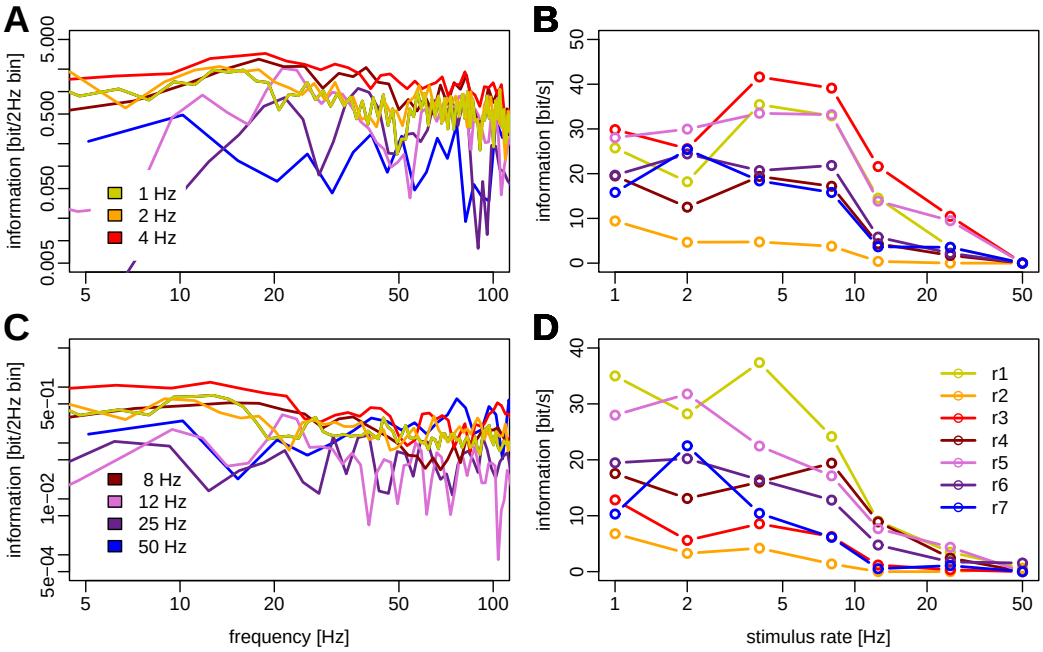

Figure 4.19: Partial and total transinformation (LFP: top; MUA: bottom) dependent on the mean *visual* stimulus rate, experiment [040]. Stimulation with 0.8° spot of light presented on a computer monitor at 100 Hz frame rate with stimulus intervals uniform distributed ±20% around mean stimulus interval. The visual stimulus appears at the same position as stimulus electrode s3 (see Fig. 4.18). **A**, **C**: *PT* of recording electrode r3. **B**, **D**: *T* depending on recording electrode and stimulus rate.

more important to have congruent positions for electrical and visual stimuli. Thus, we can directly compare the results for both modes of stimulation in the same experiment. Obviously, the achieved information rates are about 1.5 times lower for electrical compared to visual stimulation at the same position (approximately 27 vs. 40 bit/s for LFP and 18 vs. 37 bit/s for MUA). Still, this cannot be a general statement because the absolute values differ considerably between experiments. E.g., experiment [060] (Fig. 4.17) yields much higher values of more than 100 bit/s using electrical stimuli. Fig. 4.19 B shows a clear maximum of *T* at a rate between 4 and 10 Hz. Similar to the above result for electrical stimulation, here *T* based on MUA starts at a relatively high offset and drops a little earlier at lower stimulation rates.

Only in case of electrical stimulation, a further difference between LFP and MUA concerns the profiles of the weaker recording electrodes. In contrast to the effective electrodes (r3 in [040] and r4 in [060]) the LFP activity registered with the other electrodes rises nearly monotonously with growing stimulation rate whereas MUA is unaffected remaining at low activity. This might be characteristic for electrical stimulation and needs to be discussed (p. 98).

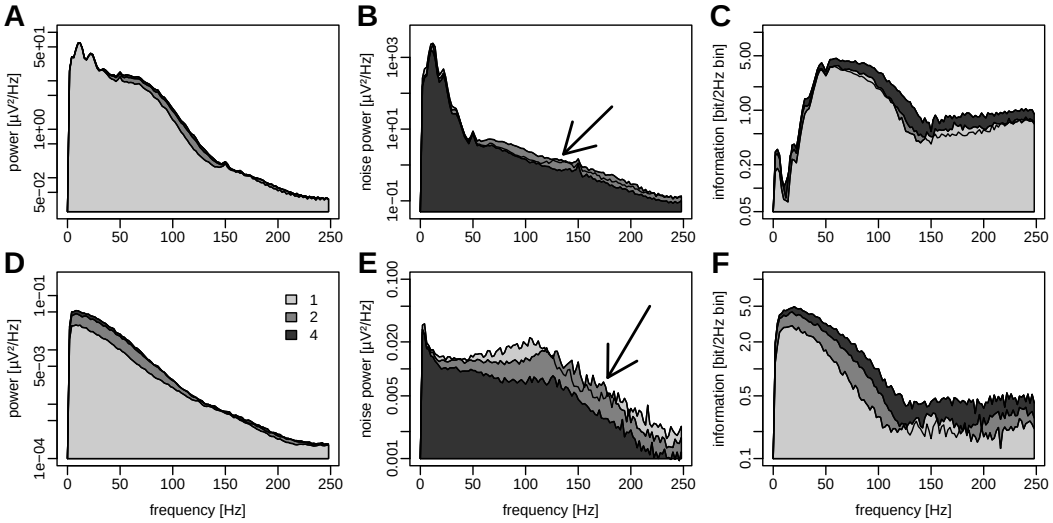

Figure 4.20: Spectral analysis of evoked responses (LFP: top; MUA: bottom) recorded with electrode r4 (see [060], Fig. 4.2) depending on the number n of stimulus intensity steps (see inset). The stimulation current is applied by electrode s6. The intensity steps are 50 µA for $n = 1$, 30 and 70 µA for $n = 2$ and 20, 40, 60, and 80 µA for $n = 4$ (Gamma distributed intervals with 50 ms mean). **A**, **D**: Response power. **B**, **E**: Deviation between original response and deterministic estimate defined as noise. **C**, **F**: Partial transinformation values. The small peaks in A and B at 50 and 150 Hz are caused by AC hum. The total transinformation rates after coordinate transformation using PCA are $T_{\mathrm{lfp}} = 127.0$, 121.3, 177.3 bit/s and $T_{\mathrm{mua}} = 82.5$, 121.5, 158.4 bit/s for $n = 1$, 2, 4 steps, respectively.

Intensity aspects. For *DS* (see p. 78) we found that with growing response strength the response variability increases for MUA but not for LFP. Now, in case of *CS* we quantify the response strength and variability according to the previous paragraph, by calculating the *PT* in the spectral domain but summing them up in the orthogonal space of principal components to yield *T*. The *PT* relate response strength and variability implicitly by using the signal-to-noise ratio between the deterministic and the non-deterministic fraction of the response. In four consecutive recordings of responses to current stimuli with 50 ms mean inter-stimulus intervals (Gamma distributed), we varied the number of intensity steps while keeping the average intensity constant. The current amplitudes were set to 50 µA for $n = 1$, 30 and 70 µA for $n = 2$ and 20, 40, 60, and 80 µA for $n = 4$. The stimulation threshold was just less than 20 µA (compare recordings within the same experiment, Fig. 4.14, B, E).

For LFP (top) and MUA (bottom) Fig. 4.20 depicts the response power, noise (non-deterministic fraction of response) and the *PT* for the first three recordings. The LFP spectrum has a maximum at about 15 Hz (A) but at the same frequency the noise is also maximal (B) suggesting that these response components are hardly stimulus correlated. The corresponding cut in C is a consequence of this weak signal-to-noise ratio and indicates that the frequency components are only poorly

used for information transmission. On the contrary, the frequency range between 50 and 100 Hz is emphasized in A and the noise is weak (B), resulting in large partial transinformation values (C). For higher frequencies the *PT* decrease again and reach a flat minimum at 140 Hz. The spectral characteristics for MUA are different. The noise is comparably weak for low frequencies leading to a maximum of the *PT* at about 20 Hz. Compared to LFP, the *PT* profile drops earlier. Already at 120 Hz the minimum level is reached.

Concerning the dependency on the number of intensity steps, especially for MUA the response power grows with increasing number of steps. At the same time the noise decreases resulting in enhanced *PT*. Note exceptions of this tendency marked by obscured areas (isolated border lines, B, E, arrows). For LFP, the transition from 2 to 4 intensity steps causes the *PT* to clearly increase whereas the transition from 1 to 2 steps is ineffective. The deviation between the different *PT* curves appears to be relatively constant (in logarithmic y coordinates) over the frequency range, indicating a constant gain factor. Thus, the additionally transmitted information by using 2 or 4 intensity steps vs. 1 intensity step does not concentrate on narrow frequency bands.

Fig. 4.21 depicts the cortical spread of different fractions of T depending on the number of intensity steps (different groups of bars). The 5 recording electrodes (r1, r4, r5, r6, r7, bars from left to right, within each group) were linearly arranged and 0.5 mm apart except for the 1.5 mm distance between r1 and r4 (cf. Fig. 4.2). One recording electrode (r4) largely exceeds the others with respect to information transmission. Analyses of LFP response signal type are shown in A, MUA in C. The total length of the bars measures the total amount of T. The dark grey sections of the bars denote the small fractions solely based on intensity aspects of the stimulus (T_{int}). The light grey sections represent the information transmitted by the temporal stimulus aspects. The unexplained rest, i.e., the total T minus pure temporal and intensity fractions corresponds to the medium grey sections.

For MUA the total amount of T increases more than 60% when 2 intensity steps (30, 70 μA) are chosen instead of 1 step (50 μA) and an additional amount of the same percentage in case of 4 steps (20, 40, 60, 80 μA). The values are averages over the five recording electrodes. The total values of T amount to 127.0, 121.3, 177.3 bit/s for LFP and 82.5, 121.5, 158.4 bit/s for MUA (recording electrode r4), corresponding to $n = 1, 2, 4$ intensity steps. Strangely, LFP keeps a relatively constant value of T for the transition from 1 to 2 steps, but a strong increase of 50% for the next transition to 4 intensity steps. In comparison to LFP, the spatial profile of MUA shows larger differences of the total values of T between different recording electrodes. On average (four recording sessions) LFP carries about 20% more information than MUA when considering the most effective electrode r4. The percentage is far higher for the other electrodes. However, the superiority of LFP compared to MUA with respect to total T is not transferable to the exclusive encoding of stimulus intensity.

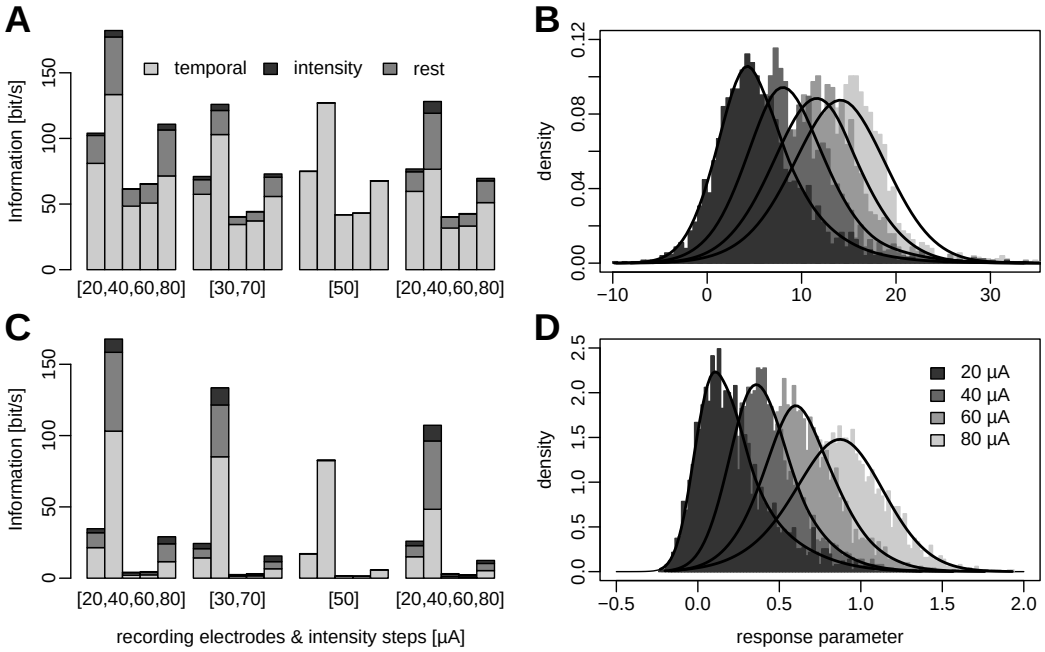

Figure 4.21: Fraction of T solely based on stimulus intensity aspects vs. total T computed on the basis of LFP (top) and MUA (bottom) activity already shown in Fig. 4.20, experiment [060]. **A**, **C**: Fractions of T in subsequent recordings with 5 cortical electrodes r1, r4, r5, r6, r7 (see Fig. 4.2) with varied number of stimulation intensity steps. Each group of bars comprises one recorded data set. The current steps are printed below. The bars within each group correspond to the different recording electrodes. **B**, **D**: Distributions of the first principal component coefficient corresponding to the four stimulus conditions (current amplitudes 20, 40, 60, 80 µA, first group of bars) and the recording electrode r4.

For recording electrode r4 and the first group of bars (four intensity steps, stimulus conditions 20, 40, 60, 80 µA), Fig. 4.21 B (LFP) and D (MUA) display the distributions of the response coefficients conditional on the stimulus intensity. The coefficients are the result of a modified PCA performed on potentially overlapping response signal epochs (see Eger and Eckhorn, 2002a, cf. Chapter 3). They compactly characterize the evoked response following each stimulus event. Even though the conditional densities more or less overlap, the coefficients still provide some information about the condition of the preceding stimulus. We quantified this information according to Appendix 4.5.1 and yielded T_{int}, the fraction of T solely based on the intensity stimulus aspect (Fig. 4.21 A, C, dark grey sections).

For all recording electrodes and recording sessions the relative and absolute amount of intensity information T_{int} is higher for MUA than for LFP. Evidently, MUA is better capable of encoding this stimulus parameter than LFP. The average (over electrodes and recording sessions) fraction of total T was less than 3% for LFP and about 15% for MUA. For these values merely the electrodes r1, r4, and

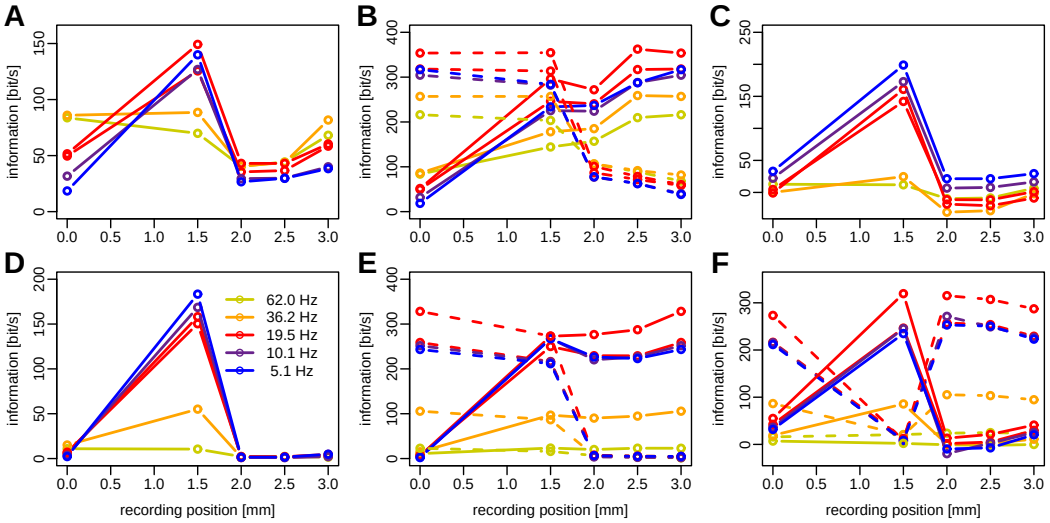

Figure 4.22: Transinformation (LFP: top; MUA: bottom) dependent on the cortical recording position in mm and *electrical* stimulus rate, experiment [060]. The recording positions correspond to electrodes r1, r4, r5, r6, r7. **A**, **D**: T between stimulus and one recording electrode at a time. **B**, **E**: Spatially cumulative transinformation from left (solid) and right (dashed). T is calculated between the stimulus and n-tupels of recording electrodes. Note that the contributions to the cumulative value may be negative due to added noise so that the curve is not rising homogeneously. **C**, **F**: Exclusive contribution of the recording electrodes. T is computed as a difference (solid) between the total joint value (according to B, E) and the value based on all but the actual recording position (dashed, shown for MUA only). Stimulation identical with Fig. 4.17.

r7 were regarded. Note that r7 yields considerable percentages for MUA. Further, we found that for both LFP and MUA the amount of information about the stimulus intensity alone cannot be increased by using more than two different stimulus intensities. We want to note that these analyses comprise extended examples based on reliable recordings. Yet, general statements should be substantiated by analyses of additional data.

