spectro-temporal characterization of auditory neurons



SPECTRO-TEMPORAL CHARACTERIZATION OF AUDITORY NEURONS

Promotor : Prof.Dr. A.J.H. Vendrik Co-referent : Dr. P.I.M. Johannesma

SPECTRO-TEMPORAL CHARACTERIZATION OF AUDITORY NEURONS

PROEFSCHRIFT TER VERKRIJGING VAN DE GRAAD VAN DOCTOR INDEWISKUNDEENNATUURWETENSCHAPPEN AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. P.G.A.B. WIJDEVELD VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN IN HET OPENBAAR TE VERDEDIGEN OP WOENSDAG 17 JUNI 1981 DES NAMIDDAGS TE 2.00 UUR PRECIES

door

ADRIANUS MARTINUS HUBERTUS JOSEPH AERTSEN geboren te Oosterhout

> mei 1981 Druk: Stichting Studentenpers Nijmegen

Graag wil ik hier iedereen bedanken die heeft bijgedragen aan hetgeen in dit proefschrift wordt gepresenteerd. Dit geldt in het bijzonder voor de leden van de werkgroep Neurofysika. Voor de prettige sfeer zoals ik die heb ondergaan op het laboratorium voor Medische Fysika en Biofysika zeg ik zijn bevolking dank. De ondersteuning ondervonden vanuit de elektronikagroep en rekenmachinegroep is bij een onderzoek als dit van essentieel belang. De afdeling medische fotografie heeft in belangrijke mate bijgedragen aan het presentabel maken van de resultaten. Tenslotte een speciaal woord van dank aan Marianne de Leng, die enige honderden malen de term 'spectro-temporal' alsmede het overgrote deel van dit proefschrift op feilloze wijze uit de typemachine wist te krijgen.

'But our everyday experiences - especially our auditory sensations - insist on a description in terms of *both* time and frequency'

> Dennis Gabor 'Theory of Communication' (1946)

CONTENTS

- 1. Introduction.
- Neural representation of the Acoustic Biotope: On the existence of stimulus-event relations for sensory neurons.
- 3. Neural representation of the Acoustic Biotope: A comparison of the response of auditory neurons to tonal and natural stimuli in the cat.
- Spectro-temporal receptive fields of auditory neurons in the grassfrog

 Characterization of tonal and natural stimuli.
- Spectro-temporal receptive fields of auditory neurons in the grassfrog II. Analysis of the stimulus-event relation for tonal stimuli.
- Spectro-temporal receptive fields of auditory neurons in the grassfrog III. Analysis of the stimulus-event relation for natural stimuli.
- The spectro-temporal receptive field: a functional characteristic of auditory neurons.
- 8. A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli.
- 9. A. Summary
 - B. Samenvatting

CHAPTER I

Introduction

The study of information processing in the sensory nervous system may be viewed as an investigation of images. Throughout the auditory nervous system, from the hair cells in the cochlea and the auditory nerve fibers, through the various stages of the auditory processor, composed of the numerous individual, neurons with their different patterns of interconnections, we have what might be called 'the neural image of sound' in its different realizations. The external world is paralleled by an internal representation (e.g. Craik, 1943; McCulloch, 1965). The acoustic environment of an animal, consisting of patterns of air pressure variations at the external ears is represented and transformed internally by a hierarchical network of neurons which communicate by complex spatio-temporal patterns of electrical activity, composed of sequences of action potentials, the all-or-none events, generated by the individual neurons. A central paradigm in the study of the sensory nervous system is that meaningful information regarding its operation principles can be gained from the experimental investigation of the functional characteristics of its elementary components i.e. the single neurons. This is reflected in the vast amount of literature based on single unit experiments (e.g. Kiang et al., 1965; Keidel and Neff, 1975). Only recently we are witnessing an increasing interest in the experimental investigation of assemblies of neurons (e.g. Dickson and Gerstein, 1974), where the availability of modern techniques has enhanced the experimental feasibility.

The characteristics of the neural image of sound at a particular level of the auditory system reflect the status of processing, performed up to that level. In the periphery the single neurons can quite adequately be described by characteristics as frequency tuning and temporal pattern of firing to stimulation with tones. This type of characteristics, grossly speaking, serves to cover all relevant aspects of the various possible sounds occurring in the external world (Evans, 1974; Van Gisbergen, 1974). Moreover the peripheral neural image can be characterized by simple organizational principles like tonotopy i.e. a spatially orderly representation of the spectral composition of the acoustic environment. The picture becomes increasingly complex and shows more and more blanks when moving towards the more central parts of the auditory system. Gradually the image starts to change and the classical experiments using 'simple' stimuli as tones and clicks loose their claim for completeness. Especially in the more central regions of the auditory system the question emerged whether 'the full capacity of the auditory processor may not be revealed except through study of its responses to stimuli that pose analytical tasks of the kind that shaped its evolutionary development' (Worden and Galambos, 1972). At

this stage ethology, the study of animal behaviour, becomes an important source of inspiration in brain research (e.g. Bullock, 1977; Ewert, 1980). Concepts on the neural base of perception gradually move towards the idea of 'action oriented perception': 'perception is oriented towards the future as much as the present, "exploring" features of the current environment which may be incorporated into an "internal model of the world" to guide future action more and more adaptively' (Arbib, 1972). The contours of a possible context of the relevant questions are emerging, a mosaic composed fragments of answers is beginning to develop.

The single neuron, at various levels of the auditory nervous system, is the subject of extensive experimental investigation. A rich repertoire of sounds is presented to the animal under investigation and the electrical activity from one particular neuron is recorded simultaneously. Various techniques have been developed to investigate the nature of the relation between acoustic stimulus and neural activity. The purpose of these techniques is to determine to what aspects of sound the neuron responds, if it responds at all which is not always a trivial question. An important question in this respect is whether the properties that are determined for a particular neuron are context-free, i.e. whether they provide an adequate neuron characteristic over a variety of different acoustic stimuli: the quest for stimulus-invariant neuron characteristics. Stimulus-invariance of a neuron characteristic is no logical a priori, it needs to be established experimentally.

A functional description of the nature of an auditory neuron's stimulus-response relation evidently should be based on a thorough description of sound as such and should furthermore rely on formal methods of quantifying the relation between input and output variables of an unknown 'black box'.

The notions of 'frequency' and 'time' have a long history in the representation of sound. This is amply illustrated by the widespread convention in musical notation, dating back as far as the 11-th century (Hindley, 1978). The mathematical theory which classically deals with these concepts, the Fourier theory, treats the description of a signal in the time domain and its description in the frequency domain as mutually exclusive representations. The basic reason for this is that in Fourier theory phenomena are considered over an infinite interval. This exclusivity, however, does not appear to fit with our perceptual experience. In 1946 Gabor, in his 'Theory of Communication' writes: 'For instance, speech and music have for us a definite "time pattern" as well as a frequency pattern. It is possible to leave the time pattern unchanged, and double what we generally call "frequencies" by playing a musical piece on the piano an octave higher, or conversely it can be played in the same key, but in different time. Evidently both views have their limitations, and they are complementary rather than mutually exclusive'. Based on this idea he developed mathematical methods of signal representation, with the purpose of rejoining the hitherto separate representations into a compound spectro-temporal image of sound, into what he calls a 'diagram of information' composed of elementary quanta of information, the 'logons'. His description, which bears a formal relation with Heisenberg's uncertainty relation in quantum mechanics, has greatly influenced certain developments in the theory of information, both in artefacts and in the nervous system (e.g. MacKay, 1969).

The study of the relation between quantities which can be described as input and output, or stimulus and response, of an unknown 'black box' forms the object of mathematical system theory. The characteristics of many biological systems are such that they require a theory which is capable of handling nonlinear systems. Based upon the work of Volterra on the integral representation of nonlinear systems with known characteristics, e.g. in the form of a differential equation, Wiener developed an approach for the investigation of nonlinear systems with unknown characteristics (Wiener, 1958; see also Barrett, 1963) This approach is inherently a stochastic one, relying on the use of Gaussian white noise as the test input ensemble. The experimental feasibility of this approach was greatly enhanced by the introduction of statistical analysis techniques, the cross correlation functions, as a tool to investigate the system's input-output relation (Lee and Schetzen, 1965; Marmarelis and Marmarelis, 1978). Inspired by this approach to system's analysis De Boer and Kuyper (1968) introduced into sensory neurophysiology the use of Gaussian noise as a stimulus ensemble and developed the 'reverse correlation' procedure to establish the formal relation between sensory stimulus and neural events. This technique appeared to be quite fruitful when applied to peripheral parts of the auditory system (e.g. Møller, 1973, De Boer and De Jongh, 1978). The 'revcor'-procedure as an analysis technique has been reformulated and generalized into the conceptual framework of the Pre-Event Stimulus Ensemble (PESE) as the ensemble of stimulus elements preceding the neural events (Johannesma, 1972; Grashuis, 1974). The statistical analysis of the properties of the PESE, which forms the neuron's selected subset of the complete stimulus ensemble, can be regarded as the formalized version of the search for those aspects in sound which affect the neuron's firing probability.

The collection of papers, presented in this thesis, forms the reflection of an attempt to reconcile the foregoing considerations into a functional analysis of the single neuron's role in the neural image of sound.

The basic questions at the single neuron level are described in Chs. 2 and 3. In the Chs. 4 to 6 the procedure of spectro-temporal analysis of the stimulus-event relation is developed and possible ways of implementation are described. Ch. 7 presents a discussion of the spectrotemporal analysis of single neurons in the context of general nonlinear system theory and, finally, Ch. 8 discusses the limitations of a particular type of realization of this approach Throughout the various papers the concepts are illustrated by results of their application to extra-cellular single unit recordings from different levels in the auditory system of the anaesthetized cat and frog. A neuron characteristic which gradually emerges as a central concept in the present approach is the 'spectrotemporal receptive field' (STRF): the particular patch in the spectro-temporal stimulus domain which characterizes the neuron's sensitivity to sound. The STRF may tentatively be regarded as the neural analogon of Gabor's elementary logon in the 'information diagram' of sound. The perspective of a neural image of the external acoustic world, composed as a patchwork of neural logons seems worthwile to be explored (Altes, 1978; Johannesma, 1981). In this investigation the contribution of interactions in the neural assembly can hardly be ignored.