The fourth recording was a repetition of the first to test stationarity (Fig. 4.21 A, C, last and first group of bars). The diagrams reveal a non-negligible difference with respect to T. Indeed, during the recording of the fourth data set a noticeable state change of the anesthetized animal had occurred: the spontaneous activity increased and the efficiency of the stimulation was reduced.[2] Similar state changes are reported by Tolhurst et al. (1981).

Spatial aspects in response space. Due to retinotopical mapping between the retina and the visual cortex (Orban, 1984) retinal activity patterns are related to corresponding cortical activity profiles. At a first glance, divergence, lateral inter-

[2]The spontaneous activity was controlled by acoustic monitoring of the ongoing activity. The efficiency was assessed on-line during the recordings by short term averages.

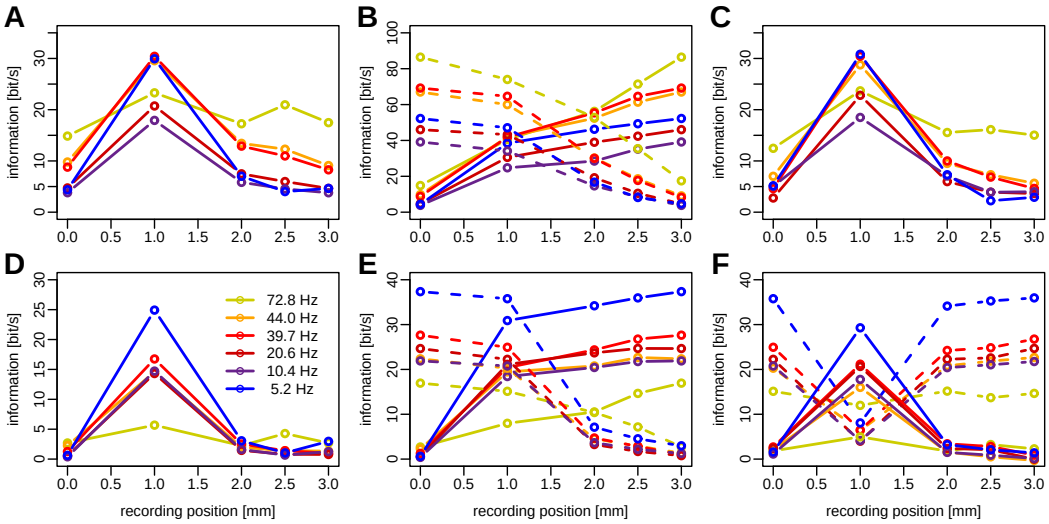

Figure 4.23: Transinformation dependent on the cortical recording position in mm and *electrical* stimulus rate, experiment [040]. The recording positions correspond to electrodes r1, r3, r5, r6, r7. Otherwise the configuration is identical with Fig. 4.22.

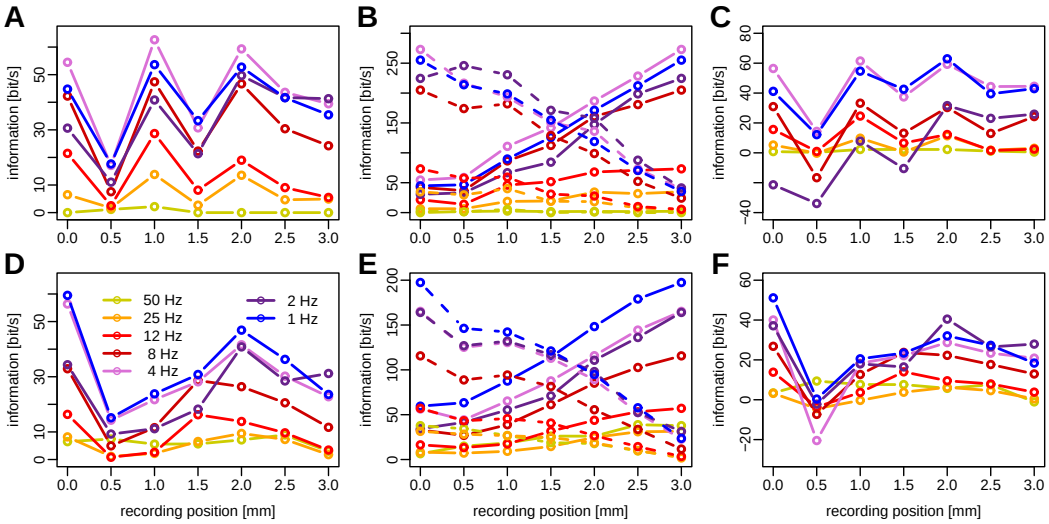

Figure 4.24: Transinformation (LFP: top; MUA: bottom) dependent on the cortical recording position in mm and *visual* stimulus rate, experiment [040]. The recording positions correspond to electrodes r1, ..., r7. Visual stimulation identical with Fig. 4.19. **A**, **D**: T between stimulus and one recording electrode at a time. **B**, **E**: Spatially cumulative transinformation from left (solid) and right (dashed). **C**, **F**: Exclusive contribution of the recording electrodes. T is computed as a difference between the total joint value (according to B, E) and the value based on all but the actual recording position.

actions, and cortical feedback mechanisms "blur" the projection in the sense that, e.g., sharply localized retinal activations turn up as relatively broad cortical activity profiles. In the following, we will investigate the transinformation between retinal stimulation electrodes and different cortical recording positions. In most cases, we had 5 cortical recording electrodes available, that were arranged as a linear array and coarsely sampled the cortical activity profile.

The diagrams A, D of Figures 4.22, 4.23, 4.24 show the transinformation dependent on different cortical positions and the stimulation rate for LFP (top) respectively MUA (bottom). While Figures 4.22 and 4.23 display the results of *electrical* stimulation from two independent experiments [060, 040], Fig. 4.24 is based on *visually* evoked responses [040]. The location of the visual stimulus was identical with the stimulation electrode position in Fig. 4.23. The shape of the cortical transinformation profiles based on electrical stimulation is very similar for the two experiments [060, 040]. In both cases – for MUA and LFP alike – we find a clear maximum which is getting the less pronounced the higher the mean stimulation rate. The cortical position of the maximum corresponds to a recording electrode whose RF is very close to the stimulus position projected into the visual space (see Fig. 4.2). Interestingly, the values of T for the other recording electrodes increase with enhanced stimulus rate only for LFP, an observation we have made already above and which is discussed on page 98. The whole profile adjusts to a mean level with reduced maximum and increased surround. On the contrary, the MUA transinformation profile, which mainly consists of the contribution of the most effective electrode, is reduced per se at growing stimulus rates. The values of T for the other electrodes do not increase. The radius of cortical activation estimated by this information theoretic measure is less than 1 mm for MUA and also for LFP, if one neglects the effect described above.

In case of *visual* information, the T profiles look differently. LFP demonstrates a zigzag course with three maxima, while the MUA profile is smoother and shows its main maximum at the boundary and another at a position where LFP has also got a peak. Unlike electrical stimulation, here the information declines when the stimulus rate is enhanced, though differently for LFP and MUA. While LFP strictly keeps the shape of its zigzag profile, the MUA contours become smoother and the lower maximum is shifted to the left. Based on these graphs it is impossible to estimate the extension of the cortical activation.

Until now, we have determined the stimulus information received at different cortical positions without taking into account any coupling of cortical positions. Yet, it is to be expected that the information revealed by one electrode overlaps with that by neighboring electrodes. In order to assess the amount of redundancy we calculated cumulative profiles by considering joint signal epochs (see Appendix 4.5.4). Diagrams B, E of Figures 4.22, 4.23, 4.24 depict the cumulative profiles for LFP (top) and MUA (bottom), again dependent on the mean stimulus rate. The solid

lines indicate that the joint epochs are built starting from the left, whereas the dashed lines denote the other direction.

In case of *visual* stimulation (Fig. 4.24 B, E) the cumulative graphs grow quite smoothly. Unlike, *electrical* stimulation leads to cumulative profiles that abruptly rise at the position of the most effective recording electrode (Fig. 4.22 B, C and Fig. 4.23, E). These jumps in the cumulative curves are often higher than the respective contributions when the electrodes are considered independently. E.g., in Fig. 4.22 B, the step in one of the dashed 19.5 Hz curves amounts to about 250 bit/s whereas the position 1.5 mm in A merely conveys 140 bit/s. Furthermore, the cumulative profiles do not grow homogeneously. In Fig. 4.22 B the slopes of the solid 19.5 Hz curves are partly negative. Obviously, the cumulative contribution of weak electrode positions in case of electrical stimulation is often zero or below, i.e., their contribution is redundant. On the other hand, "better" positions tend to be overpronounced, i.e., they are synergisticly amplified by the weak activity of nearby positions. Another example is Fig. 4.22 E. Here, the cumulative profile of the solid 19.5 Hz curve demonstrates a clear upward bend at position 3 mm, although this electrode hardly conveys stimulus information at all when handled independently (D). Apparently, this electrode merely contributes synergistic information, i.e., information which is accessible only in combination with signals of other electrodes.

The diagrams C, F of Figures 4.22, 4.23, 4.24 show T conveyed by all except for the actual position (dashed lines, for clarity reasons only partly shown) and the value subtracted from the total cumulative transinformation (solid). The solid curves look similar to the respective ones in diagrams A, D (especially in Fig. 4.23), but looking more closely, we find some differences. As observed for the cumulative curves, due to synergistic effects the peaks often are higher in C, F compared to those in A, D and some values are negative. The leveling out in case of LFP for higher rates (Fig. 4.22 A) leads to an overall decline in C very similar to the result of MUA (F). In case of visual stimulation (Fig. 4.24 C, F) the original lower peak (D) for MUA loses its predominance in F. C shows a largely negative profile for a rate of 2 Hz. E.g., the position 0 yields 30 bit/s when regarded independently (A) and surprisingly, simply changes its sign in C. Thus, if electrode position 0 is included into the joint measure then the overall information declines by more than 20 bit/s! A similar effect at the same position can be seen for MUA at 4 Hz stimulus rate (D vs. F). In these analyses we did not find synergistic encoding of visual stimulus information in contrast to electrical stimulation.

Spatial aspects in stimulus space. Different from the analyses presented above, Fig. 4.25 depicts the amount of T in case of spatio-temporal stimulus patterns. Electrical stimulation was performed simultaneously with 7 retinal electrodes arranged hexagonally as shown in Fig. 4.2 [040]. The electrode tips positioned on the retinal surface each had an about 0.5 mm distance to the closest neighbors. 7-channel independent, jittered m-sequences (see Fig. 4.16 A) were used as spatio-temporal stimulation signal. The stimulation electrodes were activated one after another.

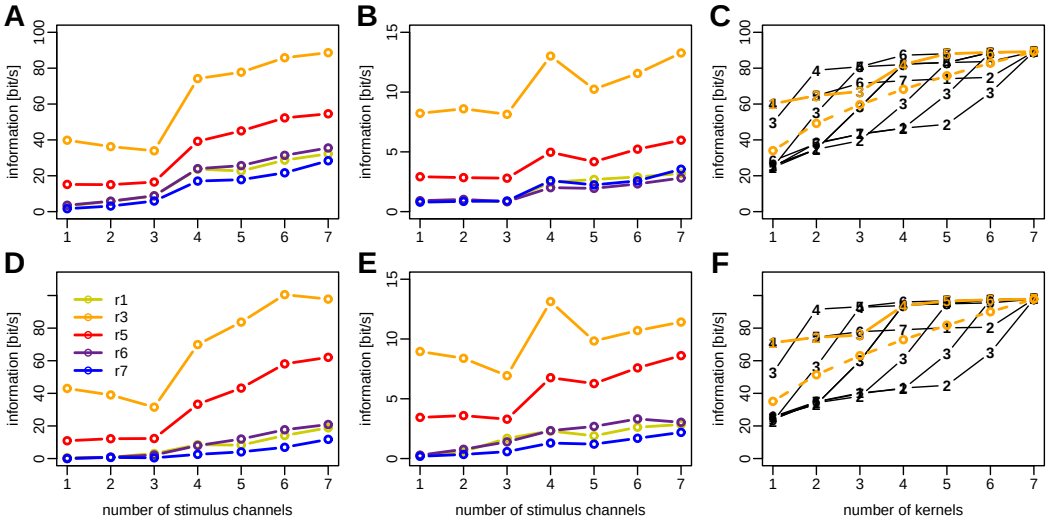

Figure 4.25: Transinformation (LFP: top; MUA: bottom) dependent on the number of independent m-sequence stimulus channels. The mean stimulus rate is 15 Hz (see Fig. 4.16 A). In A, D; B, E subsequently the number of stimulus electrodes is increased by activating a new electrode at each step. According to Fig. 4.2 [040] the *n*-tupels of stimulation electrodes are [4], [4,5], [2,4,5], [2,3,4,5], [2,3,4,5,6], [1,2,3,4,5,6], [1,2,3,4,5,6,7]. While A, D depict the total transinformation, in B, E only the non-temporal, transient stimulus pattern specific transinformation is assessed (see Methods in Eger and Eckhorn, 2002a, cf. Chapter 3) accounting for 10–20% of the total *T*. C, F shows a different analysis on the same data as was used for the last step in A, D; B, E. Here, *T* is calculated by exclusively assigning kernels to specific stimulus conditions. The number of kernels is systematically increased from 1 to 7 covering several possible traces of *n*-tupels. The solid yellow line indicates the trace that corresponds to the sequence of *n*-tupels given above for A, D; B, E. The dashed yellow line is the average of the black traces.

Fig. 4.25 A, D depict the dependency of total *T* between the multi-dimensional stimulus signal and the respective recording electrode on the number of active stimulus electrodes. Except for a drop at 3 channels[3], the graphs are mostly rising and suggest the beginning of a saturation at a number of 5 stimulation electrodes. Diagrams B, E accordingly present the dependency of the fraction of *T*, which is solely due to non-temporal stimulus aspects (see Eger and Eckhorn, 2002a, cf. Chapter 3). The graphs do not indicate a saturation effect. The fraction amounts to about 10–20% of total *T* for the given spatio-temporal stimulus patterns.

The last diagrams C, F demonstrate a re-analysis of the data evoked with 7 independent channels (A, D, last value, recording electrode r2). Here we kept the number of active electrodes constant (7) but varied the number of kernels. Thus, we included only a restricted amount of stimulus information into the response model

[3]This is most probably due to a minor state change, because the errors yielded by the applied jackknife procedure are below 1%.