'.....Denken wir uns eine weisze Fläche, auf der unregelmäszige schwarze Flecken wären. Wir sagen nun: Was für ein Bild immer hierdurch entsteht, immer kann ich seiner Beschreibung beliebig nahe kommen, indem ich die Fläche mit einem entsprechend feinen quadratischen Netzwerk bedecke und nun von jenem Quadrat sage, dasz es weisz order schwarz ist. Ich werde auf diese Weise die Beschreibung der Flache auf eine einheitliche Form gebracht haben. Diese Form 1st beliebig, denn ich hatte mit dem gleichen Erfolge ein Netz aus dreieckigen oder sechseckigen Maschen verwenden konnen. Es kann sein, dasz die Beschreibung mit Hilfe eines Dreiecks-Netzes einfacher geworden ware; das heiszt, dasz wir die Flache mit einem groberen Dreiecks-Netz genauer beschreiben konnten als mit einem feineren guadratischen (oder umgekehrt) usw. Den verschiedenen Netzen entsprechen verschiedene Systeme der Weltbeschreibung..... Dasz sich ein Bild, wie das vorhin erwähnte, durch ein Netz von gegebener Form beschreiben laszt, sagt uber das Bild nichts aus. (Denn dies gilt fur jedes Bild dieser Art.) Das aber charakterisiert das Bild, dasz es sich durch ein bestimmtes Netz von bestimmter Feinheit vollstandig beschreiben läszt' (Wittgenstein, 'Tractatus logicophilosophicus', 6.341, 6.342 (1921)).

- Arbib, M.A.: The metaphorical brain. An introduction to cybernetics as artificial intelligence and brain theory. New York, London, Sydney, Toronto: Wiley Interscience 1972
- Altes, R.A.: Possible reconstruction of auditory signals by the central nervous system (abstract). J. Acoust. Soc. Am. 64 (Suppl. I), S 137 (1978)
- Barrett, J.F.: The use of functionals in the analysis of non-linear physical systems. J. Electronics and Control 15, 567-615 (1963)
- Boer, E. de, Kuyper, P.: Triggered correlation. IEEE Tr. on Biomed. Eng. BME-15, 169-179 (1968)
- Boer, E. de, Jongh, H.R. de: On cochlear encoding: potentialities and limitations of the reverse-correlation technique. J. Acoust. Soc. Am. 63, 115-135 (1978)
- Bullock, T.H. (ed.): Recognition of complex acoustic signals. Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Craik, K.: The nature of explanation. Cambridge: Cambridge University Press 1943
- Dickson, J.W., Gerstein, G.L.: Interactions between neurons in auditory cortex of the cat. J. Neurophysiol. 37, 1239-1261 (1974)
- Evans, E.F.: Neural processes for the detection of acoustic patterns and for sound localization. In: The neurosciences, third study program, pp. 131-145. Schmidt, F.O., Worden, F.G. (eds.). Cambridge, MA: M.I.T. Press 1974
- Ewert, J.-P.: Neuroethology. An introduction to the neurophysiological fundamentals of behavior. Berlin, Heidelberg, New York: Springer 1980
- Gabor, D.: Theory of communication. J.I.E.E. (London), pt. III, 93, 429-457 (1946)
- Gisbergen, J.A.M. van: Characterization of responses to tone and noise stimuli of neurons in the cat's cochlear nuclei. Ph.D. Thesis, Nijmegen, the Netherlands 1974
- Grashuis, J.L.: The pre-event stimulus ensemble: an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons. Ph.D. Thesis, Nijmegen, the Netherlands 1974

- Johannesma, P.I.M.: The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: Proc. of the IPO Symp. on Hearing Theory, pp. 58-69. Cardozo, B.L. (ed.). Eindhoven, the Netherlands 1972
- Johannesma, P.I.M.: Neural representation of sensory stimuli and sensory interpretation of neural activity. In: Proc. Symp. 'Neural communication and control; facts and theories'. Debrecen (Hungary), in press 1981
- Keidel, W.D., Neff, W.D. (eds.): Handbook of sensory physiology. Vol. V/2, Auditory System: Physiology (CNS), Behavioral Studies, Psychoacoustics. Berlin, Heidelberg, New York: Springer 1975
- Kiang, N.Y.-S., Watanabe, T., Thomas, E.C., Clark, L.F.: Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, MA: M.I.T. Press 1965
- Lee, Y.W., Schetzen, M.: Measurement of the Wiener kernels of a nonlinear system by cross-correlation. Int. J. Control 2, 237-254 (1965)
- Lindley, G. (ed.): The Larousse encyclopedia of music. Based on: 'La musique: les hommes; les instruments; les oeuvres', Dufourq, N. (ed.). London, New York, Sydney, Toronto: Hamlyn (1978)
- MacKay, D.M.: Information, mechanism and meaning. Cambridge, MA: M.I.T. Press (1969)
- Marmarelis, P.Z., Marmarelis, V.Z.: Analysis of physiological systems. The white-noise approach. New York, London: Plenum Press 1978
- McCulloch, W.S.: Embodiments of mind. Cambridge, MA: M.I.T. Press 1965
- Møller, A.R.: Statistical evaluation of the dynamic properties of cochlear nucleus units using stimuli modulated with pseudo random noise. Brain Res. 57, 443-456 (1973)
- Wiener, N.: Nonlinear problems in random theory. Cambridge, MA: M.I.T. Press; New York: Wiley 1958
- Wittgenstein, L.: Tractatus logico-philosophicus. First published in: Annalen der Naturphilosophie (1921). Quotation from 9-th edition by Suhrkamp Verlag, Frankfurt 1973
- Worden, F.G., Galambos, R. (eds.): Auditory processing of biologically significant sounds. Neurosci. Res. Prog. Bull., Vol. 10. Brookline, Mass.: Neurosciences Research Program 1972

CHAPTER 2

Neural Representation of the Acoustic Biotope

On the Existence of Stimulus-Event Relations for Sensory Neurons

```
Aertsen, A.M.H.J., Smolders, J.W.T. and
Johannesma, P.I.M.
Biol. Cybernetics <u>32</u>, 175-185 (1979)
```



Neural Representation of the Acoustic Biotope: On the Existence of Stimulus-Event Relations for Sensory Neurons

A. M. H. J. Aertsen, J. W. T. Smolders*, and P. I. M. Johannesma Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Abstract. In the context of a behaviourally oriented investigation of neural information processing we have studied single unit responses from cochlear nucleus (12 units) and auditory cortex (23 units) of the cat. Attention was focused on two aspects, the composition of a wide stimulus ensemble and the influence of this ensemble on the single unit neural activity: action potentials or neural events. The stimulus ensemble was based on sounds occurring in the natural environment of the animal. the Acoustic Biotope A realization of such an ensemble was constructed and used in the experiments. A systematic analysis of the effect of this stimulus ensemble on the neural activity comprises different questions with respect to a possible stimulus-event relation · existence and nature. In the present paper emphasis is given to the question of existence of a s-e relation. The existence of such a relation has been investigated by presenting the stimulus ensemble twice. The degree of reproducibility of neural activity indicates to what extent a s-e relation indeed exists. A formal measure of testing the reproducibility of two sequences of neural events is defined: the Cross Coincidence Function. Results are presented, interpretation and formal problems discussed Results obtained so far indicate that on the whole the existence of a s - e relation can be established more definitely for cochlear nucleus units than for auditory cortex units Applications of the presented procedure arise in exploratory investigations in central regions of the sensory nervous system and in regions with presumed multimodal inputs.

1. Introduction

The study of sensory neural systems has in recent years attracted considerable attention from work inspired by

nonlinear system theory as outlined by Volterra and Wiener This theory, applicable to a wide class of nonlinear systems, states that such a system can be characterized by a set of different order kernels. The determination of these Wiener kernels forms a major problem in the Wiener theory A relatively simple method of measuring these kernels was presented by Lee and Schetzen (1965).

This method, based on crosscorrelation techniques, is an extension of the crosscorrelation method as applied in linear system theory (Lee, 1960) In this context Gaussian wideband noise plays a central rôle in the composition of an adequate stimulus ensemble

Although differing in actual terminology and methods, various studies on the visual system (Marmarelis and Naka, 1973, Marmarelis and McCann, 1973, Von Seelen and Hoffman, 1976) and the auditory system (De Boer and Kuyper, 1968; Johannesma, 1972; Grashuis, 1974, Van Gisbergen et al, 1975b, De Boer and De Jongh, 1978) all trace back to this same methodological approach

In this same period the question emerged whether the full capacity of the sensory processor "may not be revealed except through study of its response to stimuli that pose analytical tasks of the kind that shaped its evolutionary development" (Worden and Galambos, 1972) This suggests that the criteria for an adequate composition of the stimulus ensemble may well be deduced from the study of animal behaviour in the natural environment i.e. ethology and ecology

In the study of the auditory nervous system this leads to a stimulus ensemble based on natural sounds. Recent reviews of this line in auditory research are given by Worden and Galambos (1972) and Bullock (1977). Results obtained so far indicate that responses of peripheral neurons (up to the ventral cochlear nucleus) to complex, natural stimuli can be relatively simply predicted from responses to technical stimuli (Evans, 1975) This clearly does not hold for neurons in

^{*} Present address Zentrum der Physiologie der J W Goethe-Universitat, D-6000 Frankfurt, FRG

more central regions like the auditory cortex (Goldstein and Abeles, 1975, Newman, 1977) This suggests that especially for central auditory regions the use of an ethologically based stimulus ensemble is indispensable in the investigation of neural functioning (Scheich, 1977, Newman, 1977)

The present work forms part of an investigation of the auditory nervous system of the cat Extracellular single unit recordings are made from both a relatively peripheral and a central part in the auditory system. the cochlear nucleus and the auditory cortex. The stimulus ensemble consists of a wide selection of natural sounds together with technical sounds. The analysis of stimulus-response relations is, in its methodological aspects, inspired by nonlinear system theory as indicated before The Pre Event Stimulus Ensemble (PESE) is defined as the set of stimuli preceding a neural event: the occurrence of an action potential. As a consequence the PESE forms a subset of the Stimulus Ensemble (SE). The basic idea is that the characteristics of the PESE, as compared to those of the SE provide a functional description of the neuron's rôle in the representation of acoustic stimuli (Johannesma, 1972; Grashuis, 1974).

The analysis of a stimulus-response or stimulusevent relation logically starts with the question whether such a relation exists for the particular neuron from which a recording is made. In a number of cases the existence of a s-e relation is fairly easily demonstrated by a significant change in the neuronal firing pattern, related to the stimulus presentation. However, in case a wide, ethologically based stimulus ensemble is used it often proves difficult or even impossible, just employing simple means (visual or aural inspection of the data), to demonstrate this change unambiguously. Several reasons may account for this problem, which arises in the case of cochlear nucleus units as well as auditory cortex units, be it that it occurs more frequently in the latter case. It may be due to the complex structure of the stimulus ensemble. which holds for both peripheral and central neurons. Furthermore the neural functional properties are involved These are known to be more complex in central regions than in the periphery (Newman, 1977)

Thus the question of the existence of a s - e relation proves to be a problem, especially when neural activity is recorded from central brain areas which are involved in a complex mode of sensory information processing As already pointed out, particularly these regions appear to be the natural candidates for an investigation with an ethologically inspired stimulus ensemble. Therefore the question of existence of a stimulus-event relation deserves some special attention The elaboration of this question forms the principal objective of this paper A behaviourally oriented stimulus ensemble comprises many different stimuli and as a consequence has a long duration. The stochastic nature of the neural activity suggests many repetitions of the same experiment. On the other hand the recording time from one neuron is limited ($\sim 10-100$ min). This constraint implies an optimization to be made between both diversity and precision of information to be gained. In this paper the attention is concentrated on the analysis of neural function using a wide ensemble of ethological stimuli, where the number of presentations is limited to the minimum of two

2. The Acoustic Biotope: An Ethologically Based Stimulus Ensemble

The natural environment in which an animal lives, as far as functionally related to the animal, is defined as its habitat or biotope (Dahl, 1921). The biotope encompasses both the biotic and abiotic environment (Odum, 1971) Important characteristics of the biotope are its geophysical properties, climate regime and variety and numbers of occurring animals and plants. The Sensory Biotope may be defined as the total of sensory aspects of the biotope. It consists of different components, each one corresponding to a specific sensory modality of the animal under investigation (visual, auditory, tactile etc.) Since our work is concerned with the auditory nervous system the attention is focused on the acoustical component of the sensory biotope the Acoustic Biotope (Smolders et al., in preparation). A more formal definition of the Acoustic Biotope as a probabilistic generator of stimulus ensembles as particular realizations, together with possible ways to characterize the Acoustic Biotope will be presented elsewhere (Aertsen et al, in preparation).

For the present investigation a realization of a wide stimulus ensemble, inspired by ethological and ecological observations was made.

This realization was not restricted to (species specific) vocalizations nor was it designed in view of remarkable behavioural functions and/or neural structures The stimulus ensemble was composed of a wide selection of sounds recorded from the natural environment of the cat Furthermore it included technical sounds, in order to be able to relate results obtained with this stimulus ensemble to the outcome of previous investigations (Grashuis, 1974, Van Gisbergen et al., 1975a-c) The realization was composed according to the scheme as shown in Diagram 1 (Smolders et al., in preparation)

The stimulus ensemble consisted of 64 different sound segments, each one with a duration of 8–9s The figures in Diagram 1 specify the number of sound segments taken from different categories. This classifi-



Diagram 1. Construction scheme for realization of a stimulus ensemble inspired by the Acoustic Biotope of the cat (From Smolders et al, in preparation)

cation of sounds was based on biological or technical characteristics, depending on the nature (biotopic or synthetic) of these sounds The 64 sound segments were concatenated in a random order on analog tape, thus constituting a stimulus ensemble lasting about 9 min, its bandwidth being 20 kHz In the following this realization will be denoted briefly as "stimulus ensemble" or "acoustic biotope" It should be born in mind, however, that it is a particular realization of the Acoustic Biotope, with additional technical sounds included in it

3. Methods

Preparation, Recording, and Histological Procedures

Nine adult cats weighing 18-25 kg were used. Extracellular single unit recordings were made in the cochlear nucleus and the auditory cortex The cochlear nucleus results were obtained from cats under nembutal anaesthesia The methods used for the preparation of the animal, recording of single unit activity and histological control of the anatomical location of neurons were the same as have been described in detail by Van Gisbergen et al (1975a) Auditory cortex experiments were performed on semichronic preparations. Only a brief outline of the preparation is given here

A few weeks before the actual recordings began the cat was anaesthetized and a hole was drilled into the skull under stereotaxic control A ring-shaped chamber was screwed and cemented onto the skull around the exposed brain area The dura remained intact The chamber could be closed A small holder was screwed onto the skull overlying the frontal sinus, it served for later pain-free fixation of the head to the stereotaxic apparatus in the actual experiment The preparation was performed aseptically The animal was then allowed to recover during a period of at least one week In the actual experiment the cat was premedicated with atropine and immobilized (Flaxedil was administered intramuscularly in an initial dose of 5 mg/kg and subsequent injections of 25 mg/kg at the first sign of muscular contractions) Xylocain ointment was applied to the eyes to prevent them from drying

Artificial respiration was performed using a 30% O₂-70% N₂O gas mixture to obtain an end tidal CO₂-percentage of about 38 Body temperature was kept at 37-38 °C

By means of a remotely controlled microstepdriver tungsten microelectrodes (typical impedance $3 M\Omega$ at 1000 Hz) were driven into the brain

In order to improve recording stability the chamber on the skull was filled with an agar solution

Stimulation and Data Collection

Experiments were performed with the cat in an electrically shielded acoustic chamber (IAC type 1202 A) The main part of the stimulus ensemble consisted of natural sounds, recorded on analog tape (recorder Uher 4200, tape speed 19 cm/s, microphones



Fig. 1. Dot display representing neural activity $z_1(t)$ recorded from unit 98-2 (auditory cortex) during presentation of the acoustic biotope

Sennheiser MD 211 N). Technical sounds were generated by standard electronic equipment.

A stimulus tape, composed of these sounds, was played back to the animal from an instrumentation recorder (Honeywell 7600, FM, tape speed 60 inch/s, bandwidth 20 kHz). The stimulus was amplified (Luxman 507 X) and transduced by two microphones (Sennheiser MD 211 N) mounted near the cat's pinnae (open sound system). Sound pressure levels were measured using a half inch condenser microphone (Bruel and Kjaer 4134) mounted near the cat's pinnae. Average intensities of the stimuli near the cat's ears were in the order of 50–70 dB SPL with considerable loss at low frequencies (up to some 20 dB loss at 500 Hz) due to the sound system being open.

Neural activity was amplified (Grass P16) and recorded on the stimulus delivering instrumentation recorder. Furthermore it was transformed into unitary pulses by means of a level discriminator and fed, together with a synchronizing clock signal from the stimulus recorder, to the laboratory's computer system (time resolution $100 \,\mu$ s).

The calculations, involved in the analysis as will be described, were carried out on the following configuration: central processor PDP 11/45, additional floating point processor, operating system RSX 11-D, memory 160 Kbytes core, disks RK 05, software Fortran IV +. The analysis of two neural recordings both containing some 2000 events (example discussed in Sect. 4) takes about 100 s.

4. Existence of Stimulus-Event Relations

Because of the all-or-none character of action potentials, single unit activity as measured during presentation of the stimulus ensemble can be modelled as a stochastic point process:

$$z_{1}(t) = \sum_{i=1}^{N_{1}} \delta(t - t_{i}) \quad (0 \le t \le T)$$
(1)

with $\delta(t)$ = Dirac delta function, $\{t_i\}$ = moments of occurrence of action potentials, N_1 = total number of action potentials, measured during the experiment (duration T).

A graphical representation of $z_1(t)$ for an auditory cortex unit is shown in the dot display in Fig. 1. Each dot represents a neural event. The dot display is composed analoguously to a page of text. Each line corresponds to a time interval of 8 s.

The acoustic biotope forms a long and complex stimulus ensemble. The neuron under investigation produces at the same time a long and usually complex sequence of neural events. In this situation the characteristics of the stimulus-event relation, portraying the influence of the sensory stimulus on the neural activity, cannot be found from direct visual inspection of the experimental data as shown in the dot display. A systematic analysis of the stimulus-event relation comprises different questions:

1. Existence: to what extent does a relation exist between the acoustic stimulus and simultaneously occurring action potentials?

2. Nature: if such a relation exists, how can it be characterized? Which stimulus properties determine the occurrence of action potentials?

As already mentioned, the nature of the stimulusevent relation is reflected in the characteristics of the Pre Event Stimulus Ensemble as compared to those of the Stimulus Ensemble. The analysis of these characteristics may have to take into account both first and second order aspects. Results from this analysis for an ethologically based stimulus ensemble will be published (Aertsen et al., in preparation).

The question regarding the existence of a stimulusevent relation for a particular neuron, when stated in more neurophysiological terms, turns into the question whether the present stimulus ensemble influences the neural activity as recorded from that neuron in an appreciable way. Basic to the notion of existence of a s-e relation is the requirement of *reproducibility*: for a s-e relation to exist in the context of a wide stimulus ensemble as discussed before, it is necessary and sufficient that presentation of this stimulus ensemble has a reproducible effect on the simultaneously recorded sequence of neural events.