(see Eger and Eckhorn, 2002b, cf. Chapter 2) and likewise, into the calculation of *T*. This approach is equivalent to using different codes. As an example, the first value of the curve starting with number 4 results from the assignment of one kernel to the stimulus events of stimulation electrode 4. In a next step, another kernel is assigned to stimulus events of electrode 5 etc. The more kernels are used, the higher is the potential precision of the response model, which is revealed by the positive slope. Obviously, stimulus electrodes 3 and 4 are highly efficient. The curves starting with these respective kernels already begin with a relatively high offset of *T* whereas other curves show a jump when kernels connected with electrodes 3 or 4 are added. E.g., the curve starting with 3 shows a jump at the next kernel assigned to stimulus electrode 4. The kernels added later hardly lead to an increase of *T*, which indicates that the cortical electrode r2 receives information mainly from stimulus electrodes 3 and 4. It would have been desirable to compare these results with those that also include the other cortical electrodes by using the joint transinformation (see Appendix 4.5.4). Thereby, one could test whether this selectivity was specific for the recording electrode r2. However, the calculations are computationally very expensive and for this reason, were not possible.

4.4 Discussion

4.4.1 Temporal resolution

Spike timing. We determined the achievable temporal resolution in two ways, using discrete (*DS*) and continuous stimulation (*CS*). In case of *DS* we performed spike analysis and determined the temporal precision of evoked cortical action potentials in response to electrical retina stimuli. We yielded an astonishingly high temporal precision with even less than 300 µs standard error for the first cluster of spikes at 6 ms following the stimulus (Fig. 4.6). This variability is about two or three times the error measured by Ferster and Lindström (1983) after electrical stimulation in the lateral geniculate nucleus (LGN) of the cat. The difference is certainly due to the extra variability introduced by the retinal spike initiation in our case. Compared to the values reported in case of visual stimulation, the standard error is at least the factor 10 lower (2–10 ms in monkey MT (Bair and Koch, 1996), 1.9 ms also in monkey MT (Buracas et al., 1998), >10 ms in monkey V1 (Wiener and Richmond, 1999), 5 ms in cat LGN (Reich et al., 1997), cf. overview in (Bair, 1999)) and due to the bypassing of slow retinal structures by directly exciting the ganglion cells (Potts and Buffum, 1968).

Concerning the example presented in Fig. 4.7, we suggest that this early activity at 6 ms comprises action potentials recorded from one layer 4 neuron n1 that receives direct input from the LGN via the X pathway. As reviewed by Orban (1984), the X ganglion cells (constitutes 45–55% of all retinal ganglion cells) are relatively slow due to low axonal velocity, have small somata and are used for the

perception of form and color. For comparison, the less frequent Y cells (5% of ganglion cells) are fast, contrast sensitive and responsible for movement perception. To substantiate the above hypothesis we first have to refine the measured delay of 6 ms by regarding the time needed for retinal spike initiation (0.2–0.3 ms, cf. Ferster and Lindström, 1983) and the conduction time between the soma of the retinal neuron stimulated and the electrode position (1.5 ms delay, caused by 10° dislocation and 1.33 m/s measured conduction velocity, see next paragraph). Thus, in our case the total conduction latency between a retinal and cortical spike amounts to $6 - 0.25 + 1.5 = 7.25$ ms.

Now we compare this value to data from the literature to identify the recorded cell type. For retinal X cells a latency of 4.56 ms has been measured at the axon terminals in the LGN and 4.8 ms including spike initiation (Stanford, 1987). Usrey et al. (1999) report a value of 4.45 ms for X cells and 2.66 ms for Y cells.[4] Usrey et al. (2000) measure an additive delay of 2–4 ms between LGN and simple cell spikes in A17. The overview given by Orban (1984) reports a pure conduction latency of 0.4–1.4 ms for Y cells and 0.9–4 ms for X cells. He suggests a best separation for the identification of unknown cells at 0.9 or 1 ms. A precise analysis of the geniculo-cortical connectivity is presented by Ferster and Lindström (1983) who quantified an extrapolated latency that exclusively captures the processing in the visual cortex. Correspondingly, a certain duration has to be added to the pure latency, dependent on whether the recorded cortical activity is mediated by one, two, or three cortical synapses: a monosynaptic connection "costs" 0.6–1.05 ms (mean 0.7 ms), a di-/ and trisynaptic connection needs 1.05–2.4 ms, respectively 2.4–3.8 ms. Additional 0.4 ms have to be considered for the delay between the registered cortical excitatory postsynaptic potential (EPSP) and the elicited spike.

Taking these pieces of information together and using realistic values of 1.3 and 0.8 ms for the unknown delay between LGN and area 17 via the X and Y path, respectively, a total relative delay between the activity of a retinal X cell and a monosynaptic connected cell in area 17 amounts to about $4.8+1.3+0.7+0.4 = 7.2$ ms compared to $4.8 - (4.45 - 2.66) + 0.8 + 0.7 + 0.4 \approx 4.9$ ms for Y cells. For disynaptic connections, a millisecond has to be added. To conclude, our value of 7.25 ms may either arise from recording a monosynaptic coupled simple cell of the X pathway or a trisynaptic connected cell of the Y pathway. The latter does not seem to be very probable because the variability measured is very low. For this reason we rather suggest that the electrically evoked cortical activity is caused by monosynaptically excited cells with X input.

By this argument we do not want to exclude that we effectively stimulated retinal Y cells. Indeed, in several cases we found extracellularly registered spikes in close accordance with the timings of monosynaptic Y activations. The broad band

[4] In contrast to others, they performed extracellular recordings and calculated the crosscorrelation between retinal and geniculate spike trains to estimate the latency. The small deviation might be due to the different method used.

recordings (BBR) in Fig. 4.3 and Fig. 4.11 reveal a small peak at 4.5 ms (marked by dashed vertical line) which precedes the dominant needle like activation 2 ms later. Yet, in raster plots the corresponding events were never very precise and could also have been intermingled with effects of the stimulus artifact. Therefore, we did not focus our attention on these very early events.

We have often registered a second cluster of spikes at 8.5 ms and at least double variability (Fig. 4.6, Fig. 4.7). We suspect that this activation belongs to a second neuron n2, which receives mono-/ or disynaptically mediated input from n1. We found that n2 fires if n1 was active about 2.5 ms before. However, since n1 emitted one spike following nearly each stimulus event, the hypothesis of direct coupling could not be proved. The histogram of the time differences of consecutive spikes with respect to the clusters at 6 and 8.5 ms, revealed about the same variance as that of the second cluster (Figure not shown).

Given these results, we can imagine – on the condition of low stimulation rates – that spatio-temporal timing differences in the millisecond range might be perceivable on the basis of electrical retina stimulation, if higher cortical areas are able to assess these temporally precise activation patterns in the primary visual cortex.

Axonal stimulation. The stimulation of retinal ganglion cell axons poses a yet unsolved problem on the development of the retina implant because of the spatial and temporal disorder introduced between retinal stimulation positions and evoked cortical activity (cf. Nowak and Bullier, 1998a,b). As shown in Fig. 4.6, 4.7, the same cortical neurons were reliably activated, both by two stimulation electrodes E1 and E2 positioned about 10° distant from the RF locations of the recorded cortical neurons (Fig. 4.2 [318], recording electrode r1). This strongly suggests that we stimulated the axons of retinal ganglion cells whose somata were located near the cortical RF positions due to retinotopy of V1 (Tusa et al., 1979; Tootell et al., 1982). The stimulation electrodes were about 400 µm apart corresponding to a visual angle of 2° which probably accounts for the relative delay of ≈300 µs of their respective cortical activation. The values for distance and delay lead to a conduction velocity of 1.33 m/s which is in close accordance with reports by Stanford (1987). Axonal stimulation violates a retinotopic bijective relation between stimulation and cortical response patterns and thus, inevitably leads to information loss. In addition, the timing of the cortical activity onset is disrupted, because the distinct latencies of the myelinated extraocular axon sections naturally compensate the different conduction latencies of the unmyelinated intraretinal part (Stanford, 1987). By electrical stimulation of retinal ganglion cell axons the natural conduction delay is shortcut, which most probably leads to temporal blurring of evoked cortical activations.

We exemplarily tested how well the stimulation electrodes can be distinguished based on the time of the first recorded spike (window 3–7 ms, Fig. 4.7). An information theoretical analysis revealed that the spike timing yielded 0.5 bit/spike about stimulus conditions E1 or E2. The calculation was performed based on density estimation (see Appendix in Eger and Eckhorn, 2002b, cf. Chapter 2 on p. 43).

Population activity. The temporal course of evoked population activity potentials is in agreement with the review of Creutzfeldt and Kuhnt (1973). We found corresponding response deflections but a slightly larger delay due to more distal stimulation (Fig. 4.3, 4.8). Yet, depending on the electrode depth in the visual cortex and the position relative to the stimulus location projected into the visual space, the size and sign of later response deflections varies. Regarding *PT* of the decorrelated response signals (bottom rows of Fig. 4.8, 4.9, and Fig. 4.10, A, C) most information about the preceding stimulus is transmitted within the first 40 ms following the event. The total cumulative values of 0.1–2.0 bit per stimulus are in agreement with with the result that only 2 or 3 stimulus intensities corresponding to 1–1.6 bit can be discriminated at most (cf. p. 102). After the 40 ms interval (at the latest, after 60 ms, note exceptions of very strong responses in Fig. 4.8 D and Fig. 4.10 A) neither LFP nor MUA reveals considerable information about the stimulus rendering later response deflections redundant. Typically, the uncorrelated fraction of MUA activity has a shorter duration. The later response deflections might be the consequence of cortical feedback loops for the post-processing of the activity evoked initially. In contrast to electrical stimulation, the *PT* profiles for visual stimulation at the same stimulus position (compare Fig. 4.10 B, D and B, D) yield a much larger delay of the first volley of activity and only a single peak for the MUA signal. LFP demonstrates a slow after-response and three uncorrelated response components at equidistant onsets. They may indicate sub-threshold rhythms of mutually uncorrelated activity. Serial decorrelation of the data had been achieved using Schmidt's orthogonalization procedure (cf. Appendix 4.5.3 and Eger and Eckhorn (2002c)) before the transinformation was calculated. Due to the restriction to linear decorrelation and the use of signal-to-noise ratios, normality was assumed and potential higher order correlation is not captured by the method (Mendel, 1991).

Transinformation dependent on stimulation rate. Apart from *DS* referred to above, we also used quasi continuous stimulation signals consisting of rapid sequences of transient stimuli. In the following we discuss the analysis of data in response to Gamma distributed inter-stimulus intervals in case of electrical stimulation (Figures 4.17, 4.18) and uniform distributed intervals in case of visual stimulation (Fig. 4.19). We found that the LFP signal has a maximum of T in the range of 20–40 Hz for electrical stimulation, which corresponds to the 40 ms time constant resulting after serial decorrelation. The maximum of T in case of visual stimulation turns up between 4 and 10 Hz. This is in accordance with Grüsser's results and may correspond with the psychophysical Brücke–Bartley effect (see review in van de Grind et al., 1973), which says that the subjective brightness of an intermittent visual stimulus yields a maximum at about 3–15 Hz stimulus rate dependent on the stimulus luminance. Still, in contrast to LFP we found that T based on MUA begins to decline at slightly lower rates for both electrical and visual stimulation.

Before we will try to explain this effect we have to clarify what kind of information these signals registered at the same electrode encode. LFP simply results

from lowpass filtering of the raw signal and reflects the slow potentials surrounding the recording electrode tip. It has a larger catchment area compared to MUA (max. 1 mm radius for LFP vs. 0.5 mm for MUA, cf. Schanze, 1995) and thus, spatially averages to a larger degree. These values include a spread caused by lateral coupling within the cortical network. Potential sources of the slow potentials are mainly excitatory and inhibitory postsynaptic potentials (IPSP and EPSP, respectively), i.e., dendritic signals reflecting the graded input to cortical neuron cell bodies. Mitzdorf (1985) estimated and compared the contributions of different neuronal activity types to current source densities (CSD) that are based on local field potential measurements. She showed that IPSP and EPSP have similar signal properties, except that the currents caused by IPSP are to a larger degree capacitive. Because of the lowpass properties of the tissue, the contribution of IPSP to CSD and LFP is relatively weak.

MUA results from highpass filtering of the raw signal and subsequent rectification and lowpass filtering (see p. 66). It can be seen as the envelope of frequency components between 0.5 and 10 kHz, a range mainly used by action potentials that typically have a time constant of about 1 ms (Nicholls et al., 1992). Thus, it represents a spike density spatially integrated over a sphere with $\approx$0.5 mm radius. Contrary to LFP it does not capture slow potentials. Since most recordings were done in the larger layer 4 of area 17 which receives direct afferent input from LGN fibers (Orban, 1984), the early and precise components of LFP reflect their postsynaptic potentials and thus – simply spoken – comprise the input to the visual cortex. In contrast, MUA captures the activity of the spikes in response to the excited postsynaptic potentials and inherits the additional variability of spike generation.

Now we are turning back to the explanation of the above described effect, i.e., T based on MUA declines at lower stimulus frequencies compared to LFP. A drop in T either originates from an increase of the response variability or from a decrease of the stimulus evoked response power, as long as the linear response model is still adequate (Eger and Eckhorn, 2002b, cf. Chapter 2 on p. 13). Here, we cannot totally exclude the latter influence because we do not have recordings in response to identical repetitions of the same stimuli available. Otherwise the performance of the linear response model could be assessed according to (Roddey et al., 2000). However, we find it plausible that the adequacy of the model does not change dramatically – and differently for LFP and MUA – dependent on the stimulus rate.

Sticking to the arguments outlined above we conclude that the input to the cortical neurons (LFP) – due to its lower variability – is better capable to follow higher stimulation rates than MUA. Apparently, the cortical neurons might not be able to follow the high stimulation rates with maintained precision. This corresponds to Grüsser's findings reviewed in (van de Grind et al., 1973, p. 527), that at visual flicker stimulation the upper frequency limit (critical flicker frequency, CFF) decreases from the retina, over the LGN to the visual cortex while traversing synaptic connections. Since the respective neurons themselves are directly electrically ex-

citable at very high stimulation rates, the reason is not seen in the properties of the synapses but rather in inhibitory feedback loops reducing the cells' excitability.

One may still assume that electrical compared to natural visual stimulation has the outstanding advantage that the slow retinal structures are bypassed and the cortex can be fed with information at much higher rates. Indeed, the cut-off frequency in the profile of T amounting to 20–40 Hz in case of electrical stimulation (Fig. 4.17, 4.18) is considerably higher than that achieved with visual stimulation (10 Hz, Fig. 4.19). Psychophysical experiments with humans proved that electrical flicker phosphenes caused by external mono-focal electrical stimulation of the eye are perceivable up to frequencies of 100 or 120 Hz (van de Grind et al., 1973, p. 465). Thus far, it is not yet clear whether the cortical neurons make use of these high frequency inputs to form useful percepts such as the binding or separation of objects represented by electrical spatio-temporal stimuli.

The following finding should be considered in this context: evoked activity in response to electrical stimulation is very sensitive to the relative position of stimulation and recording electrodes projected into the visual space. The two Figures 4.17 and 4.18 show one very efficient recording electrode (r3 in [040] and r4 in [060]) apart from several weaker electrodes. Interestingly, exclusively for LFP the activity of the weaker electrodes increases with growing stimulation rate. One reason might may be the smaller catchment area of MUA. However, the activity profile of Fig. 4.17 D, r5 shows a very small increase at higher rates for MUA only in one example. Another interpretation which may be more plausible is that the higher stimulation rates lead to saturation of the cortical activity. The increased LFP activity may rather correspond to inhibition (IPSP) than to excitation. The abrupt drop of T for MUA at the respective stimulation rates is in accordance with this view.