This statement, in fact, defines the notion of "auditory neuron" in operational terms: each neuron for which, upon stimulation with a wide acoustic stimulus ensemble, a s-e relation is shown to exist is considered to be an "auditory neuron". This interpretation will receive further consideration in the discussion. In view of the stochastic nature of neural activity the use of the term "reproducibility" suggests repetition of experiments: the stimulus ensemble is presented at least twice and the resulting sequences of neural events have to be tested for similarity. Since the stimulus ensemble has a long duration and the recording time is limited, the stimulus ensemble is presented only twice.

Quite analoguously to (1) the neural activity $z_2(t)$ recorded during second presentation of the stimulus ensemble is given by

$$z_2(t) = \sum_{j=1}^{N_2} \delta(t - t'_j).$$
⁽²⁾

Figure 2 shows $z_2(t)$ for unit 98–2 (i.e. the same unit as shown in Fig. 1).

Due to the complex structure of the sequences of action potentials it appears not to be possible in general to decide on the similarity by mere visual inspection of the dot displays. The notion of similarity needs a more formal elaboration.

The (non-normalized) crosscorrelation function $R_{z_1z_2}(t)$ of $z_1(t)$ and $z_2(t)$ is defined as

$$R_{z_1 z_2}(\tau) = \frac{1}{T} \int_0^T dt \, z_1(t) z_2(t+\tau)$$
(3)

with T=duration of the experiment (i.e. of the stimulus ensemble). In order to avoid difficulties at the boundaries (t=0 and t=T) both $z_1(t)$ and $z_2(t)$ are continuated in a cyclic way.

Substitution of the expressions (1) and (2) gives

$$R_{z_1 z_2}(\tau) = \frac{1}{T} \int_0^T dt \sum_{i=1}^{N_1} \delta(t-t_i) \sum_{j=1}^{N_2} \delta(t-t'_j+\tau)$$

Using the convolution of two impulses (Papoulis, 1962)

$$\delta(t-\sigma) * \delta(t-v) = \delta[t-(\sigma+v)]$$
(4)

we obtain

$$R_{z_1 z_2}(\tau) = \frac{1}{T} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \delta[\tau - (t'_j - t_i)].$$
 (5)

It is clear that some type of smoothing has to be applied to $R_{z_1z_2}(\tau)$ in order to obtain a practically useful estimate.

To this end we define the Cross Coincidence Function

$$C_{z_1 z_2}(\tau, \Delta) = \frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} d\sigma R_{z_1 z_2}(\sigma) \qquad (\Delta > 0).$$
(6)

 $C_{z_1z_2}(\tau, \Delta)$ counts the number of coinciding events from the spike sequence $z_1(t)$ and the set of time-shifted sequences $\{z_2(t+\sigma)\}$ with $\tau < \sigma \le \tau + \Delta$.



Fig. 2. Dot display representing neural activity $z_2(t)$ recorded from unit 98-2 during second presentation of the acoustic biotope This dot display should be compared to Fig 1

It is easily verified that $C_{z_1z_2}(\tau, \Delta)$ formally equals $R_{z_1Z_2}(\tau)$ with the smoothed sequence $Z_2(t) = (z_2*h)(t)$ and the weight function

$$h(t) = \frac{1}{\Delta} (0 < t \le \Delta) \qquad 0 \text{ (otherwise)}.$$
(7)

Furthermore we note that (for $\tau \neq 0$) it holds that

$$\lim_{\Delta \downarrow 0} C_{z_1 z_2}(\tau, \Delta) = \frac{N_1}{T} I_{z_1 z_2}(\tau)$$
(8)

with $I_{z_1z_2}(\tau)$ the Cross Intensity Function (Cox and Lewis, 1966), for $\tau \neq 0$ defined as

$$I_{z_1 z_2}(\tau) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \operatorname{prob} \left(\begin{array}{c} \operatorname{event} \text{ in sequence } z_2 \\ \operatorname{between} t + \tau, t + \tau + \Delta \end{array} \middle| \begin{array}{c} \operatorname{event} \text{ in} \\ \operatorname{sequence} z_1 \text{ at } t \end{array} \right).$$
(9)

Inspired by these results we propose an approach to the question of reproducibility, based on the cross coincidence function $C_{z_1z_2}(\tau, \Delta)$ as an estimate of the cross intensity function $I_{z_1z_2}(\tau)$.

For reasons of symmetry we prefer to use a modified cross coincidence function $C_{z_1z_2}(\tau, \Delta)$, defined as

$$\tilde{C}_{z_1 z_2}(\tau, \Delta) = C_{z_1 z_2}(\tau - \Delta/2, \Delta)$$
(10)

for which it holds that

$$\tilde{C}_{z_1 z_2}(\tau, \Delta) = \tilde{C}_{z_2 z_1}(-\tau, \Delta).$$
(11)

This implies that the (modified) autocoincidence function is an even function of τ .

The behaviour of $\tilde{C}_{z_1z_2}(\tau, \Delta)$ in the region around $\tau = 0$, as compared to its behaviour for larger values of $|\tau|$, indicates to what extent the event sequences $z_1(t)$ and $z_2(t)$ are correlated (synchronized).

A detailed correspondence in the instants of events in both sequences, as revealed by a relatively sharp



Fig. 3. Cross Coincidence Histogram of neural event sequences $z_1(t)$ and $z_2(t)$, recorded from unit 98-2 during subsequent presentations of the acoustic biotope

peak in the cross coincidence function around $\tau = 0$ is attributed to the effect on neural activity exercised by effective (i.e. acoustic) stimulation. In that case the existence of a stimulus-event relation has been established.

The actual analysis of experimental data is performed by means of the Cross Coincidence Histogram $\{C_n(\Delta)\}$, with

$$C_n(\Delta) = \tilde{C}_{z_1 z_2}(n\Delta, \Delta) \qquad n = 0, \pm 1, \pm 2, \dots$$
 (12)

as a sampled representation of the cross coincidence function.

Figure 3 shows the form of the cross coincidence histogram as calculated for the spike sequences recorded from unit 98-2 (Figs. 1 and 2). In this case Δ has been taken 1 ms. Note that the result has not been divided by T or Δ ; ordinate values $N_{1,2}(\tau)$ refer to numbers of coincidences in one bin Δ . Numbers of action potentials in sequences $z_1(t)$ and $z_2(t)$ (N_1, N_2) are given for reference.

The arrow in Fig. 3 indicates the expectation $E(N_{1,2}) = N_1 N_2 \Delta/T$ for independent stationary point processes (Cox and Lewis, 1966).

It is clear from Fig. 3 that $z_1(t)$ and $z_2(t)$ are highly correlated. The distinct peak around $\tau = 0$ (width some 5 ms) against the flattened background for larger values of $|\tau|$ implies a detailed correspondence between both recordings. From this we infer that the firing pattern of unit 98-2 is influenced in a reproducible way by presentation of a wide acoustic stimulus ensemble. For this neuron the existence of a stimulusevent relation has now been established.

5. Results

Experiments were performed on 9 cats, recordings were made from 23 auditory cortex neurons and 12 cochlear nucleus neurons. As in a number of cases some doubt existed about the quality of the recordings, the analysis as described in this paper was applied to 9 auditory cortex units and 9 cochlear nucleus units.

Figures 4 and 5 show the results for four neurons, two of them from the cochlear nucleus (Fig. 4) and two from the auditory cortex (Fig. 5). Note the different time scales in the histograms. As in Fig. 3 the arrow in the histogram indicates the expectation $E(N_{1,2})$ for independent stationary point processes.

Results for all cochlear nucleus units proved to be very similar to the examples shown in Fig. 4: a distinct peak around $\tau = 0$, its width some 5–10 ms. A conclusion regarding the existence of a s - e relation for these neurons is clearly positive.

Results for neurons recorded from the auditory cortex can be divided into two groups. One group (5 neurons) showed results which lead to a positive conclusion regarding the existence of a s-e relation. Compared to cochlear nucleus neurons peaks around $\tau = 0$ tend to be less distinct: width up to some 40 ms (Fig. 5). Unit 98-2 (Fig. 3) forms an exception to this.

For the second group (4 neurons) the cross coincidence histograms did not show a clear structure, which implies that on the basis of these results no conclusion can be drawn regarding their possible s-e relation. For one of these neurons the average firing rate under zero-stimulus conditions appeared to be considerably lower than during presentation of the stimulus ensemble, which suggests that a s-e relation existed nevertheless. For the remaining 3 neurons (from different cats) in the second group no additional evidence permitted to draw a positive conclusion about a possible s-e relation.

For the cochlear nucleus experiments histological analysis revealed that in all cases considered, the neuron indeed was located within the target area. For the auditory cortex experiments histological data were not available.

6. Discussion

In the context of a behaviourally oriented investigation of neural information processing we have studied single unit responses from different auditory regions of the cat: cochlear nucleus and auditory cortex. To this end a wide stimulus ensemble was constructed, based on arguments adopted from ethology and ecology. The Acoustic Biotope is intended to create a natural stimulus environment for the animal; an environment to which the auditory nervous system, in its phylogenetical and ontogenetical development structurally and functionally has adapted.

Biological and physical characterization of sounds from the Acoustic Biotope can reveal which properties of sound signals are behaviourally important. Results



Fig. 4. Dot displays and cross coincidence histograms of neural activity recorded from two cochlear nucleus units during subsequent presentations of the acoustic biotope. Left column: unit 101-1, right column: unit 107-1

from this analysis have to be related to the neural representation of (natural) sounds as deduced from single unit studies, thus providing a possible frame of reference for meaningful interpretation of neuronal functional properties. Moreover such an analysis of the Acoustic Biotope may give clues to the problem of a suitable parametrization of complex natural sounds, ultimately resulting in a stimulus grammar, capable of generating a synthesized acoustic pseudo-biotope. By systematic variation of the stimulus parameters, preferably on an interactive basis, we would be able to study the auditory system in a quasi-natural acoustic environment without having to give up stimulus control. During presentation of a wide, mainly natural, partly technical stimulus ensemble a neuron produces a long and complex sequence of action potentials. Aspects of a possible stimulus-event relation then have to be subjected to systematic analysis: does it exist anyhow and, if so, what is the nature of it. In this paper we have concentrated on the problem of existence, the analysis of its nature will be the subject of a forthcoming paper (Aertsen et al., in preparation).

The question regarding the existence of a s-e relation was reformulated into the question whether presentation of a wide stimulus ensemble has a reproducible effect on the recorded single unit activity. The stimulus ensemble therefore is presented twice and an



Fig. 5. Dot displays and cross coincidence histograms of neural activity recorded from two auditory cortex units during subsequent presentations of the acoustic biotope. Left column: unit 96-4, right column: unit 97-4

analysis of the two resulting sequences of neural events has to reveal whether the hypothesis of independent firings during subsequent presentations of the stimulus ensemble is to be rejected. The rejection of independentness is considered to prove the existence of a s-e relation.

We have used an analysis which is based on the cross correlation function: the cross coincidence function (or -histogram). This approach was inspired by, amongst others, a similar approach used in the evaluation of synaptic interactions between different neurons from which simultaneous recordings were made (Griffith and Horn, 1963; Perkel et al., 1967; Moore et al., 1970; Gerstein, 1970; Bryant et al., 1972; Dickson and Gerstein, 1974). Some important distinctions, however, have to be made. The first one concerns the neuronal data, subjected to the analysis. In the multiunit studies the correlation analysis is applied to stationary spike sequences, recorded under zerostimulus conditions and on sequences, recorded during repetitive presentation of a stimulus, the stimulus necessarily being short because the generation of post stimulus time histograms is involved. The present paper describes the application of correlation methods to two, in general nonstationary, spike sequences recorded under identical complex stimulus conditions. A second distinction which is to be made, concerns the actual form of the coincidence histogram and the

information which is derived from it Apart from being used in testing the hypothesis of independent firings, the auto- and cross coincidence histograms, supplemented with results from detailed model studies, are used in multi-unit investigations to give information about the possible nature of synaptic interaction between the investigated neurons. The detailed structure of the histograms is essential in that case. It is obvious that in our case the shape of the coincidence histogram reflects properties of the stimulus ensemble as well as characteristics of the s-e relation of the investigated neuron. (Note for instance the wide "tails" in the histogram for unit 107-1 in Fig 4.) These effects, however, are interwoven in a complicated way. In the present stage the coincidence histogram is primarily intended to test the existence of a possible s - e relation, additional evidence and hypotheses regarding the nature of the s - e relation are necessary to be able to draw more specific conclusions from the detailed structure of the histogram

Results obtained up till now suggest that on the whole cochlear nucleus units tend to have a narrower peak in the cross coincidence histogram than auditory cortex units have.

Some formal problems associated with the analysis as presented here have to be considered the choice of binwidth Δ and the problem of statistical significance of features in the cross coincidence histogram.

Purely formal arguments for a particular choice of Λ cannot be derived from the theory It is possible however to restrict the range of possible values of Λ on the basis of knowledge about the physiological mechanisms associated with the point processes to be analyzed

On the one hand it appears to be rather pointless to take a value for Δ which is smaller than some 0.1 ms, since such a value suggests a time resolution in the cat's auditory system which contradicts the experimentally found loss of phase-lock at frequencies above a few kHz in the auditory nerve (Kiang et al, 1965) and the cochlear nucleus (Lavine, 1971) There appears to be no evidence for phase-locking by auditory cortex units (Goldstein and Abeles, 1975) On the other hand the available data about integration times for neurons in the cat's cochlear nucleus (some 10 20 ms, Van Gisbergen et al, 1975c) and auditory cortex (up to 200 ms, Abeles and Goldstein, 1972, De Ribaupierre et al., 1972) suggest an upper limit for Δ Both the lower and upper limit thus appear to vary with the level in the auditory system, with smaller values for the periphery For various neurons we have varied Δ in the range from 0.2 to 10 ms the global structure of the cross coincidence histograms (a "peak" around $\tau = 0$) was not seriously affected by this In case the histogram did not show a clear structure, variation of Δ was not able to improve this For various reasons a value for Δ of about 1–4 ms seems to be a reasonable choice

We have not been able to develop satisfactory statistical tests of the null hypothesis of independent firings This problem is also mentioned in the case of stationary spike sequences as studied in multi-unit investigations (Moore et al, 1966, Perkel et al, 1967) Theoretical problems in this respect are aggravated by the nonstationary character of the point processes to be analyzed here

In practice this implies that the histograms were judged by the – quite arbitrary – criterion whether a peak around $\tau = 0$ "clearly" stood out against a flattening background for larger values of $|\tau|$ In this respect unit 96-4 (Fig 5) may be considered as a limiting case, only just passing the test.

We realize that the problems of significance and of the choice of binwidth Δ are critical ones as they are of fundamental importance for a judgement regarding the reliability of results of this analysis Further theoretical analysis on these points as well as on related problems is currently in progress (Barndorff-Nielsen and Pedersen, Mathematical Institute, University of Århus, Denmark, in collaboration with the authors of this paper). This theoretical analysis explicitly takes into account the *nonstationary character* of the point processes to be analyzed here.

Results obtained so far suggest that the procedure as described in this paper is capable of demonstrating the existence of a stimulus-event relation, including a number of cases in which this was hard or even impossible to demonstrate by mere (visual or aural) inspection of the experimental data

For a number of neurons in the auditory cortex the existence of a s - e relation could not be demonstrated Apart from a possible insensitivity of the procedure several types of shortcomings of the stimulus ensemble may very well be the cause of this For instance it should be realized that the stimulus ensemble as it was used in this investigation did not contain frequencies exceeding 20 kHz, whereas neurons in the cat's auditory cortex with characteristic frequencies as high as 45 kHz have been reported (Evans and Whitfield, 1964).

It is clear that more theoretical insight as well as experimental evidence to be gained in further investigations will help to obtain a more complete evaluation of the procedure as proposed

A fundamental problem, often met in single cell sensory electrophysiology, is the question whether a particular neuron should, on the basis of physiological data, be considered to be an "*auditory*" neuron (or a visual one in another context) The notion of "auditory" commonly is operationalized by requiring that the neuron's firing pattern should be affected reproducibly (within statistical limits) by presentation of a proper acoustic stimulus: "auditory" in these terms means "liable to be influenced by adequate acoustic stimulation".

This functional definition works both ways. Any neuron, the firing pattern of which is influenced reproducibly by acoustic stimulation, is considered to be an "auditory" one. It is obvious that these findings, on the basis of physiological evidence, should be confronted with other, i.e. anatomical, evidence. Furthermore, it should be realized, that an answer to the question whether a particular neuron is an "auditory" one, is valid only within the reach of the presented stimulus ensemble, i.e. within the range of questions posed to the neuron. It may very well be possible that for different stimulus ensembles the question of the neuron being "auditory" is answered in different ways. This is related to the notion of "adequate" stimulation. Especially in areas of the central nervous system, involved in a more complex mode of sensory information processing, this may form a serious problem, the more so as available recording times are finite. The approach of using a stimulus ensemble containing only a small selection of biologically significant sounds (usually from the animal's vocal repertoire), presented many times, may very well, wrongly and unintentionally, give rise to the idea of highly specific neurons (Scheich, 1977).

Therefore we have used a wide stimulus ensemble, inspired by ethological and ecological observations. not restricted to species-specific vocalizations but containing all sorts of sounds from the natural environment of the animal (cat). Consequently, a neuron is considered to be "auditory" if for this wide stimulus ensemble a stimulus-event relation is shown to exist. If such a relation cannot be demonstrated, we have no arguments to call the neuron an "auditory" neuron in a functional sense. This point has been stressed somewhat amply because, to our opinion, it is of fundamental importance to define the subject of investigation as precise as possible. Once a stimulus-event relation has been shown to exist, the nature of it has to be clarified, which may not in all cases be easy or even possible with the available methods and present-day understanding of neural functioning. This, however, does not detract anything from the existence of the stimulusevent relation.

The procedure as it is described should be able to demonstrate the influence of external factors on a possible s - e relation. In this respect one might think of influences from other modalities (e.g. visual, tactile) or external factors like attention, level and type of anaesthesia, hormonal condition etc. Therefore this approach may be a useful tool in an exploratory investigation of complex regions in the central nervous system, which are known to receive inputs from different modalities or the function of which is not clear yet.

Acknowledgements. This investigation forms part of the research program of the Working Group Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (the Netherlands) and is supported by the Netherlands Organization for Advancement of Pure Research (Z.W.O). Discussions with Jan van Gisbergen and Ton Vendrik helped to shape the ideas. The expertise of Koos Braks, Jan Bruijns, Paul van Dongen, and Hans Krijt were of great help in the experiments and data analysis