Efficient stimulation intervals. Dawson and Radtke (1977) demonstrated the effects of the retinal stimulation rate depending on the height of the peak or area under the intra-cortical response curve. They found a maximum of the amplitude at an inter-stimulus interval of 500 ms and a decline for shorter intervals, though using very high charge injections (2 mA, 0.8 ms phase width) compared to our studies. The effect of declining response strength with shorter intervals is partly in agreement with our results. Fig. 4.14 C, F reveals that an electrical stimulation is most effective if it follows a facilitating stimulus within 10–30 ms or after a longer delay of more than 150 ms. On the other hand, it is less efficient if it follows a preceding stimulus after 70 ms. This effect has shown to be relatively independent of the stimulation current of the second stimulus. It has to be investigated how the effect depends on the stimulation current of the preceding stimulus. The depressive effect of a preceding stimulus could be explained by cortical inhibition. The example in Fig. 4.11 demonstrates that inhibition depends on the strength of the previous activation. In B, top panel, after a strong activation peak, inhibition arises about 20 ms after the stimulus onset and vanishes another 20 ms later (note range indicated). Another example is Fig. 4.9 A. The inhibition begin is in accordance

with the downward bend of the curves in Fig. 4.14 C, F at about the same time. However, the depression endures much longer than the registered period of spiking inhibition which suggests that other suppressive mechanisms not revealed by the present analyses are at work.

4.4.2 Intensity resolution

Mean variance relation. In Results (see Fig. 4.12 and Table 4.1) we have shown that the variance of MUA activity grows with increasing response strength while the LFP variability is comparably independent of it. The regression exponent for MUA lies in the range between 0.49 and 1.1 and thus, is mostly lower compared to values for single unit activity reported in the literature (1.1 for simple cells, cat A17 in (Dean, 1981), 1.18–1.43 for complex cells, monkey V1 in (Gershon et al., 1998), 1.0–1.4 for simple and complex cells, cat A17 in (Tolhurst et al., 1981)). Concerning the different characteristics of LFP and MUA described above – corresponding to the very simplified view of cortical input and output, respectively – we can say that the afferent input from LGN (LFP) does hardly change its variability with increased response strength, whereas the output (MUA) does. This finding is in accordance, e.g., with (Kara et al., 2000), who recorded visually evoked spike activity simultaneously from the retina, the LGN, and V1 in anesthetized cats. They find the variability doubling from retina to LGN and again from LGN to V1. However, Kara et al. (2000) find the variability decreasing with increased firing rate, i.e., response strength, contrary to our results and, e.g., those of Gur et al. (1997); Gershon et al. (1998). Above all, one should keep in mind that we have not analyzed spike but population activity. Population activity signals might be behaviorally and with respect to perception more relevant than single unit spike trains (Mitzdorf et al., 1994). A systematic comparison of these different signals, especially in case of electrical retina stimulation, is still to be done.

Response strength vs. current. Dawson and Radtke (1977) have shown that the amplitude of responses evoked by electrical retinal stimulation in awake cats does not grow linearly with the stimulation current but saturates with higher currents. However, they used huge currents from 0.1 to 2 mA and large phase widths of 0.8 ms in comparison to 100 µA and 200 µs respectively, in our case. Nonetheless our studies revealed qualitatively similar results (see Fig. 4.15).

Response model. As mentioned before and according to Fig. 4.12, the response variance is almost independent of the response strength for the LFP signal type. Even for MUA the dependence is rather weak. This observation may justify the application of a simple model that assumes constant variance in order to coarsely estimate the maximum number of discriminable intensity steps. The question of how many different intensity steps can be discriminated by analyzing evoked cortical activity is equivalent to the intensity resolution and thus, important for the development of the retina implant.

With respect to the information capacity of a randomly discharging model neuron, Stein (1967) calculated the maximum transinformation dependent on the number of stimuli to be discriminated. He utilizes the spike count within a window of varied duration as response code. A similar but more complex approach is presented by Gershon et al. (1998) for the calculation of the channel capacity. He examines the discriminability of visual patterns based on recorded spike activity in monkey V1.

Here, the principal idea consists in the consideration of the response distributions within the limited response range as schematically indicated by a red bar in Fig. 4.26 A and C. Now, the transinformation between n equally probable stimulus conditions (or symbols) and the response parameter distributions only depends on their respective positions r_i within the response range R and the variance σ^2 if normality is assumed[5]. It can be numerically computed as

$$\begin{aligned} T(S;R) &= H(R) - H(R|S) \\ &= -\int_{-\infty}^{+\infty} p_r(r) \log_2 p_r(r) dr \\ &\quad -\frac{1}{2}\log_2[2\pi e\sigma^2] \end{aligned} \tag{4.3}$$

with

$$p_r(r) = \frac{1}{\sqrt{2\pi}\sigma n} \sum_{i=1}^{n} e^{-\frac{1}{2}\frac{(r-r_i)^2}{\sigma^2}} \tag{4.4}$$

(see Appendix 4.5.1). In case of three stimulus conditions Fig. 4.26 C depicts the value of T dependent on the position of the third distribution within the response range R on the condition that the other distributions are located at the boundaries. As long as the normalized standard deviation $s/R < 0.28$, then T is maximized if the third symbol is located in the center of the range R. Yet, if s/R increases beyond 0.28, the third distribution has to be positioned at either boundary to maximize T. In this case, two stimulus conditions are sufficient. It is not possible to increase T by adding further symbols. Equivalently, one can say that the system is capable to discriminate only between two stimulus conditions. The critical value of s/R is reflected in the bend of the curve D indicated by arrows. Fig. 4.26 B gives the maximum T for 2, 3, 4, 6, and 9 stimulus conditions dependent on s/R. For larger values of s/R, small symbol ensembles perform as well as larger ones. In this simulation they may perform even better because the symbols were assumed to be equally probable. In the example Fig. 4.26 A, C two symbols outperform three symbols for larger values of s/R. In fact, although three symbols were assumed, they were actually reduced to two symbols with different frequencies (1/3 vs. 2/3) due to the third symbol's shift to the boundary. This explains the downward deviation of curve $N = 3$ in Fig. 4.26 B.

[5]This assumption is relatively well justified regarding the measured distributions of the principal component coefficients shown in Fig. 4.12, distributions not shown.

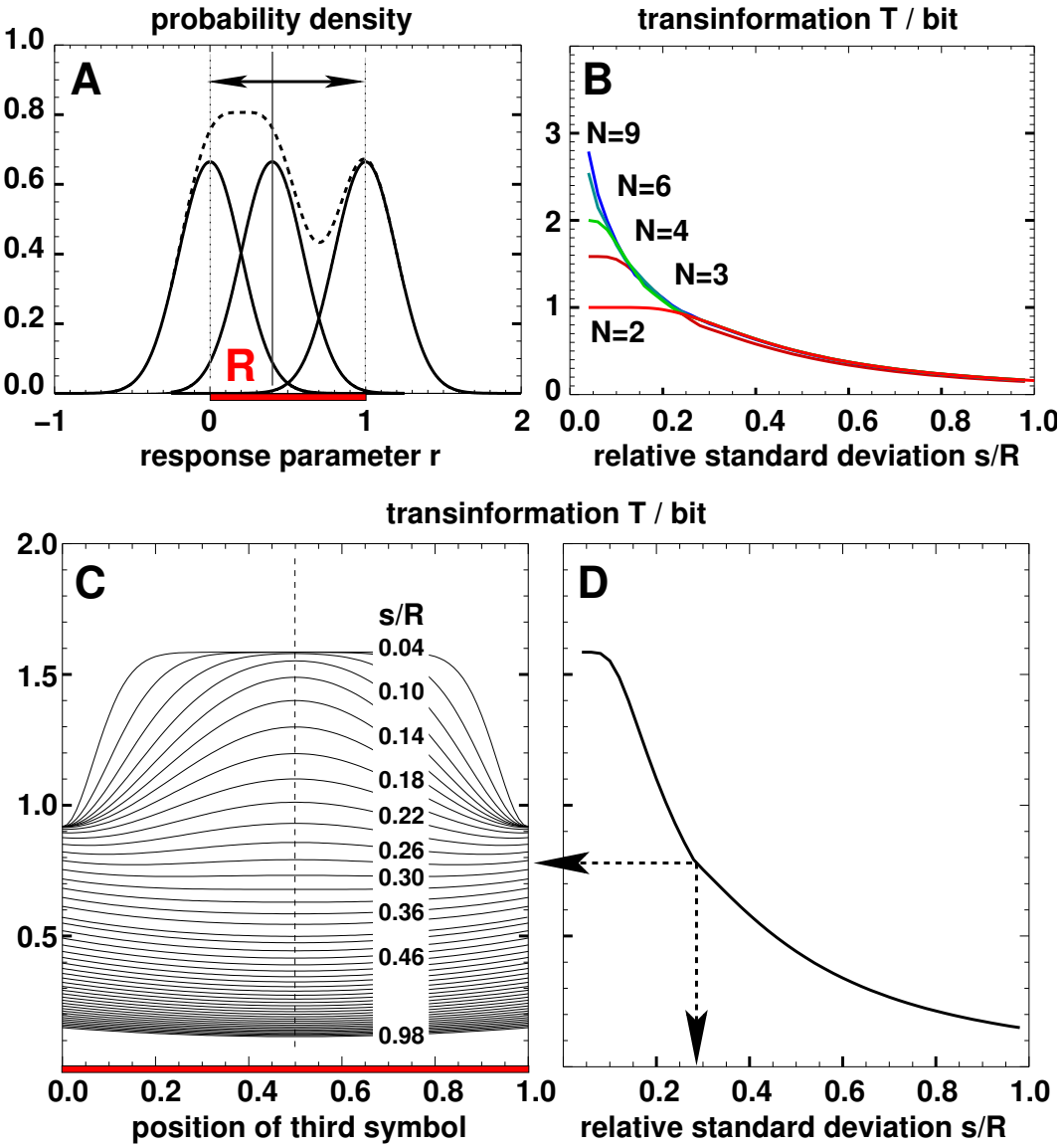

Figure 4.26: Maximized transinformation depending on the response variability and the number of equally probable stimulus symbols. **A**: Example of response range R and conditional distributions for three stimulus symbols. The question is, where the third symbol has to be positioned in order to maximize T? **C** quantifies T dependent on the position of the third symbol. When s/R is small, the maximum is reached when it is located in the center of R. However, if $s/R > 0.28$ the transinformation is minimized when it is located in the center. It has to be positioned at either boundary to maximize T. D depicts the maximized value of T dependent on s/R. The critical point, where two symbols outperform three symbols, is reflected in the bend of the curve. **B**: Maximized T for 2, 3, 4, 6, and 9 stimulus symbols. If $s/R > 0.15$ then $N = 3$ symbols can be discriminated.

High stimulation rate. We found that while LFP often reflected a larger amount of total T compared to MUA, its capability to capture information based on stimulus intensity is inferior. In the example Fig. 4.21 the absolute and relative amount of information about stimulus intensity was larger for MUA in all cases (recording electrodes and sessions). Its average fraction of total T was less than 3% for LFP and about 15% for MUA.

What are the reasons behind the different behavior of MUA and LFP? Any growth of T in response to an increased number of stimulus intensities may be explained by an enhanced effectiveness of the stimulation and a diminished response variability. The mean stimulation current was kept constant in all cases and each current intensity was over-threshold (threshold just less than 20 μA). One has to consider that the profiles for response vs. stimulation current are not linear (see Fig. 4.15). In particular, the profile for LFP showed a stronger effect of saturation compared to MUA. The response variability is relatively constant for LFP in contrast to MUA. Its variability increases with growing response strength. These differences contribute to the distinct behavior of LFP and MUA with respect to the encoding of stimulus intensity. It is to be expected that the variance of the stimulus interval distribution and the exact settings of quantization steps in relation to the stimulation threshold to some degree affects the results. This still needs to be investigated.

Thus far, we have not focused our attention to the fraction of T that is neither explained by the independent encoding of temporal nor intensity information (Fig. 4.21 A, C, "rest"). This fraction is larger for MUA than for LFP. We suspect a joint encoding of temporal and intensity stimulus aspects. Apparently, MUA captures this joint encoding to a larger degree than LFP. Reich et al. (2000) report a remaining information rate of similar percentage – neither explained by visual stimulus contrast nor spatio-temporal pattern – in single unit recordings of monkey V1. They attributed the effect to the confounded encoding of the respective stimulus aspects. Furthermore, we found that the additional amount of information transmitted by increasing the number of intensity steps cannot be assigned a narrow frequency range, neither for LFP nor for MUA. Instead, it spreads over most parts of the spectrum as a rather constant gain factor. Evidently, each frequency band proportionally contributes to the encoding process of the additional stimulus information.

To conclude, stimulus intensity does not seem to be a very effective stimulus parameter with respect to information transmission. About 7 (MUA) or even 20 (LFP) times the information is transmitted using the temporal stimulus aspects. Thus, a code based on them would be more efficient than an intensity code as investigated above. The stimulus current intensity modulates the size of the retinal activity region (Ranck, 1975) rather than the number of elicited spikes per ganglion cell as with normal visual stimulation. If the stimulation field extends to the position of other electrodes, spatial resolution is lost. Humayun et al. (1996) report accordingly that both size and brightness of phosphene percepts in humans grow with

increasing stimulation current. In contrast, the variation of stimulation rate solely leads to brightness modulation.

Intensity quantization. Apparently, a small number of different stimulus intensities, say 2 or 3, is sufficient for *CS*. It is not possible to discriminate between more conditions when the response variance is large. Using the *DS* results and the response model we now address the question, which intensity steps to choose to maximize T. Applying Fig. 4.26 B to the experimental results given in Fig. 4.12 and Table 4.1, a best s/R value of 0.16 suggest $N = 3$ stimulus conditions. It is not possible to increase T when more stimulus conditions are used. If now three current steps are to be chosen, we proceed as follows. According to Fig. 4.26 C two stimulus conditions should be adjusted to yield responses at the boundaries of the range R. The third condition should evoke a response equivalent to the center of R in order to maximize T. Considering Fig. 4.15, the mean vs. current profiles are not linear. Thus, it is not correct to simply adjust the intensity of the third stimulus to be the mean of the intensities of the boundary stimulus conditions. E.g., for the LFP signal A the current amplitudes of the boundary conditions are 10 and 100 μA. Still, the amplitude of the center condition has to be adjusted to about 15 μA just over stimulation threshold.

In practice, the lower boundary condition that corresponds to a zero stimulus may not be useful because zero stimuli transmit information if their time of occurrence is known, e.g., in discrimination task experiments. Thus, with respect to the retina implant only two stimulus conditions remain to be adjusted. The low intensity stimulus should slightly exceed the stimulation threshold whereas the high intensity stimulus should be adjusted to evoke very strong responses, but within the tissue safety range. It should be noted that this rationale applies to information theoretic analyses of evoked responses registered at *one* cortical position at a time. The brain may assess a spatio-temporal code and capture more information if stimulus intensity is encoded that way.