References

- Abeles, M, Goldstein, Jr., M.H.: Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. Brain Res. 42, 337-352 (1972)
- Aertsen, A.M.H.J., Olders, J.H.J., Johannesma, P.I M.: Neural representation of the Acoustic Biotope: Spectro-temporal analysis of auditory neuron responses to tonal and natural stimuli (in preparation)
- Boer, E de, Kuiper, P.: Triggered correlation. IEEE Tr. on Biomed. Eng. BME-15, 169-179 (1968)
- Boer, E. de, Jongh, H. R. de. On cochlear encoding: potentialities and limitations of the reverse-correlation technique. J. Acoust. Soc. Am. 63, 115-135 (1978)
- Bryant, Jr., H.L., Ruiz Marcos, A., Segundo, J.P : Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs J Neurophysiol 36, 205-225 (1973)
- Bullock, T.H (ed.): Recognition of complex acoustic signals Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Cox, D. R, Lewis, P A.W.: The statistical analysis of series of events London: Methuen 1966
- Dahl, F., Grundlagen einer ökologischen Tiergeographie. Erster Teil. Jena: Fischer 1921
- Dickson, J.W., Gerstein, G.L.: Interactions between neurons in auditory cortex of the cat J. Neurophysiol 37, 1239–1261 (1974)
- Evans, E F, Whitfield, I.C.: Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat J. Physiol. 171, 476-493 (1964)
- Evans, E.F. Cochlear nerve and cochlear nucleus. In Handbook of sensory physiology Vol V/2 Auditory system, pp. 1–108. Keidel, W.D., Neff, W.D., eds. Berlin, Heidelberg, New York. Springer 1975
- Gerstein, G L.: Functional association of neurons. detection and interpretation. In: The Neurosciences · second study program, pp. 648 661 Quarton, G C., Melnechuk, T., Adelman, G A., eds New York. Rockefeller University Press 1970
- Gisbergen, J.A.M. van, Grashuis, J.L., Johannesma, P.I.M., Vendrik, A J H : Spectral and temporal characteristics of activation and suppression of units in the cochlear nuclei of the anaesthetized cat. Exp. Brain Res 23, 367–386 (1975a)
- Gisbergen, J A M. van, Grashuis, J.L., Johannesma, P I.M., Vendrik, A J H · Neurons in the cochlear nucleus investigated with tone and noise stimuli Exp Brain Res. 23, 387–406 (1975b)
- Gisbergen, J.A.M van, Grashuis, JL, Johannesma, PIM, Vendrik, AJ.H. Statistical analysis and interpretation of the initial response of cochlear nucleus neurons to tone bursts Exp Brain Res. 23, 407–423 (1975c)
- Goldstein, Jr., M H., Abeles, M., Single unit activity of the auditory cortex In. Handbook of sensory physiology Vol V/2, Auditory system, pp 199 218 Keidel, W.D., Neff, W D, eds. Berlin, Heidelberg, New York. Springer 1975

- Grashuis, I L The pre-event stimulus ensemble an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons Ph D Thesis, Nijmegen, The Netherlands 1974
- Griffith J S, Horn, G Functional coupling between cells in the visual cortex of the unrestrained cat Nature **199**, 878, 893-895 (1963)
- Johannesma, P I M The pre-response stimulus ensemble of neurons in the cochlear nucleus In Proc of the IPO Symp on Hearing Theory, pp 58-69 Cardozo, B L, ed Eindhoven, The Netherlands 1972
- Kiang, N Y S, Watanabe, T, Thomas, E C, Clark L F Discharge patterns of single fibers in the cat's auditory nerve Cambridge, Mass M I T Press 1965
- Lavine R A Phase-locking in response of single neurons in cochlear nuclear complex of the cat to low-frequency tonal stimuli J Neurophysiol 34, 467–483 (1971)
- Lee, Y W Statistical theory of communication New York, Wiley 1960
- Lee, Y W, Schetzen, M Measurement of the Wiener kernels of a nonlinear system by cross-correlation Int J Control 2, 237–254 (1965)
- Marmarelis, PZ, McCann, GD Development and application of white noise modeling techniques for studies of insect visual nervous system Kybernetik 12, 74 89 (1973)
- Marmarelis, PZ, Naka, K-I Nonlinear analysis and synthesis of receptive-field responses in the catfish retina I Horizontal cell→ganglion cell chain J Neurophysiol **36**, 605 618 (1973a)
- Marmarelis, P.Z., Naka, K.-I. Nonlinear analysis and synthesis of receptive-field responses in the catfish retina II One-input white-noise analysis J. Neurophysiol. 36, 619–633 (1973b)
- Marmarelis, P Z, Naka, K-I Nonlinear analysis and synthesis of receptive-field responses in the catfish retina III Two-input white-noise analysis J Neurophysiol 36, 634 648 (1973c)
- Moore GP, Perkel, DH, Segundo, JP Statistical analysis and functional interpretation of neuronal spike data Annu Rev Physiol 28, 493-522 (1966)
- Moore G P, Segundo, J P, Perkel D H, Levitan, H Statistical signs of synaptic interaction in neurons Biophys J 10, 876–900 (1970)

Newman, J D, rapp Biological filtering and neural mechanisms group report In Recognition of complex acoustic signals pp 279 306 Bullock, T H, ed I ife Sciences Res Rep, Vol 5 Berlin, Heidelberg, New York Springer 1977

Odum, E.P. Fundamentals of ecology Philadelphia Saunders 1971

- Papoulis, A The Fourier integral and its applications New York McGraw-Hill 1962
- Perkel, D H, Gerstein, G L, Moore, G P Neuronal spike trains and stochastic point processes II Simultaneous spike trains Biophys J 7, 419 440 (1967)
- Ribaupierre, F de, Goldstein, Jr, M H, Yeni-Komshian, G Intracellular study of the cat's primary auditory cortex Brain Res 48, 185 204 (1972)
- Scheich, H Central processing of complex sounds and feature analysis In Recognition of complex acoustic signals pp 161 182 Bullock, T H, ed Life Sciences Res Rep, Vol 5 Berlin, Heidelberg, New York Springer 1977
- Scelen, W von, Hoffman, K -P Analysis of neuronal networks in the visual system of the cat using statistical signals Biol Cybernetics 22, 7-20 (1976)
- Smolders, J W T, Aertsen, A M H J, Johannesma, P I M Neural representation of the acoustic biotope on the usefulness of biologically relevant sounds in the investigation of single auditory neurons (in preparation)
- Worden, F.G., Galambos, R. (eds.) Auditory processing of biologically significant sounds. Neurosci. Res. Prog. Bull., Vol 10 Brookline, Mass. Neurosciences Research Program 1972

Received November 25, 1978

A M H J Aertsen Dept of Medical Physics and Biophysics University of Nijmegen Geert Grooteplein Noord 21 Nijmegen The Netherlands

CHAPTER 3

Neural Representation of the Acoustic Biotope

A Comparison of the Response of Auditory Neurons to Tonal and Natural Stimuli in the Cat

Smolders, J.W.T., Aertsen, A.M.H.J. and Johannesma, P.I.M. Biol. Cybernetics <u>35</u>, 11-20 (1979)



Neural Representation of the Acoustic Biotope

A Comparison of the Response of Auditory Neurons to Tonal and Natural Stimuli in the Cat

J. W. T. Smolders*, A. M. H. J. Aertsen, and P. I. M. Johannesma Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Abstract. Cats were stimulated with tones and with natural sounds selected from the normal acoustic environment of the animal. Neural activity evoked by the natural sounds and tones was recorded in the cochlear nucleus and in the medial geniculate body. The set of biological sounds proved to be effective in influencing neural activity of single cells at both levels in the auditory system. At the level of the cochlear nucleus the response of a neuron evoked by a natural sound stimulus could be understood reasonably well on the basis of the structure of the spectrograms of the natural sounds and the unit's responses to tones. At the level of the medial geniculate body analysis with tones did not provide sufficient information to explain the responses to natural sounds. At this level the use of an ensemble of natural sound stimuli allows the investigation of neural properties, which are not seen by analysis with simple artificial stimuli. Guidelines for the construction of an ensemble of complex natural sound stimuli, based on the ecology and ethology of the animal under investigation are discussed. This stimulus ensemble is defined as the Acoustic Biotope.

1. Introduction

In the neurophysiological study of the peripheral auditory system the stimuli used are mostly pure tones or tone bursts, clicks and continuous noise or noise bursts (Evans, 1975). Natural sounds have been used seldomly, and mainly for illustrative purposes (Watanabe, 1972; Kiang and Moxon, 1974). Representation of the experimental results in terms of stimulus-response relationships has mostly been low-dimensional: one aspect of the neuronal spike trains vs. one or two parameters of the stimulus. Examples are the tuning-

curves (Kiang et al., 1965), PST-histograms (Kiang et al., 1965) and reverse correlation functions (De Boer, 1968; De Boer and De Jongh, 1978). Arguments for this approach are the simplicity of these stimuli and the ease with which they are generated in a laboratory. Moreover the quasilinear behaviour of peripheral parts of the auditory system allows an explanation of neural responses to complex and behaviourally interesting sounds from the responses to simple artificial stimuli. In higher parts of the auditory system the approach with this type of stimuli is less profitable; they are less effective and information on the responses to simple stimuli may not help in understanding the responses to complex stimuli (Evans, 1974). More recently investigations with more complex sounds have been made. A problem is the choice of the complex stimuli. One may start for instance with pure tones and gradually increase the degree of complexity by introducing amplitude and frequency modulation, addition of harmonics and noise. In most cases the complexity induced is relatively simple: modulation with linear or logarithmic sweeps, ramps or sinusoidal frequency and intensity variations (see e.g. Møller, 1973, 1978; Evans, 1974). In some special cases (e.g. FM bats (Suga, 1972)) such signals may resemble natural stimuli, but in most cases their structure does not resemble that of natural sounds. Another way is to start from a wide ensemble of complex sounds that may play a role in the natural behaviour of the experimental animal. This approach may be based on the argument that in phylogenesis the acoustic system has adapted structurally and functionally to process the sounds which occur in the environment of the animal and which are important for behaviour (Worden and Galambos, 1972). On the other hand species vocalisations have adapted in phylogenesis and ontogenesis to the properties of the auditory system (Marler, 1961; Sebeok, 1968; Gersuni and Vartanian, 1972; Worden and Galambos, 1972; Suga, 1978). A recent collection of papers related to this

^{*} Present address Zentrum dcr Physiologie dcr JW Goethe-Universitat Frankfurt, Federal Republic of Germany



Fig. 1. The frequency response of the closed sound system The graph is representative for all experiments and shows the sound pressure level (re $2\,10^{-5}$ N/m²) as a function of frequency at 0dB attenuation Relative sound levels in the Figures represent dB attenuation relative to this level



Fig. 2. A segment from a sequence of smooth tone bursts Lxplanation in the text

approach is provided by the Dahlem report (Bullock, 1977)

In the present investigation a proposal is made for the construction of a wide stimulus ensemble on an ecological and ethological basis, not limited to specialised behavioural functions and neural structures (Suga, 1972) or animal vocalisations (Gersuni and Vartanian, 1972) A necessary condition for the usefulness of such a stimulus ensemble for neurophysiological investigation of single cell properties is that biological sounds influence neural activity in a significant and reproducible manner