4.4.3 Spatial resolution

The analysis of the achievable spatial resolution was performed within the *CS* paradigm at different mean stimulation rates. For *DS* results we want to refer to (Wilms, 2001) and a paper in preparation. The obtained transinformation profiles quantify the spatial spread of transmitted information based on the whole data set separately for LFP and MUA. In case of single electrode stimulation, the profiles may be similar to cortical “point spread” functions (Das and Gilbert, 1995). Here, they demonstrate one clear maximum whose predominance diminishes at high stimulation rates. The stimulus rate dependence has already been discussed above. The width of the profiles suggests that cortical activation centers have a radius smaller than 1 mm. On the contrary, the profiles for visual stimulation do not show a unique maximum. The LFP performs a zigzag course with three maxima independent of

the mean stimulus rate, while MUA otherwise similar, leaves out the highest peak of LFP completely and becomes smoother the higher the rate (Fig. 4.24 A, B). Apparently, the distinct properties of the profiles for the signal types reflect different cortical processes which encode the stimulus information. These processes will be connected to the anatomical and functional structure of area 17, ocular dominance stripes and orientation columns. Layer 4 of area 17 receives patchy afferent input from the LGN separately for both eyes (Orban, 1984). These zones comprise stripes of ocular dominance being 0.25–0.5 mm wide (Nicholls et al., 1992). At the same time, area 17 is covered with circularly organized orientation columns. A band of 1 mm width represents all orientations 0–180° (Orban, 1984). The orientations are arranged in "pinwheels" that have a singular point in their center, where all orientation slabs join (Nicholls et al., 1992). The density of the pinwheels amounts to about $2/\mathrm{mm}^2$ cortex. The zigzag shape of the LFP transinformation profile (Fig. 4.24 A) may reflect the pattern of cortical ocular dominance or the orientation columns. Since we electrically stimulated one eye only with one retinal stimulation electrode, a stimulus orientation did not exist. Thus, it could not have been encoded by the neurons. Therefore, we rather suspect that the transinformation profile reflects cortical ocular dominance stripes.

There still remains the difficulty to define a clear cortical activation radius by our analysis. This must not lead to the conclusion that the visual resolution is worse than the electrical. The cortical network – consisting of many nonlinear nodes multifoldly connected – has many potentialities to precisely encode localized stimulus information, e.g., in a distributed way. Thus, it seems necessary to also take into account the information hidden in the coupling between cortical positions. For these reasons, we determined transinformation profiles based on various joint recording positions. E.g., the cumulative profiles assess the cumulative amount of T when one electrode after another is included into the joint multi-dimensional recording signal. Alternatively, the exclusive contribution of single electrode positions is captured by taking the amount of T received at all electrodes, minus the information received by all electrodes except the one observed. Thereby, we have removed the linear part of spatio-temporal correlation.

By comparing the independent transinformation profiles with the joint ones we made the following observations: we found some degree of (1) *redundancy*, especially for "weaker" cortical positions in case of electrical stimulation, and slightly for MUA in case of visual stimulation. We noticed a lack of increase in the joint transinformation when we included a certain cortical electrode. This electrode conveyed substantial information when regarded independently. We observed a considerable degree of (2) *synergy*, but only for electrical stimulation. The joint transinformation was enhanced by a larger amount when a certain electrode was joined in, compared to the electrode's independent contribution. Mainly in case of visual stimulation we found quite the (3) *opposite of synergy*: the joint transinformation

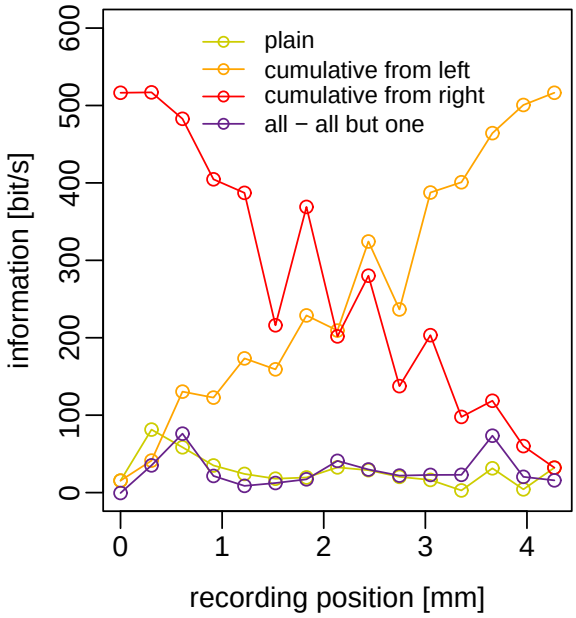

Figure 4.27: Transinformation dependent on the cortical recording position in mm and *electrical* stimulus rate, experiment [140], LFP signal type. The recording positions correspond to electrodes r1, ..., r15. The yellow graph describes the plain profile of T dealing with the electrode positions independently. The orange and red graphs show the cumulative profiles. The violet curve depicts the difference between the information conveyed by all electrodes minus the information of all but the one observed electrode. The electrical stimuli were supplied at random by seven stimulation electrodes with only one being active at a time (100 μA amplitude). The inter-stimulus delay was uniformly distributed between 140 and 190 ms.

was largely reduced when an electrode was added that independently conveyed information.

Redundant and synergistic encoding is well-known in neuroscience (Gawne and Richmond, 1993; Brenner et al., 2000b). But, how can information decrease when an effective recording electrode is added into the joint measure? In case the additional electrode merely produces noise, i.e., non-relevant information with respect to the given stimulus, it is clear that a joint measure is degraded and loses precision. However, if the additional electrode contributes relevant stimulus information one would expect that a joint measure should profit from it. In the example Fig. 4.24 A, C the degree of the opposite of synergy strongly varies with the stimulus rate: it only turns up for a rate of 2 and 8 Hz while rates of 1, 4, 12 Hz and above are without effect.

Fig. 4.27 presents the different types of T profiles based on experiment [140] for one mean stimulus rate of 6 Hz in case of electrical stimulation. Here, 15 record-

ing electrodes linearly positioned at equal distances (305 µm) were available and allowed a precise sampling of the cortical activity. The plain profile (yellow line) does not differ largely from the one with correlation removed (violet line). Interestingly, the cumulative profiles (orange and red graphs) show systematic peaks and drops indicating synergy and its opposite taking turns.

At present we do not have a suitable explanation for these effects except that they seem to be important for the encoding of sensory information. Higher order statistics not captured by linear analysis methods may hide behind. We had expected that the transinformation profiles generally got narrower after removing the spatio-temporal correlation. Possibly, the removal of the linear fraction is not sufficient. As stated above, an alternative view is the notion that localized sensory information is coded in a non-redundant distributed way. Then the widths of activity or transinformation profiles do not reveal much about the achievable spatial resolution. If one uses the radius of 1 mm estimated above and takes into account the eccentricity dependent magnification factor between the visual angle and its corresponding cortex area (approximately 0.8 mm^2/deg^2 at 5° eccentricity, cf. Orban (1984)) a spatial resolution of about 2° results.

Alternatively to these analyses we tested how T depends on the number of independent stimulus channels on a retinal area of approximately 1 mm^2. For both signal types LFP and MUA, the curves (especially for r2 and r3) seem to saturate at the maximum number of stimulus channels, suggesting a resolution of 7 electrodes per mm^2 or 1 electrode per 2° visual angle[6] which is the typical size of the RFs we have measured. Interestingly, the fraction of T based on non-temporal stimulus aspects amounts to 10–20% of total T but does not show a saturation effect. The value is double the fraction of intensity information presented above. A further observation is that LFP is not less efficient in encoding the non-temporal information than MUA – again contrary to the results above. Obviously, spatial aspects of electrical stimuli are more efficiently encoded in the cortex than intensity stimulus aspects. The lack of a saturation effect may indicate that there could still be room in the coding space.

4.4.4 Conclusion

Summarizing, we yield the following results and resolutions possibly achievable by a retina implant:

Temporal resolution. We found extremely precise (300 µs) cortical spike responses belonging to the relatively slow X pathway. We also registered responses of the Y path but did not quantify the variability due to their early occurrence – they were partly obscured by the stimulus artifact. We suspect that the relative timing of single stimulus events at different locations may be perceivable with high precision in the millisecond dimension if the higher cortical circuits are capable to exploit

[6] 1 mm retina corresponds to 5° visual angle.

the precise spatio-temporal patterns in area 17. In case of continuous stimulation the most efficient mean rate with respect to information transmission lies between 20 and 40 Hz. For higher rates strong inhibition is suspected to degrade the transmission of information and prevent useful percepts. Accordingly, single stimulus events showed a stimulus intensity dependent duration of the decorrelated response, which typically was shorter than 40 ms. Stimuli have shown to be most effective if they follow a facilitating stimulus within 10–20 ms or after a very long delay of more than 150 ms. They lose their efficiency if the delay has a medium length.

Intensity resolution. We found that only 2 or 3 stimulus intensities can be encoded efficiently. A finer quantization does not improve the rate of *T*. The stimulus intensity is not a very efficient parameter capturing not more than 3% (LFP) or 15% (MUA) of total *T*. A better parameter would exploit the temporal precision of responses. Here, we suggest to utilize informative bursts of stimulus events. The inter-event intervals (instantaneous rate code) or impulse density (rate code) will more efficiently code the intensity or contrast parameter of the visual scene. We recommend to implement one scarcely over-threshold intensity and another value rather at the other boundary of the physiological useful range evoking very strong responses. If only one intensity step is to be chosen, one should select a well over-threshold but not too high value. The stronger the cortical activation, the higher is the variability of the action potentials and the larger the percept, impairing the spatial resolution. Interestingly, the variance of LFP does not grow with increasing response strength.

Spatial resolution. The halfwidth of cortical transinformation profiles in case of electrical stimulation amounted to about 1 mm suggesting an achievable spatial resolution of 2° visual angle. The notion that an equivalent measure could not be found for visual stimulation may be due to distributed cortical encoding of localized sensory information. Accordingly, despite the findings of some degree of redundancy, we have not observed a shrinking of the profile widths after removing the linearly correlated fraction of spatio-temporal information. However, we found synergistic encoding only for electrical stimulation and the opposite of synergistic encoding mainly in case of visual stimulation. Furthermore, we tested the dependence of *T* on the number of stimulus electrodes per retinal area. Our findings yield that roughly a maximum of 7 stimulation electrodes per mm^2 could efficiently contribute to *T* corresponding to a resolution of 2° visual angle. The fraction of *T* solely based on spatial stimulus aspects accounts for 10–20% of total *T*.

Depending on the exact positioning of the retina implant – especially with respect to eccentricity – and the specific properties of the patient's visual system, individually achievable resolutions will vary. The most convenient stimulation parameters will certainly be found by individual adjustment in direct dialogue with the patient (Humayun et al., 1996, 1999; Eckmiller et al., 1999).

4.5 Appendix

4.5.1 Transinformation for discrete stimuli and continuous response parameters

For n stimulus conditions with probabilities $p_s(s_i)$ and continuous scalar conditional response distributions $p_{r|s_i}(r)$ the transinformation is calculated as

$$\begin{aligned} T(S;R) &= H(R) - H(R|S) \\ &= -\int_{-\infty}^{+\infty} p_r(r) \log_2 p_r(r) dr \\ &\quad + \sum_{i=1}^{n} p_s(s_i) \int_{-\infty}^{+\infty} p_{r|s_i}(r) \log_2 p_{r|s_i}(r) dr \end{aligned} \tag{4.5}$$

(Shannon, 1948) with the total response density $p_r(r)$ consisting of the weighted superimposed conditional densities

$$p_r(r) = \sum_{i=1}^{n} p_s(s_i) p_{r|s_i}(r) \; . \tag{4.6}$$

When the stimulus conditions are equally probable, i.e., $p_s(s_i) = 1/n$ and the conditional densities are normal we have

$$p_{r|s_i}(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(r-r_i)^2}{\sigma^2}} \tag{4.7}$$

and

$$p_r(r) = \frac{1}{\sqrt{2\pi}\sigma n} \sum_{i=1}^{n} e^{-\frac{1}{2}\frac{(r-r_i)^2}{\sigma^2}} \; . \tag{4.8}$$

Then, the calculation of transinformation according to (4.5) is simplified

$$\begin{aligned} T(S;R) &= -\int_{-\infty}^{+\infty} p_r(r) \log_2 p_r(r) dr \\ &\quad - \frac{1}{2} \log_2 [2\pi e \sigma^2]. \end{aligned} \tag{4.9}$$

The first integral term has to be solved numerically. It may be solved analytically in the trivial case when the conditional densities do not overlap. In that case it becomes

$$\begin{aligned} & -\int_{-\infty}^{+\infty} p_r(r) \log_2 p_r(r) dr \\ = \; & -\sum_{i=1}^{n} p_s(s_i) \int_{-\infty}^{+\infty} p_{r|s_i}(r) \log_2 p_s(s_i) p_{r|s_i}(r) dr \\ = \; & -\sum_{i=1}^{n} p_s(s_i) \log_2 p_s(s_i) \int_{-\infty}^{+\infty} p_{r|s_i}(r) dr \\ & + \frac{1}{2} \log_2 [2\pi e \sigma^2] \; . \end{aligned} \tag{4.10}$$

Since the integral over any density equals to one, the transinformation yields

$$T(S;R) = -\sum_{i=1}^{n} p_s(s_i) \log_2 p_s(s_i) \,, \tag{4.11}$$

which is identical with the stimulus entropy $H(S)$. This result is in accordance with the notion that the stimuli can be discriminated without error when the conditional response distributions do not overlap. Thus, no information gets lost.

4.5.2 Transinformation in case of identical stimuli

We regard stimulus evoked response epochs R assuming identical stimuli S at epoch begin. According to Borst and Theunissen (1999) we get an upper bound of T when we apply

$$T_{\text{upper}} = \frac{1}{2} \log_2 \left[1 + \frac{|\langle R \rangle|^2}{\left\langle |R - \langle R \rangle|^2 \right\rangle} \right] \tag{4.12}$$

to each bin. The braces indicate the expectation with respect to stimulus S and the term

$$\frac{|\langle R \rangle|^2}{\left\langle |R - \langle R \rangle|^2 \right\rangle}$$

is equivalent to the signal-to-noise ratio in case of an additive noise model with the signal $\tilde{S} = \langle R \rangle$ and the noise $N = R - \langle R \rangle$. Since

$$\left\langle |R|^2 \right\rangle = \left\langle |N + \langle R \rangle|^2 \right\rangle = \left\langle |N|^2 \right\rangle + |\langle R \rangle|^2 \tag{4.13}$$

the formula (4.12) may be rewritten as

$$T_{\text{upper}} = \frac{1}{2} \log_2 \frac{\left\langle |R|^2 \right\rangle}{\left\langle |N|^2 \right\rangle} \,. \tag{4.14}$$

Even though the input $\tilde{S}$ to our fictitious additive noise channel is constant we obtain T_{upper} as a measure of transmitted information. Obviously, in this scenario no information is transmitted at all. An explanation is that T_{upper} is equivalent to the transinformation resulting from the following situation: the same channel is assumed to be linear, i.e., the noise independent of the input signal amplitude and its input is GWN with variance $\sigma_S^2 = |\langle R \rangle|^2$.

Now, it is possible to compare the efficiency of different single stimuli. The variability of their responses is related to the amplitude of their mean response yielding a bin-wise signal-to-noise ratio. Although neither in the temporal nor frequency domain the *PT* are additive, their temporal course indicates informative sections in the response signal, that may contribute to the overall value of T.

4.5.3 Decorrelation by means of Schmidt's orthogonalization

Here, we consider epochs of output signal sections in response to an identically repeated input. Let $\vec{x}_i$ be a sample vector at the i^{th} position within the epoch comprised of n samples. If samples $\vec{x}_i$ and $\vec{x}_j$ are linearly correlated, i.e.,

$$\vec{x}_j = \vec{\xi}_j + k_{j,i}\,\vec{x}_i\,, \tag{4.15}$$

they may be decorrelated by subtracting the linearly correlated fraction $k_{j,i}\vec{x}_i$ which is equivalent to the projection of $\vec{x}_j$ into the direction of $\vec{x}_i$ with

$$k_{j,i} = \frac{\vec{x}_j^T \vec{x}_i}{\left|\vec{x}_i\right|^2}\,. \tag{4.16}$$

Now, $\vec{x}_i$ and $\vec{\xi}_j$ are orthogonal: $\vec{\xi}_j$ comprises novel information. This orthogonalization may be performed for the whole vector space spanned by the signal section epochs (Rice, 1966)

$$\begin{aligned}
\vec{x}_0 &= \vec{\xi}_0 \\
\vec{x}_1 &= \vec{\xi}_1 + k_{1,0}\,\vec{\xi}_0 \\
\vdots &= \vdots \\
\vec{x}_n &= \vec{\xi}_n + k_{n,0}\,\vec{\xi}_0 + \cdots + k_{n,n-1}\,\vec{\xi}_{n-1}\,,
\end{aligned} \tag{4.17}$$

resulting in a mutually orthogonal set of sample vectors $\vec{\xi}_i$ with $i \in \{1,\ldots,n\}$. The advantage of this approach is that causality may be taken into account if the orthogonalization process is carried out subsequently in the direction of time. Fig. 4.8 is an example of how it may be applied to sensory responses to identical stimuli with respect to the calculation of T. Without an orthogonalization the PT are not additive. A summation leads to an overestimated cumulative transinformation (middle row, cumulative curve not shown). After an orthogonalization (bottom row) fewer samples contribute to T because later samples do not contribute additional information. By the repeated subtraction of correlated data vectors a downward bias is introduced since correlation may occur by chance. To prevent that, we performed a statistical test for zero correlation (Kendall test, implemented in R language, Ihaka and Gentleman (1996)) and set the correlation factor (4.16) to zero if it was not significantly different from zero ($p \leq 0.05$). A detailed derivation and discussion of the approach is presented by Eger and Eckhorn (2002c).

4.5.4 Spatially joint transinformation

Using an indirect method based on a deterministic model we first calculated independent predictions for the available response signal channels (see Chapter 2).

Thereby, T can be computed separately between the stimulus and one response channel at a time. This is achieved by performing PCA on each pair of deterministic response estimate and corresponding original response signal before partial values of transinformation PT are summed up in the principal component coordinate system. The first column of the Figures 4.17, 4.18, 4.19, 4.22, 4.23, 4.24 shows these values of T.

The second and third columns however, yield results of joint transinformation between the stimulus and different n-tupels of recording channels. This was achieved as follows: The n-tupels of responses and deterministic estimates were joined by means of PCA in a combined response space, respectively. In practice, we serially aligned the n-tupels of individual response epochs (or *estimate* epochs, respectively) and performed PCA on the result. Thus, any linear correlation between different samples of the aligned n-tupels is removed. Usually, we did not normalize the amplitude of the signals of different recording channels before serially aligning their data epochs. We suspect that the cortex itself has to tackle its own activity of different power at various locations. Yet, in several examples (Figures not shown) we beforehand normalized the data to unit variance, which did not lead to qualitatively different results.

5 Conclusion and Outlook

5.1 Summary of methods and results

For the information theoretical analysis of the acquired neurophysiological data, it was necessary to develop new methods because existing ones were constrained to unacceptable conditions such as stimulation or recording configurations and thus, were not applicable. To assess the encoding capabilities of the neuronal system on *continuous stimulation* conditions a *three step method* was developed bearing advantages in the following aspects. It is (1) not restricted to single channel Gaussian stimuli, (2) due to the use of PCA for coordinate transformation it is easily applicable to multi-output systems, (3) the calculation of multiple kernels in the frequency domain is computationally less costly then in the time domain, (4) it provides a lower bound of transinformation at least if the signal inherent correlation is linear and the samples are Gaussian distributed, and finally (5) due to its forward direction and adjustable model properties, it allows to investigate the encoding of stimulus aspects.

A second method applies to rapid series of transient *discrete stimuli* and assesses the exclusive encoding of non-temporal stimulus aspects. An advantage is that both the three step method and the second approach may be applied to the same data. By quantifying the fraction of transinformation with respect to different stimulus aspects, the encoding capabilities of the system can be selectively investigated. The application of the developed methods yields the following results and conclusions:

The achievable *temporal* resolution with a retina implant amounts to about 40 ms, i.e., electrical stimulation with a mean impulse rate much higher than 20 Hz will not provide perceptually useful information. Since the fraction of transmitted information based on temporal stimulus aspects accounts for most (50–80%) of the total transinformation T, the temporal coding space should be utilized. The temporal resolution with respect to the detection of timing differences between phosphenes at different locations may even be much higher, probably reaching the millisecond range. This capability may be used to define identity and segregation of objects depending on whether potential object features are closely synchronized or not. Due to the smaller time scale this object coding process may run in parallel to the plain temporal coding of stimulus intensity changes.

The achievable resolution using the stimulus *intensity*, i.e., the injected charge as parameter is only 2–3 quantization steps. The implementation of more than three

different current values per stimulation electrode is not advantageous with respect to information transmission. However, the values have to be adjusted properly: when using two quantization steps, the first has to be set to just produce an over-threshold response, the second to evoke a very strong response. The coding of contrast and intensity of the visual scene cannot sufficiently be encoded alone by using the injected charge as coding parameter. Here, the high temporal resolution should be exploited, e.g., by using a rate code that modulates the stimulus impulse density in proportion to the local intensity or contrast of the visual scene.

The achievable *spatial* resolution amounts to about 2° visual angle or the equivalent percept of a 35 cm object at 10 m distance. The resolution corresponds to an electrode distance of about 0.4 mm. Smaller distances will hardly lead to an increase of T or improve perception except if the brain adjusts its receptive field properties. However, the value will likely depend on the eccentricity and thus, the positioning of the implant on the retina. Towards larger eccentricities the magnification factor between a corresponding cortex and retina area decreases which may lead to a decline of the resolution. Spatial information accounts for 10–20% of total T.

A retina implant that considers the above design rules will probably make a good guess with respect to the adjustment of its parameters. However, individual fine tuning in direct dialogue with the patient may improve perception considerably.

5.2 A vision for blinds

The remarkable success of the cochlear implant for deaf people since the first implantation in 1957 has been inspiring (Hambrecht, 1990). Will a similar success be attainable with the retina implant? It goes without question, a retina implant faces a more difficult task. In case of the cochlear implant, the frequency properties of sound stimuli have to be tonotopically mapped to the auditory cortex. In a similar way, the retina implant has to ensure that two-dimensional spatial properties of visual stimuli are retinotopically mapped to the visual cortex. The number of different frequency channels are not very critical for the cochlear implant because the timing of acoustic events indirectly transmits frequency information. In contrast, the retina implant needs many electrodes to sample the visual scene and thus, deals with a two-dimensional array of electrodes difficult to fixate within the moving eye.

Even though the developmental process of retina implants has advanced considerably in the last years, some problems are still to be solved. Until now it is not yet clear how stimulation of ganglion cell axons can be avoided in favor of stimulation of cell bodies. This is an important issue because the cell bodies encode spatial information using a labeled line code. Axonal stimulation violates the retinotopical mapping to the visual cortex and therefore, leads to information loss. An attempt to correct the mapping in direct dialogue with the patient cannot solve the problem of ambiguous phosphenes (Eckmiller et al., 1999). A possibility to overcome

this difficulty could be to use hybrid implants (Jimbo and Kawana, 1992; Ito et al., 1999) being equipped with cultured neurons directly on the stimulation contacts. The application of growth factors that specifically favor the connection to retinal cell bodies would finally solve the problem of axonal stimulation.

Individual improvement of the performance of the implant with respect to independent navigation of the patient may be expected due to the plasticity of the brain and its capability to adjust. A next step in the development of the retina implant needs to take the "binding problem" into account (Eckhorn and Frien, 1995; Treisman, 1996): the natural capability of the visual system to bind different features of the visual scene together or segregate others into separate objects is fundamental for advanced perception. In this field there is plenty of room left for further research.

5.3 Ethical remarks

The connection of an electronic device to the central nervous system rises ethical questions. Compared to other neuroprostheses such as cortical implants that directly link to brain areas, the retina implant may seem gentle. It can hardly be misused, e.g., to intrude on the personality of the patient. Yet, one should be aware that in any case a technology is being developed which allows to take intentional influence on cortical activity – be it for the good or for the bad – if the neuronal code is better understood. Here, science has an ethical responsibility that is easily neglected under the every day pressure of success and financial constraints. In a society in which information is more valuable than money, science may be lured to prepare the way for dubious innovations. Undoubtedly, it is advisable to envision new possibilities – many things are possible today that were never dreamt of before. Yet, not everything that is possible is really needed and many needs are merely created.

A vision aid for blinds may be utterly helpful for independent navigation. Yet, there may be people who do not want to have it. The individual benefit has to outweigh the costs. In any case it is important to encounter the disabled patients as humans with dignity; the step from subjects to objects is easily taken. The study of feasibility ("Machbarkeitsstudie", Eckmiller et al., 1995) that preceded the foundation of the two branches of the German retina implant project has devoted a full chapter to medical ethics.

Nonetheless, there is no protection against tactless reports. As an example, on 24th of November 1996 the German weekly newspaper Bild am Sonntag printed an article with the bold headline "Blinde lesen Briefe"[1]. Later, on November 18th 1999, the Bildzeitung announced "Kamera-Brille – die neue Hoffnung für viele Blinde"[2]. At the present stage of the project these doubtful *news* stir unjustified hope and may cause disappointment.

[1] The Blind read mail

[2] Camera glasses – the new hope for many blinds

Yet, science is drawn into the role of the creator. A verse from the Bible is often quoted: “the blind receive sight, the lame walk, the deaf hear” (Matt. 11, 5). However, the quoted verse is abbreviated and continues: “the dead are raised and the good *news* is preached to the poor”. The second part is not less important than the first. The true creator is concerned with more dimensions of life than physical health. This is really good *news*.

A Appendix

A.1 Lower bound of transinformation on the basis of an indirect forward method

Let the deterministic model P of a system's response R be the input and respective output of a fictitious information channel disturbed by additive noise $N = R - P$. The transinformation is given by

$$T(P;R) = H(R) - H(R|P) \ .$$

- In case the noise is independent of the model we have

$$T(P;R) = H(R) - H(N) \ .$$

- If the noise is not independent of the model, the irrelevance is decreased, i.e.,

$$H(R|P) = H(N|P) < H(N)$$

  (Cover and Thomas, 1991). Thus, the real amount of T is larger than $H(R) - H(N)$ though not violating a lower bound demand.

- If the noise is not normal but normality is assumed, e.g., if the Shannon formula

$$H = 0.5 \log_2 2\pi e \sigma_n^2$$

  is used, its real entropy $H(N)$ is overestimated which again leads to an underestimate of T.

- If the noise contains nonlinear dependencies between samples, its entropy is overestimated in case these dependencies are not taken into account by the entropy calculation. Hereby, also T is underestimated.

- If the signal R contains dependencies between samples that are not assessed, or if the samples are not Gaussian distributed though this is implicitly assumed, the entropy $H(R)$ is overestimated which results in an overestimation of T. Here, the lower bound condition may be violated.

The same arguments apply if the estimation of T is based on the effective noise (see Section 2.5.8).

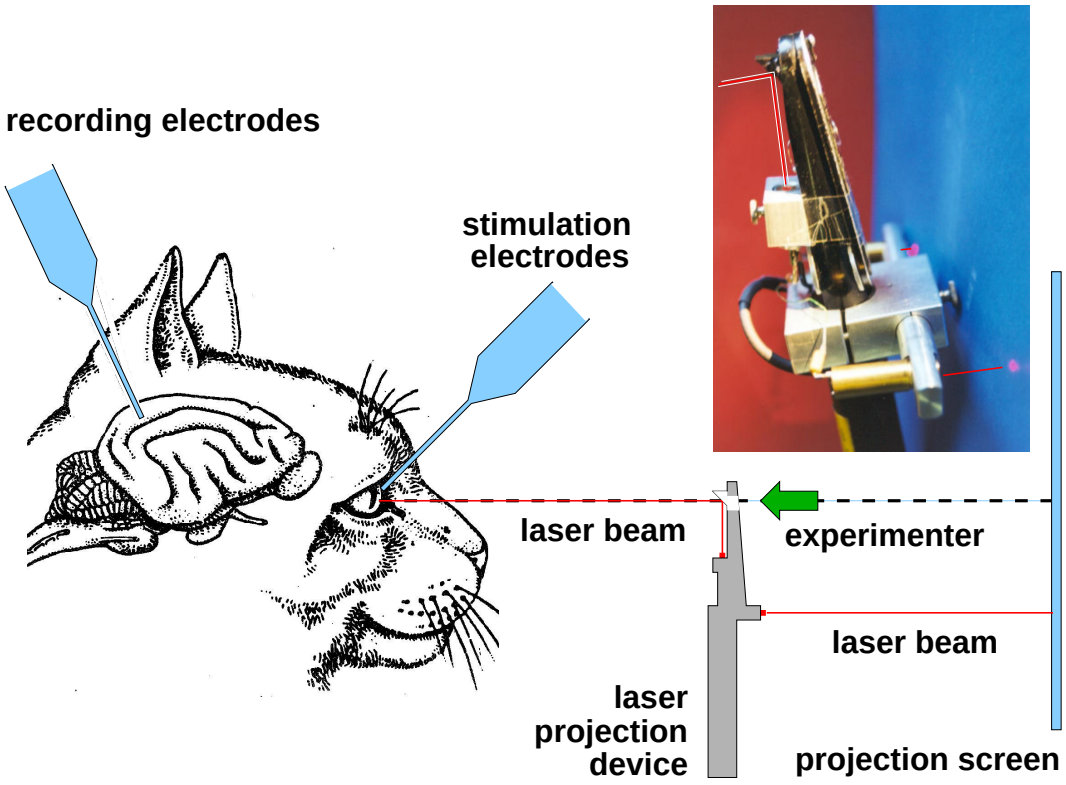

Figure A.1: Scheme for mapping the retinal stimulation positions. The experimenter (green arrow) directs the target laser beam onto the retina while viewing the retinal surface with the ophthalmoscope. He focuses the location of the stimulation electrodes with the laser pinpoint. Another person maps the positions of the back projecting laser beams on the projection screen using a stencil to correct for the translation offset.

A.2 Development of a laser projection device

A requirement for efficient electrical stimulation is the electrical stimulation at a position corresponding to the receptive fields of the recording electrodes. Therefore, the author developed a laser projection device. The finding of an appropriate stimulation position is ensured by the scheme depicted in Fig. A.1. While the experimenter focuses the laser beam onto the retinal stimulation site another person marks the position of the back projecting beams on the screen. The accuracy of the method is about 1° visual angle.

A.3 Development of miniature current sources

For the exact positioning of the retinal stimulation electrodes it is necessary to also be able to record from them. The measurable increase of “noise” while lowering the electrode tips onto the retina signals the proximity to the ganglion cells. For

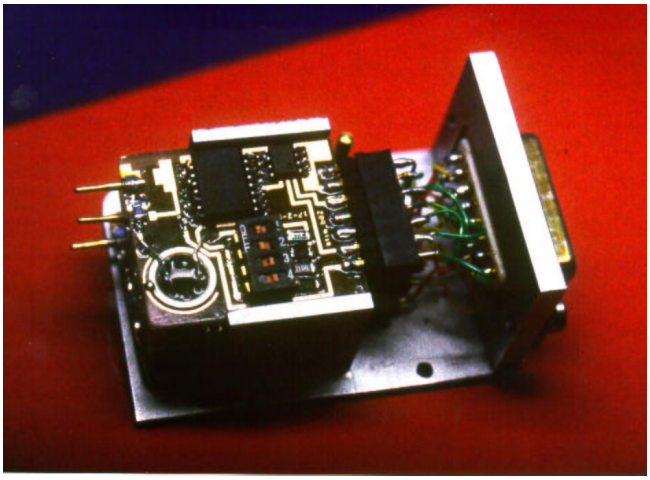

Figure A.2: Opened electronics part of 3 channel fiber electrode matrix with three current source devices.