The methodological aspects of testing whether neuronal responses are indeed effected by a long complex ensemble of natural sounds in a reproducible way have been treated recently (Aertsen et al, 1979) It is the aim of the present paper to investigate whether responsiveness to natural sounds can be explained on the basis of the responsiveness to tones, or if some new information is gained from the use of such a wider stimulus ensemble. To this end we have compared single unit responses to natural sounds with those to tonal stimuli in a peripheral structure (cochlear nucleus CN) as well as in a more centrally located nucleus (medial geniculate body MGB) of the cat

2. Methods

21 Preparation, Recordings and Histological Procedures

Five yound adult laboratory grown cats were used for the experiments. The methods used for the preparation of the animal record ing of single unit activity and histological control of the anatomical location of neurons were the same as have been described in detail by Van Gisbergen et al., (1975). Only a brief outline is given here

The animals were premedicated with 0.25 mg atropine sulphate and anaesthetised with an intraperitoneal injection of Nembutal (30 mg/kg) A cannula was placed in the trachea. The head of the animal was fixed in a stereotaxic apparatus by means of a small holder screwed to the skull. A small hole was drilled in the skull over the desired brain area. Rectal temperature was kept at 37-38 °C

Neural activity was recorded from two levels in the auditory system the cochlear nucleus and the medial geniculate body For experiments at the level of the cochlear nucleus the anaesthesia was maintained by additional intravenous injection of Nembutal (6 mg/kg) whenever reflexes to pinching of the skin between the toes of the hindlegs occurred For experiments at the level of the medial geniculate body local anaesthesia of wounds was applied by infiltration with 0.5 % Marcaine before the administration of Nembutal was stopped and the animals were respirated with a mixture of 30% O₂ and 70 ° N₂O. The cats were immobilized by continuous infusion of a Flaxedil/glucose solution (0.2 % Flaxedil, 5 % glucose in Ringers solution) at a rate of 5 ml/h. The artificial respiration was adjusted for an end tidal CO₂ percentage of 3.8

Neural activity was recorded extracellularly using glass micropipettes filled with 0.5 M sodiumacetate and 2% Pontamine Sky Blue. Tip diameters were approximately 1 μ m and impedances varied from 7 to 20 M Ω Neural activity was amplified and action potentials were transformed to unitary pulses by means of a level discriminator for on line computer analysis At characteristic points in an electrode track dye deposits were made for later histological identification of the recording sites. For this purpose, at the end of the experiment the animals were deeply anaesthetised and perfused with a 10% formaline solution buffered to pH 4. Histological sections were made in a sagittal plane through the cochlear nucleus and in a frontal plane through the medial geniculate body. The sites of recording were reconstructed from the dye deposits made in each electrode track and the registration of recording depth with a micrometer.

22 Acoustic Stimulation

During the experiments the cat was placed in an electrically shielded acoustic chamber (IAC type 1202 A) Acoustic stimuli were applied by means of half inch condensor microphones (B&K 4134) and acoustic couplers, filled with damping material, sealed into the cat's external ear canals (Van Gisbergen et al., 1975) The sound pressure level was monitored using an identical B&K microphone. The frequency response of the system was measured in situ at the beginning of each experiment. A representative result is shown in Fig. 1.

Relative sound levels mentioned in the text refer to the sound pressure level (SPL re $2 \, 10^{-5} \, \text{N/m}^2$) obtained with an attenuator setting of 0dB Harmonics in the acoustic output of the sound generating system were at least 40 dB down Stimulation was to both ears (I + C) unless stated otherwise (I for stimulation ipsilaterally C for stimulation contralaterally with respect to the recording electrode)

The stimulus ensemble consisted of two classes of sound stimuli Natural sounds (vocalisations of animals and abiotic sounds gener ated by wind and water) were recorded on analog tape (recorder Uher 4200, tape speed 19 cm/s, microphone Sennheiser transistor condensor microphone MKH 805) Selected sounds were repeatedly presented from a Sony TC 252 recorder in order to compile PST-histograms

Artificial stimuli consisted of sequences of smooth tone bursts The envelope of the bursts (squared cosine) was approximately Gaussian in order to minimize the product of the duration and bandwidth (Gabor, 1946) The frequency and peak intensity of subsequent tones were varied in a pseudorandom manner (Fig 2) Standard parameters were 127 frequency values (range 1 to 8 octaves) and 63 peak intensity values (range 40 dB) Both the frequency and the peak intensity values were distributed uniformly on log scale. The time intervals between the onsets of the Gaussian shaped tone bursts as well as the duration of the bursts were constant in one sequence (standard values 80 ms each)

23 Analysis of Natural Sound Stimuli

The set of natural sounds used for the present experiments consisted of 9 elements, arbitrarily chosen at this stage of the investigation The sounds were classified into three groups a) animal vocalisations with a low background noise level, b) animal vocalisations with a noisy background, c) noisy sounds (possessing no definite tonal structure) For a more detailed analysis spectrograms of each sound were made The bandwidth of the sounds was limited to 10kHz by low-pass filtering (Krohn-Hite 3343, 96 dB/oct) and the sounds were sampled by a digital computer at a rate of 20kHz. The power spectral density as a function of time (the dynamic power spectrum) was calculated from these samples The mewing sound of the tomcat (Fig 3) may serve as an example Figure 3-1 shows a representation equivalent to conventional sonograms (Potter, 1947) In Fig 3-2 the same data are represented in a threedimensional way. Power as a function of frequency and time is presented as an amplitude on the time-frequency plane (Singleton and Poulter, 1967) allowing a more detailed inspection of the sound structure. The following properties of the sound become clear in this analysis. It has a tonal structure and shows many harmonics The fundamental frequency is at about 600 Hz, the dominant frequency at 12kHz (Fig 3-3) There is only moderate modulation of frequency and amplitude A substructure of noisy and harmonic parts is clearly discernible (compare Figs 3-3 and 3-4)

24 Analysis of the Stimulus-Response Relation

241 Stimulation with Natural Sounds The relation between the natural stimulus and the associated neural activity was studied by means of peristimulus time histograms (PSTH's) neural responses to repeated presentation of the stimulus were summated (time axis was divided in 1000 bins) The PSTH's were compared with the dynamic spectrum of the sounds, presented on the same time scale. The beginning and the end of the PSTH's are indicated by vertical bars in the spectrograms, the arrow S in the PSTH indicates the level of spontaneous activity, which is 0 in this case (Fig 4-1) This representation allows simple visual comparison of the stimulus Figure 4-1 illustrates this for a simple cochlear nucleus neuron

242 Stimulation with Smooth Tone Bursts The frequency-intensity response area in the classical sense was investigated with a sequence of smooth tone bursts, pseudorandomly varied in frequency and intensity (see Sect 22)

Each action potential was represented by a dot in a cubic display *IFT-cube* The position of the dot was determined by three stimulus parameters

I The maximum *intensity* (in dB attenuation) of the tone burst during which the action potential occurred

F The carrier *frequency* of this tone burst (in kHz on a logarithmic scale)



Fig. 3. The spectrogram of the mewing sound of the tomcat, illustrating the substructuring of noisy and harmonic parts of the sound The resolution of the spectrogram is 40 Hz, 25 ms 3-1. The conventional representation as a sonogram (intensity resolution 10dB) 3-2. The same data in a threedimensional representation 3-3. Power spectrum of time sect A of the sound (282-307 ms) showing the harmonic structure of this section. The dominant component is at $1 \ 2 \ Hz$ 3-4. Power spectrum of another time sect B of the sound (896 921 ms) where harmonic structure is lost. The location of sects A and B is indicated in 3-1, 3-2.

T The time difference between the onset of the tone burst and the occurrence of the action potential (in ms) The length of the time axis equals the interval between onsets of tones

For the purpose of good visualisation the IFT-cube (see Fig 4 3) was divided into 5 intensity windows, with equal width and equidistant on the dB attenuation scale, and neural activity within each window was projected onto a FT-plane in the middle of that window. So each FT plane represents the responses of the neuron to tone bursts of different frequencies and approximately the same peak intensity (within $\pm 4 dB$ in Fig 4-3) i.e. an isointensity scan Projection of the data on a vertical IF-plane presents the response area of the neuron as a function of frequency and intensity (Fig 4-2) The outer border of the cloud of dots represents the "tuning curve"

3. Results

The data were obtained from 5 cats Neural activity was recorded from 6 units in the cochlear nucleus and from 19 units in the medial geniculate body. The sites of measurement were verified histologically

3 1. Cochlear Nucleus

Results from the analysis of unit 93-3, located in the central region of the dorsal cochlear nucleus (nomen-



Fig. 4. Unit 93-3. Analysis of a cochlear nucleus neuron with natural and artificial stimuli 4-1 Spike responses as a function of time (top) and power spectral density of the sound of the coot (Fulica atra) as a function of time. The sound was presented with a maximum peak intensity of $-40 \, dB$ relative sound level and repeated 64 times. Presentation was to both ears Beginning and end of the PSTH are indicated with vertical bars. Resolution is 1000 bins. The time mark corresponds to 1 s. The vertical axis represents the total number of action potentials per bin. Spontaneous activity is indicated by the arrow S. The resolution of the spectrogram is 40 Hz, 25 ms, 10 dB 4-2 Neural response as a function of frequency and intensity summated over the duration of the tone bursts (80 ms). Action potentials are represented as dots. The outer border of the dotted area represents the "tuning curve" of the neuron. The characteristic frequency is approximately $2 \, kHz$. $4-3 \, IFT$ -cube: neural activity as a function of tone burst maximum intensity I, tone burst frequency F and time T after tone burst onset. Explanation in the text

clature from Osen, 1969a, 1969b) are presented in Fig. 4. Comparison of the PSTH and the spectrogram of a vocalisation of the coot (Fulica atra) shows that stimulus evoked action potentials occur only when there is appreciable energy at frequencies above 1 kHz (Fig. 4-1). The level of spontaneous activity for unit 93-3 is zero. The sound of the coot has a harmonic structure: it consists of three "tweets", with tonal components at 1.8, 2.7 and 3.6 kHz. Apparently these tweets activate the neuron whereas the background noise in the vocalisation is confined to frequencies responses are higher at the on- and offset of each tweet. The maximum peak relative sound level at presentation time was -40 dB.