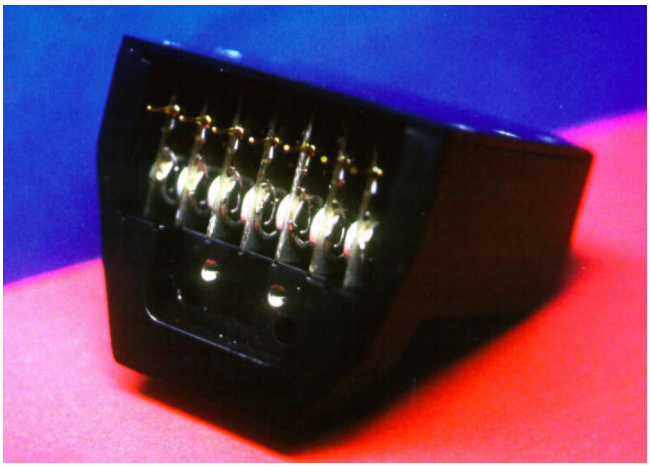

Figure A.3: Electronics part of 7 channel fiber electrode matrix with seven current source devices.

this reason, low-bias voltage-controlled current sources in combination with high impedance recording amplifiers were developed. The idea to allow for simultaneous low noise recording stems from T. Schanze who holds the copyright for the circuit diagram (therefore not shown). The layout, the implementation, and part of the dimensioning was done by the author. Fig. A.2, A.3 show the current sources assembled in the electronics part of the electrode drives (Eckhorn and Thomas, 1993).

Bibliography

Agnew WF, McCreery DB, editors: Neural Prostheses: Fundamental Studies. Prentice Hall, New Jersey, 1990.

Arieli A, Sterkin A, Grinvald A, Aertsen A: Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science, 1996, 273: 1868 – 1871.

Bair W: Spike timing in the mammalian visual system. Curr. Opin. Neurobiol., 1999, 9: 447–453.

Bair W, Koch C: Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Comput., 1996, 8: 1185–1202.

Bercher JF, Vignat C: Estimating the entropy of a signal with applications. IEEE Trans. Signal. Process., 2000, 48: 1687–1694.

Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D: Reading a neural code. Science, 1991, 252: 1854–1857.

Bode HW, Shannon CE: A simplified derivation of linear least square smoothing and prediction theory. Proc. I. R. E., 1950, 38: 417–425.

Borst A, Theunissen FE: Information theory and neural coding. Nat. Neurosci., 1999, 2: 947–957.

Brenner N, Bialek W, de Ruyter van Steveninck R: Adaptive rescaling maximizes information transmission. Neuron, 2000a, 26: 695–702.

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck R: Synergy in a neural code. Neural Comput., 2000b, 12: 1531–1552.

Brindley GS, Lewin WS: The sensations produced by electrical stimulation of the visual cortex. J. Physiol., 1968, 196: 479–493.

Brummer SB, Turner MJ: Electrical stimulation with Pt electrodes: II. Estimation of maximum surface redox limits. IEEE Trans. Biomed. Eng., 1977, 24: 440–443.

Buracas GT, Albright TD: Gauging sensory representation in the brain. Trends Neurosci., 1999, 22: 303–309.

Buracas GT, Zador AM, DeWeese MR, Albright TD: Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. Neuron, 1998, 20: 959–969.

Buß R, Groß M, Prämaßing F, Püttjer D, Jäger D: Optical signal und energy transmission for implantable intraocular microsystems. In Proc. World Microtechnologies Congress MICRO.tec 2000, volume 2, Hannover, 2000 pages 461–465.

Chow AY, Chow VY: Subretinal electrical stimulation of the rabbit retina. Neurosci. Lett., 1997, 225: 13–16.

Comon P: Independent component analysis, a new concept? Signal Process., 1994, 36: 287–314.

Cover TM, Thomas JA: Elements of Information Theory. John Wiley and Sons, Inc., New York, 1991.

Creutzfeldt OD, Kuhnt U: Central Processing of Visual Information. B: Visual Centers in the Brain, volume VII/3 of *Handbook of Sensory Physiology*. Springer, Berlin, Heidelberg, New York, 1973 pages 595–646.

Das A, Gilbert CD: Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. Nature, 1995, 375: 780–784.

Dawson WW, Radtke ND: The electrical stimulation of the retina by indwelling electrodes. Invest. Ophtalmol. Visual Sci., 1977, 16: 249–252.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W: Reproducibility and variability in neural spike trains. Science, 1997, 275: 1805–1808.

de Ruyter van Steveninck R R Laughlin SB: The rate of information transfer at graded-potential synapses. Nature, 1996, 379: 642–645.

Dean AF: The variability of discharge of simple cells in the cat striate cortex. Exp. Brain Res., 1981, 44: 437–440.

deCharms RC, Zador A: Neural representation and the cortical code. Annu. Rev. Neurosci., 2000, 23: 613–647.

Deco G, Obradovic D: An Information-Theoretic Approach to Neural Computing. Springer, 1997.

DeWeese MR, Meister M: How to measure the information from one symbol. Network, 1999, 10: 325–340.

Dimitrov AG, Miller JP: Neural coding and decoding: Communication channels and quantization. Network, 2001, 12: 441–472.

Dimitrov AG, Miller JP, Aldworth Z, Parker AE: Spike pattern-based coding schemes in the cricket cercal sensory system, 2002, in press.

Dobelle WH: Artificial vision for the blind by connecting a television camera to the visual cortex. ASAIO Journal, 2000, 46: 3–9.

Doty RW: Central Processing of Visual Information. B: Visual Centers in the Brain, volume VII/3 of *Handbook of Sensory Physiology*. Springer, Berlin, Heidelberg, New York, 1973 pages 483–541.

Eckhorn R, Frien A: Neural signals as indicators of spatial and temporal segmentation coding in the visual cortex. In R Moreno-Diaz, J Mira-Mira, editors, Proc. Int. Conf. on Brain Processes, Theories and Models, 1995, MIT Press, 1995 pages 381–390.

Eckhorn R, Grüsser OJ, Kröller J, Pellnitz K, Pöpel B: Efficiency of different neuronal codes: Information transfer calculations for three different neuronal systems. Biol. Cybern., 1976, 22: 49–60.

Eckhorn R, Krause F, Nelson JI: The rf-cinematogram. a cross-correlation technique for mapping several visual receptive fields at once. Biol. Cybern., 1993, 69: 37–55.

Eckhorn R, Pöpel B: Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. Kybernetik, 1974, 16: 191–200.

Eckhorn R, Pöpel B: Rigorous and extended application of information theory to the afferent visual system of the cat. II. Experimental results. Biol. Cybern., 1975, 17: 7–17.

Eckhorn R, Pöpel B: Responses of cat retinal ganglion cells to the random motion of a spot stimulus. Vision Res., 1981, 21: 435–443.

Eckhorn R, Querfurth H: Information transmission by isolated frog muscle spindle. Biol. Cybern., 1985, 52: 165–176.

Eckhorn R, Stett A, Schanze T, Gekeler F, Schwahn H, Zrenner E, Wilms M, Eger M, Hesse L: Physiologische Funktionsprüfungen von Retinaimplantaten an Tiermodellen. Ophthalmologe, 2001, 98: 369–375.

Eckhorn R, Thomas U: A new method for the insertion of multiple microprobes into neural and muscular tissue, including fiber electrodes, fine wires, needles, and microsensors. J. Neurosci. Methods, 1993, 49: 175–179.

Eckmiller R, Eckhorn R, Hömberg V, Hosticka B, Jäger D, Klar H, Linke D, Gersonde K, Zimmer G, Noth J, Samii M, Ehrfeld W, Heuberger A: Neurotechnologie-Report Gesamt-Kurzfassung. Technical report, Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, Germany, Bonn, 1995.

Eckmiller R, Hünermann R, Becker M: Exploration of a dialog-based tunable retina encoder for retina implants. Neurocomputing, 1999, 26-27: 1005–1011.

Efron B: The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, 6th edition, 1994.

Efron B, Tibshirani R: Statistical data analysis in the computer age. Science, 1991, 253: 390–395.

Eger M, Eckhorn R: Temporal, spatial, and intensity information in cat visual cortex evoked by a retina-implant, 2001a, 4th Neural Coding Workshop, Plymouth, UK, 10-15 Sept. 2001 (Abstract).

Eger M, Eckhorn R: Three-step-approach for the analysis of information transmission via neural systems with multiple inputs and outputs, 2001b, 4th Neural Coding Workshop, Plymouth, UK, 10-15 Sept. 2001 (Abstract).

Eger M, Eckhorn R: Assessing the encoding of stimulus attributes with rapid sequences of stimulus events. J. Comput. Neurosci., 2002a, in press.

Eger M, Eckhorn R: A model-based approach for the analysis of neuronal information transmission in multi-input and -output systems. J. Comput. Neurosci., 2002b, 12: 175–200.

Eger M, Eckhorn R: Quantification of sensory information transmission using time-series decorrelation techniques. Biosystems, 2002c, accepted.

Eger M, Eckhorn R, Schanze T: Methods of calculating retino-cortical information transmission for multi-dimensional visual and electrical stimuli in the cat's visual system. In Göttingen Neurobiology Report. Proc. 28th Göttingen Neurobiology Conf., 2001 (Abstract #583).

Eger M, Schanze T, Wilms M, Eckhorn R: Real-time suppression of stimulus artifacts by subtraction of adaptive artifact estimate. In Göttingen Neurobiology Report. Proc. 26th Göttingen Neurobiology Conf., 1998 (Abstract #477).

Eger M, Schanze T, Wilms M, Eckhorn R: Electrical micro-stimulation of cat retina evokes precisely timed cortical spikes: A possibility for information transmission using a retina implant. In Göttingen Neurobiology Report. Proc. 27th Göttingen Neurobiology Conf., 1999 (Abstract #462).

Ferster D, Lindström S: An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. J. Physiol., 1983, 342: 181–215.

Finn R: Neural prosthetics come of age as research continues. The Scientist, 1997, 11: 13–16.

Freeman JA: An electronic stimulus artifact suppressor. Electroencephalogr. Clin. Neurophysiol., 1971, 31: 170–172.

Gawne TJ, McClurkin JW, Richmond BJ, Optican LM: Lateral geniculate neurons in behaving primates. III. Response predictions of a channel model with multiple spatial-to-temporal filters. J. Neurophysiol., 1991, 66: 809–823.

Gawne TJ, Richmond BJ: How independent are the messages carried by adjacent inferior temporal cortical neurons? J. Neurosci., 1993, 13: 2758–2771.

Gershon ED, Wiener MC, Latham PE, Richmond BJ: Coding strategies in monkey V1 and inferior temporal cortices. J. Neurophysiol., 1998, 79: 1135–1144.

Glaser EM, Ruchkin DS: Principles of Neurobiological Signal Analysis. Academic Press, New York, 1st edition, 1976.

Golomb D, Hertz JA, Panzeri S, Treves A, Richmond BJ: How well can we estimate the information carried in neuronal responses from limited samples? Neural Comput., 1997, 9: 649–665.

Gray RM: Toeplitz and circulant matrices: A review. http://ee.stanford.edu/∼gray/toeplitz.pdf, 2001.

Groß M, Alder T, Buß R, Heinzelmann R, Meininger M, Jäger D: Micro photovoltaic cell array for energy transmission into the human eye. In Proc. 14th European Photovoltaic Solar Energy Conf., volume 1, Barcelona, 1997 pages 1165–1167.

Grumet AE, Wyatt JL, Rizzo JF: Multi-electrode stimulation and recording in the isolated retina. J. Neurosci. Methods, 2000, 101: 31–42.

Guenther E, Tröger B, Schloßhauer B, Zrenner E: Long-term survival of retinal cell cultures on retinal implant materials. Vision Res., 1999, 39: 3988–3994.

Gur M, Beylin A, Snodderly DM: Response variability of neurons in primary visual cortex (V1) of alert monkeys. J. Neurosci., 1997, 17: 2914–2920.

Haag J, Borst A: Encoding of visual motion information and reliability in spiking and graded potential neurons. J. Neurosci., 1997, 17: 4809–4819.

Hambrecht FT: Neural Prostheses: Fundamental Studies. Prentice Hall, New Jersey, 1990 pages 2–23.

Hesse L, Schanze T, Wilms M, Eger M: Implantation of retina stimulation electrodes and recording of electrical stimulation responses in the visual cortex of the cat. Graefe's Arch. Clin. Exp. Ophthalmol., 2000, 238: 840–845.

Hinich M: Testing for Gaussianity and linearity of a stationary time series. J. Time Ser. Anal., 1982, 3: 169–176.

Humayun MS, de Juan E, Dagnelie G, Greenberg RJ, Propst RH, Phillips DH: Visual perception elicited by electrical stimulation of retina in blind humans. Arch. Ophthalmol., 1996, 114: 40–46.

Humayun MS, de Juan E, Weiland JD, Dagnelie G, Katona S, Greenberg RJ, Suzuki S: Pattern electrical stimulation of the human retina. Vision Res., 1999, 39: 2569–2576.

Ihaka R, Gentleman R: R: A language for data analysis and graphics. J. Comput. Graph. Statist., 1996, 5: 299–314.

Ito Y, Yagi T, Tanaka S, Kanda H, Watanabe M, Uchikawa Y: Cultures of neurons on micro-electrode array in hybrid retinal implant. In Proc. 1999 IEEE Int. Conf. on Systems, Man and Cybernetics, IEEE SMC'99, 1999 pages 414–417.

Jimbo Y, Kawana A: Electrical stimulation and recording from cultured neurons using a planar electrode array. Bioelectrochemistry and Bioenergetics, 1992, 29: 193–204.

Kandel ER, Schwartz JH, Jessell TM, editors: Essentials of Neural Science and Behaviour. Appleton & Lange, Stamford, 1995.

Kara P, Reinagel P, Reid RC: Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. Neuron, 2000, 27: 635–646.

Karhunen J, Wang L, Vigario R: Nonlinear PCA type approaches for source separation and independent component analysis. In IEEE Proc. ICNN '95, 1995 pages 995–1000.

Kato S, Saito M, Tanino T: Response of the visual system evoked by an alternating current. Medical and Biological Engineering and Computing, 1983, 21: 47–50.

Kjaer TW, Hertz JA, Richmond BJ: Decoding cortical neuronal signals: Network models, information estimation and spatial tuning. J. Comput. Neurosci., 1994, 1: 109–139.

Kohler K, Hartmann JA, Werts D, Zrenner E: Histologische Untersuchungen zur Netzhautdegeneration und zur Gewebeverträglichkeit subretinaler Implantate. Ophthalmologe, 2001, 98: 364–368.

Krausz HI: Identification of nonlinear systems using random impulse train inputs. Biol. Cybern., 1974, 19: 217–230.

Lee YW, Schetzen M: Measurement of the Wiener kernels of a non-linear system by cross-correlation. Int. J. Contr., 1965, 2: 237–255.

Linsker R: Self-organization in a perceptual network. IEEE Computer, 1988, 21: 105–117.

Loader C: Local Regression and Likelihood. Statistics and Computing, Springer, 1st edition, 1999.

Loader CR: Locfit: An introduction. Statistical Computing and Graphics Newsletter, 1997.

Loeb GE: Neural prosthetic interfaces with the nervous system. Trends Neurosci., 1989, 12: 195–201.

MacKay DM, McCulloch WS: The limiting information capacity of a neuronal link. Bulletin of Mathematical Biophysics, 1952, 14: 127–135.

MacWilliams FJ, Sloane NJA: Pseudo-random sequences and arrays. Proceedings of the IEEE, 1976, 64: 1715–1729.

Makhoul J: On the eigenvectors of symmetric Toeplitz matrices. IEEE Trans. Acoust. Speech Signal Process., 1981, 29: 868–872.

Marmarelis PZ, McCann GD: Development and application of white-noise modeling techniques for studies of insect visual nervous system. Biol. Cybern., 1973, 12: 74–89.

Mendel JM: Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. Proceedings of the IEEE, 1991, 79: 278–305.

Mitzdorf U: Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and EEG phenomena. Physiol. Rev., 1985, 65: 37–100.

Mitzdorf U, Li BH, Pöppel E: Mass-action view of single-cell responses to stimulation of the receptive field and/or beyond: Exemplification with data from the rabbit primary visual cortex. Electroencephalogr. Clin. Neurophysiol., 1994, 92: 442–455.

Moddemeijer R: On estimation of entropy and mutual information of continuous distributions. Signal Process., 1989, 16: 233–248.

Moddemeijer R: A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations. Signal Process., 1999, 75: 51–63.

Nicholls JG, Martin AR, Wallace BG: From Neuron to Brain. Sinauer Associates, Inc., Sunderland, Massachusetts, 3rd edition, 1992.

Nowak KG, Bullier J: Axons, but not cell bodies are activated by electrical stimulation in cortical gray matter. I. evidence from chronaxie measurements. Exp. Brain Res., 1998a, 118: 477–488.

Nowak KG, Bullier J: Axons, but not cell bodies are activated by electrical stimulation in cortical gray matter. I. Evidence from selective inactivation of cell bodies and axonal initial segments. Exp. Brain Res., 1998b, 118: 489–500.

Oppenheim AV, Schafer RW: Discrete-time signal processing. Prentice-Hall, 1989.

Optican LM, Richmond BJ: Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J. Neurophysiol., 1987, 57: 162–178.

Orban GA: Neuronal Operations in the Visual Cortex, volume 11 of *Studies of Brain Function*. Springer, Berlin, Heidelberg, New York, Tokyo, 1984.

Palm G, Pöpel B: Volterra representation and Wiener-like identification of nonlinear systems: Scope and limitations. Quart. Rev. Biophys., 1985, 18: 135–164.

Panzeri S, Treves A: Analytical estimates of limited sampling bias in different information measures. Network, 1996, 7: 87–107.

Parsa V, Parker PA, Scott RN: Adaptive stimulus artifact reduction in noncortical somatosensory evoked potential studies. IEEE Trans. Biomed. Eng., 1998, 45: 165–179.

Plumbley MD: On information theory and unsupervised neural networks. Technical report, Cambridge University Engineering Department, UK, 1991.

Potts AM, Buffum D: The electrically evoked reponse of the visual system (EER). Investigative Ophthamology and Visual Science, 1968, 7: 269–278.

Priestley MB: State-dependent models: A general approach to non-linear time series analysis. J. Time Ser. Anal., 1980, 1: 47–71.

Ranck JB: Which elements are excited in electrical stimulation of mammalian central nervous system: A review. Brain Res., 1975, 98: 417–440.

Rao TS, Gabr MM: A test for linearity of stationary time series. J. Time Ser. Anal., 1980, 1: 145–158.

Reich DS, Mechler F, Victor JD: Formal and attribute-specific information in primary visual cortex. J. Neurophysiol., 2000, 85: 305–318.

Reich DS, Victor JD, Knight BW, Ozaki T, Kaplan E: Response variability and timing precision of neuronal spike trains in vivo. J. Neurophysiol., 1997, 77: 2836–2841.

Reinagel P, Reid RC: Temporal coding of visual information in the thalamus. J. Neurosci., 2000, 20: 5392–5400.

Reitboeck HJ: Fiber microelectrodes for electrophysiological recordings. J. Neurosci. Methods, 1983, 8: 249–262.

Rice JR: Experiments on Gram-Schmidt orthogonalization. Math. Comput., 1966, 10: 325–328.

Richmond BJ: The Relationship Between Neuronal Codes and Cortical Organization. Wiley, New York, 1st edition, 1998 .

Richmond BJ, Optican LM: Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. J. Neurophysiol., 1987, 57: 147–161.

Richmond BJ, Optican LM, Podell M, Spitzer H: Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. J. Neurophysiol., 1987, 57: 132–146.

Rieke F, Bodnar DA, Bialek W: Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc. R. Soc. Lond. B Biol. Sci., 1995, 262: 259–265.

Rieke F, Warland DK, de Ruyter van Steveninck RR, Bialek W: Spikes: Exploring the Neural Code. The MIT Press, Cambridge, 1st edition, 1998.

Rife DC, Boorstyn RR: Single-tone parameter estimation from discrete-time observations. IEEE Trans. Inf. Theory, 1974, 20: 591–598.

Roby RJ, Lettich E: A simplified circuit for stimulus artifact suppression. Electroencephalogr. Clin. Neurophysiol., 1975, 39: 85–87.

Roddey JC, Girish B, Miller JP: Assessing the performance of neural encoding models in the presence of noise. J. Comput. Neurosci., 2000, 8: 95–112.

Schanze T: Struktur und Kopplung reizabhängiger rhythmischer Aktivität der Sehrinde. Ph.D. thesis, Fachbereich Physik der Philipps-Universität Marburg, 1995.

Schanze T, Eckhorn R, Hesse L, Eger M, Wilms M, Kossler R, Nebeling B: Experimental setup for assessing the efficacy and quality of retina implant stimulations by retinal and cortical recording in cat. In Göttingen Neurobiology Report. Proc. 26th Göttingen Neurobiology Conf., 1998 (Abstract #476).

Shannon CE: A mathematical theory of communication. Bell Syst. Tech. J., 1948, 27: 379–423, 623–656.

Shannon CE: Communication in the presence of noise. Proc. I. R. E., 1949, 37: 10–21.

Shimazu K, Miyake Y, Watanabe S: Retinal ganglion cell response properties in the transcorneal electrically evoked response of the visual system. Vision Res., 1999, 39: 2251–2260.

Snedecor GW, Cochran WG: Statistical Methods. Iowa State University Press, 8th edition, 1989.

Stanford LR: Conductionvelocity variations minimize conduction time differences among retinal ganglion cell axons. Science, 1987, 238: 358–360.

Stein RB: The information capacity of nerve cells using a frequency code. Biophys. J., 1967, 7: 797–825.

Stein RB, French AS: Excitatory Synaptic Mechanisms. Universitetsforlaget, Oslo, 1970 pages 247–257.

Stett A, Barth W, Weiss S, Hämmerle H, Zrenner E: Electrical multisite stimulation of the isolated chicken retina. Vision Res., 2000, 40: 1785–1795.

Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W: Entropy and information in neural spike trains. Phys. Rev. Lett., 1998, 80: 197–200.

Sutter EE: Imaging visual function with multi-focal m-sequence technique. Vision Res., 2001, 41: 1241–1255.

Theunissen F, Miller JP: Temporal encoding in nervous systems: A rigorous definition. J. Comput. Neurosci., 1995, 2: 149–162.

Theunissen FE: An Investigation of Sensory Coding Principles Using Advanced Statistical Techniques. Ph.D. thesis, University of California at Berkeley, Berkeley, 1993.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL: Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network, 2001, 12: 289–316.

Tolhurst DJ, Movshon JA, Thompson ID: The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Exp. Brain Res., 1981, 41: 414–419.

Tootell RBH, Silverman MS, Switkes E, De Valois RL: Deoxyglucose analysis of retinotopic organization in primate striate cortex. Science, 1982, 218: 902–904.

Treisman A: The binding problem. Curr. Opin. Neurobiol., 1996, 6: 171–178.

Troy JB, Robson JG: Steady discharges of X and Y retinal ganglion cells of cat under photopic illuminance. Vis. Neurosci., 1992, 9: 535–553.

Tsodyks M, Kenet T, Grinvald A, Arieli A: Linking spontaneous activity of single cortical neurons and the underlying functional architecture. Science, 1999, 286: 1943–1946.

Tusa RJ, Palmer A, Rosenquist AC: The retinotopic organization of area 17 (striate cortex) in the cat. J. Comput. Neurosci., 1979, 177: 213–236.

Unser M: On the approximation of the discrete Karhunen-Loeve transform for stationary processes. Signal Process., 1984, 7: 231–249.

Usrey WM, Alonso JM, Reid RC: Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. J. Neurosci., 2000, 20: 5461–5467.

Usrey WM, Reppas JB, Reid RC: Specifity and range of retinogeniculate connections. J. Neurophysiol., 1999, 82: 3527–3540.

van de Grind WA, Grüsser OJ, Lunkenheimer HU: Central Processing of Visual Information. A: Integrative Functions and Comparative Data, volume VII/3 of *Handbook of Sensory Physiology*. Springer, Berlin, Heidelberg, New York, 1973 pages 431–573.

Venables WN, Ripley BN: Modern Applied Statistics with S-Plus. Statistics and Computing, Springer, 3rd edition, 1999.

Victor JD: Temporal aspects of neuronal coding in the retina and lateral geniculate. Network, 1999, 10: R1–R66.

Warland DK, Reinagel P, Meister M: Decoding visual information from a population of retinal ganglion cells. J. Neurophysiol., 1997, 78: 2336–2350.

Werner G, Mountcastle VB: Neural activity in mechanoreceptive cutaneous afferents: Stimulus-response relations, Weber functions, and information transmission. J. Neurophysiol., 1965, 28: 359–397.

Wiener MC, Richmond BJ: Using response models to study coding strategies in monkey visual cortex. Biosystems, 1998, 48: 279–286.

Wiener MC, Richmond BJ: Using response models to estimate channel capacity for neuronal classification of stationary visual stimuli using temporal coding. J. Neurophysiol., 1999, 82: 2861–2875.

Wiener N: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The Technology Press of MIT and John Wiley & Sons, 1950.

Wiener N: Nonlinear Problems in Random Theory. The Technology Press of MIT and John Wiley & Sons, 1958.

Wilms M: Electrical Receptive Fields and Cortical Activation Spread in Response to Electrical Retina Stimulation. Assessment of Spatio-Temporal Resolution for a Retina-Implant. Ph.D. thesis, Physics Department, Philipps-University of Marburg, Germany, Marburg, Germany, 2001.

Wilms M, Schanze T, Eger M, Eckhorn R: Cortical activity distributions in cat area 17/18 elicited by short visual and electrical retinal point stimuli: Investigations for a retina-implant. In Göttingen Neurobiology Report. Proc. 28th Göttingen Neurobiology Conf., 2001 (Abstract #588).

Yang HH, Amari S, Cichocki A: Information theoretic approach to blind separation of sources in non-linear mixture. Signal Process., 1998, 64: 291–300.

Zrenner E: “Retina Implantate” – Ersatz von Retinafunktionen durch technische Implantate: Ein gangbarer Weg zur Wiederherstellung des Sehens? Ophthalmologe, 2001, 98: 353–356.

Zrenner E, Gekeler F, Gabel VP, Graf HG, Graf M, Guenther E, Hämmerle H, Höfflinger B, Kobuch K, Kohler K, Nisch W, Sachs H, Schloßhauer B, Schubert M, Schwahn H, Stelzle M, Stett A, Tröger B, Weiss S: Subretinales Mikrophotodioden-Array als Ersatz für degenerierte Photorezeptoren? Ophthalmologe, 2001, 98: 357–363.

Zrenner E, Stett A, Weiss S, Aramant RB, Guenther E, Kohler K, Miliczek KD, Seiler MJ, Hämmerle: Can subretinal photodiodes successfully replace degenerated photoreceptors? Vision Res., 1999, 39: 2555–2567.

Abbreviations

A17	area 17, belongs to primary visual cortex (physiologically)
AR	autoregressive
BBR	broad band recordings
CPT	cumulative transinformation values
CS	continuous stimulation
CSD	current source density
DS	discrete stimulation
EPSP	excitatory postsynaptic potential
GWN	Gaussian white noise
ICA	independent component analysis
IPSP	inhibitory postsynaptic potential
LFP	local field potentials
LGN	lateral geniculate nucleus
MUA	multi unit activity
PCA	principal component analysis
PT	partial transinformation values
RF	receptive field
SNR	signal-to-noise ratio
SUA	single unit activity
T	transinformation
V1	primary visual cortex (functionally)
X	denotes the X pathway of the visual system
Y	denotes the Y pathway of the visual system

Glossary

Action potentials Rapid, transient, all-or-none nerve impulses of about 1 ms duration. They have been elicited by a depolarization of the cell membrane and travel along the axon* at 1 to 100 meters per second.

Area 17, A17 cf. primary visual cortex.

Axon The output of a neuron. Typically, a tubular process that may split into branches and terminates at synapses connecting with dendrites of other neurons. The diameter ranges from 0.2 to 20 μm. The velocity for traveling action potentials along the axon is between 1 and 100 meter per second (Kandel et al., 1995). Retinal ganglion cell* axons are translucent and cross over the retina towards the optic disk (blind spot). They constitute the closest functional neuronal elements to an epi-retinal implant and may easily be erroneously stimulated instead of the ganglion cells bodies.

Current source density, CSD Neuronal field potentials originate from inward and outward cell membrane currents. In analogy to electrostatics, the CSD method localizes current sources and sinks by spatial differentiation of neuronal field potential profiles (cf. Mitzdorf, 1985).

Excitatory postsynaptic potential, EPSP In chemical synapses, a presynaptic neuron liberates a transmitter substance into the synaptic gap which causes an inward current into the postsynaptic cell if the synapse is excitatory. The current causes an EPSP driving the postsynaptic neuron towards threshold.

Ganglion cells Cells of a semi-transparent cell layer of the retina. The light travels through this inner layer and other layers of the retina and is absorbed by the black pigment of the photoreceptors. The evoked photoreceptor potentials are preprocessed by horizontal, bipolar, and amacrine cells before the signal is received by the ganglion cells. Ganglion cells have concentric receptive fields* (RF) with antagonistic center-surround properties. On-ganglion cells respond best to a spot o flight focused on the RF center, whereas off-ganglion cells prefer a dark spot with bright surround. The ganglion cell output is formed by the optic nerve and terminates in the LGN. X type ganglion cells have small RFs and produce a sustained response. Their axons* have a low conduction velocity for action potentials*. In contrast, Y cells have larger RFs and respond transiently to rapid stimulus motion. When the ganglion cells are electrically instead of visually stimulated, the preprocessing performed in the more distal retinal layers is skipped.

Independent component analysis, ICA May be thought of being an extension of PCA*, which does not only yield decorrelated coefficients to represent the analyzed data, but statistically independent (or as independent as possible). Second order statistics such

as variance or correlation, are not sufficient to provide an independent representation. Higher order statistics have to be employed (cf. Mendel, 1991).

Inhibitory postsynaptic potential, IPSP In contrast to the excitatory postsynaptic potential*, the transmitter substance causes an outward current in the postsynaptic cell, which leads to its hyperpolarization.

Lateral geniculate nucleus, LGN Subcortical structure receiving direct input from the respective contralateral half of both retinae via the optic nerves. The LGN comprises several layers that distinctly process information from retinal X cells* and Y cells. LGN relay cells pass on the incoming activity to the primary visual cortex* (V1).

Local field potentials, LFP Signals resulting from bandpass filtering of extracellularly recorded signal. Captures mainly slow potentials, such as EPSP or IPSP.

Multi unit activity, MUA Spiking activity of a population of neurons. Technically, the extracellularly recorded signal is highpass filtered (0.5–10 kHz) and subsequently rectified and lowpass filtered. Thus, one yields the envelope of spiking activity, an estimate for the population spike density.

Primary visual cortex, V1 Defined as that part of the cortex that receives direct input from the dorsal LGN*. In primates, primary visual cortex, V1, and area 17 are the same. In cat, area 18 and 19 also get input from LGN and thus, functionally belong to the primary visual cortex. The terms area 17*, 18, 19, etc. refer to a physiological classification of cortical areas (cf. Orban, 1984).

Principal component analysis, PCA Achieves a linear representation of the data with a minimum set of coefficients by implicitly using a least squares criterion. The data-specific basic vectors (principal components) result from the orthogonalization of the covariance matrix of the data. Both, the principal components and the principal component coefficients (representation of the data) are orthogonal (cf. Glaser and Ruchkin, 1976).

Receptive field, RF Defined as the restricted region of the sensory surface (e.g. retina or skin) that influences the activity of a neuron. In analyses of the visual system the RF is usually projected onto a tangent screen.

Spikes cf. action potentials.

Single unit activity, SUA Refers to action potentials of single neurons, cf. action potentials.

V1 cf. primary visual cortex.

X cells cf. ganglion cells.

Y cells cf. ganglion cells.

Index