A comparison of these data with the results of tonal analysis shows that for this cochlear nucleus neuron the responses to natural sounds can be understood quite well from tone burst responses. The background noise in the vocalisation is confined to frequencies below 1 kHz; from tonal analysis it is seen that the neuron is not activated by tone bursts with frequencies below 1 kHz at intensities below $-40 \, \text{dB}$ relative sound level (Figs. 4-2, 4-3). This explains why the PSTH shows no events inbetween the coot tweets. The tweets have energy up to $-40 \, \text{dB}$ relative sound level at 1.8 and 2.7 kHz and the neuron is activated by tone bursts of these frequencies and intensities (Figs. 4-2, 4-3). This explains the high event density at time moments corresponding to the occurrence of the tweets.

One discrepancy between the analysis with natural sounds and the analysis with tone bursts is very obvious: the response to the coot tweets is higher at



Fig. 5-10. IFT-cube neural activity as a function of tone burst maximum intensity I, tone burst frequency F and time T after tone burst onset Explanation in the text


Fig. 5. Unit 93-4. Analysis of a cochlear nucleus neuron with a variety of natural sounds. The maximum peak intensity for all sounds was $-40 \, \text{dB}$ relative sound level. Each sound was presented 64 times to both ears. The resolution of the PSTH's is 1000 bins. The time mark corresponds to 1 s. The vertical axis represents the total number of action potentials per bin. Spontaneous activity was 16.5 imp/s (2.7 imp/bin in the PSTH) as indicated by S The resolution of all spectrograms is 40 Hz, 25 ms, 10 dB. The 9 biological sounds are. 5-1. Mewing of tomcat (Felix catus). 5-2. Sound of cuckoo (Cuculus canorus). The vocalisation occurs twice in the sonogram 5-3. Sound of blackbird (Turdus merula) 5-4. Song of a robbin (Erithacus rubecula). 5-5. Vocalisation of coot (Fulica atra) 5-6. Alarm sound of wren (Troglodytes Aroglodytes) 5-7. Wind, rustling trees, tutmouses (Parus spec.) 5-8. Barking of dog (Canis), preceeded and followed by growling 5-9. Wind, rustling leaves, twigs, breaking twigs



Fig. 6. Unit 90-3. Analysis of a neuron from the medial geniculate body with mewing of a tomcat and with tone bursts. The maximum peak intensity of the mewing sound was 0dB; it was repeated 256 times. 6-1 Sonogram of mewing sound and PSTH of neural activity evoked by contralateral stimulation (C). Further details as in Fig. 4-1. 6-2 IFT-cube: neural activity as a function of tone burst maximum intensity I, tone burst frequency F and time T after tone burst onset. Explanation in the text. Stimulation was bilaterally (I + C). 6-3 IFT-cube upon contralateral stimulation (C). 6-4, 6-5, 6-6 Random FT-scans upon bilateral (I + C), ipsilateral (I) and contralateral (C) stimulation. Explanation in the text

the on- and offset (Fig. 4-1) whereas the responses to the smooth tone bursts show gradual build up and decline (Fig. 4-3). The temporal structure of the stimulus apparently plays a role here.

Figure 5 presents the experimental results for unit 93-4, which was located in the posteroventral cochlear nucleus at the border of the pyramid cell layer. The unit was tested with a variety of natural sounds.

In all cases except in 5-4 the neural activity on stimulation with a natural sound is quite different from the spontaneous activity which amounts to 16.5 impulses/s and is indicated by the arrows in the PSTH's. So for eight out of nine natural sounds a stimulusresponse relation clearly exists.

The responses to tone burst stimuli show simple activation for intensities below -60 dB attenuation, at frequencies of about 1.3 kHz (Fig. 5-10). At higher

intensities a complex pattern of activation and suppression occurs for tone burst frequencies in the range of 0.7-2 kHz (Fig. 5-10). If we compare the results from the stimulation with tone bursts to those of the stimulation with natural sounds some conclusions on the stimulus-response relation can be made. For those natural sound stimuli which have, at certain moments of time, considerable energy in the 0.7-2 kHz frequency range (Figs. 5-1, 5-2, 5-3, 5-5, 5-8) there is a high correlation between the time structure of the PSTH and the spectral density in this frequency range. Sometimes simple activation appears to occur (Fig. 5-1), sometimes both activation and suppression (Figs. 5-2, 5-3, 5-5). Note in Fig. 5-2 the differences in high frequency composition of both vocalizations and the similarity of the corresponding neural responses in the PSTH. The result in Fig. 5-3 appears to be con-

sistent with these observations, however, at this level of analysis it remains unclear whether indeed the energy in the 0.7-2 kHz frequency band is responsible for the strong response between 1.0 and 1.3 s. In the remaining natural sound stimuli there is less energy in the 0.7-2 kHz frequency range and the stimulus-response relation is less clear. This is seen best in Fig. 5-4 where the sound evoked activity nearly equals the spontaneous level and in Fig. 5-6. It must be emphasized however that, although its nature is not immediately clear from Figs. 5-6, 5-7, 5-8, 5-9, a stimulus-response relation certainly exists, as concluded from the effectivity of all of these stimuli. The structure of the PSTH's in these cases is too complex to investigate the stimulus-response relation in a qualitative way only. The explanation or prediction of the responses to the natural sounds from the responses to tone bursts can be only partly successful since in the natural sounds complex spectro-temporal patterns are presented and in most cases the time envelopes are different from the 80 ms smooth envelope of the tone bursts. From the analysis with the tone bursts no information can be obtained on interactions of more frequency components with different intensities nor any information on the effect of envelopes other than the particular one used. This may be illustrated by the "sustained" response in PSTH 5-1 to the tonal harmonic vocalisation of the tomcat and the "suppression" response in PSTH 5-5 to the short tone "tweets" in the vocalisation of the coot. Intermediates can be seen in Figs. 5-2 and 5-3.

A more detailed and more elaborate evaluation of the data is necessary to gain more than the general impression outlined above (see Discussion).

3.2. Medial Geniculate Body

In Figure 6-1 the response of unit 90-3 from the medial geniculate body is presented. This unit is located at the border of the magnocellular and the principal nucleus of the MGB (Berman, 1968). The stimulus is the mewing sound of the tomcat which was presented to the contralateral ear. The PSTH clearly indicates that there is a stimulus-response relation, although its nature is not obvious. The peaks in the PSTH are larger where the energy in the spectrogram is lower and the bandwidth smaller, but no simple relation of the time courses exists. The analysis with tonal stimuli reveals that the unit responds to tones of frequencies between 1 and 4kHz (Figs. 6-2, 6-3). There is an interaction of ipsilateral and contralateral stimuli, and there are differences in the response properties with only contralateral or ipsilateral stimulation (Figs. 6-4, 6-5, 6-6). The neuron is at least 20 dB more sensitive to contralateral than to bilateral stimulation, while the response type is completely different to both ways of stimulation (Figs. 6-2, 6-3). The unit was also investigated with long tone bursts (800 ms) (random FT-scans: Van Gisbergen et al., 1975). Although there is hardly any response to ipsilaterally presented tone bursts (Fig. 6-5), there is a clear influence of the ipsilateral stimulus (suppression) in the bilateral case (Figs. 6-4, 6-6). Additional information could be gained from other findings. The spontaneous discharge rate of the unit was 80 impulses/min. If stationary Gaussian wideband noise was presented to both ears the discharge rate fell to 3 impulses/min. Stimulation of only the ipsilateral ear appeared not to influence the discharge rate (75 impulses/min), whereas stimulation of the contralateral ear raised the discharge rate to 363 impulses/min. The unit responded better to frequency sweeps around 3kHz in the upward than in the downward direction.

The responses to stimulation with the noise stimulus and with 800 ms tone bursts are somewhat puzzling. When noise stimulation is used, the lowest discharge rate is obtained upon bilateral stimulation, whereas with tonal stimuli ipsilateral stimulation yields the lowest discharge rate (Fig. 6-5). Apart from the obvious differences which exist between these stimuli, it should be noted that there is a time difference of about two hours between these recordings. It is clear that the response to the natural sound cannot be explained by the responses to tone bursts not even with the additional information gained. There is some similarity, however, between the nature of the responses to tone bursts of 800 ms and the response to the tonal mewing sound of the tomcat, both responses are burstlike (Fig. 6-1, 6-3).

The results for unit 94-1 are shown in Fig. 7. This unit was located in the medial geniculate body, as concluded from the electrode track, but could not be localised precisely.

The response to repeated bilateral presentation (256 times) of the call of the cuckoo is shown in Fig. 7-1. The PSTH shows a clear structure, with lowest event density at moments corresponding with high power spectral density in the stimulus. This structure clearly indicates that the neural activity is influenced by the stimulus. Results from the analysis with binaurally presented tone bursts are shown in Fig. 7-2. A comparison of the neural activity during binaural stimulation (Fig. 7-3) with the activity during monaural stimulation (Figs. 7-4, 7-5) was made with tone bursts of 80 ms duration presented at 250 ms intervals. There was no detectable influence of the tone burst stimuli on the neural activity. Evidently it is impossible in this case as well to explain the responses to the natural sound from the neural activity during tone burst stimulation, at least at this level of analysis.



Fig. 7. Unit 94-1 Analysis of a medial geniculate body neuron with the vocalisation of the cuckoo and with tone bursts. The maximum peak intensity of the vocalisation was -20 dB relative sound level. 7-1 Sonogram of cuckoo vocalisation and PSTH of neural activity evoked by bilateral stimulation (I + C). Further details as in Fig. 4-1. 7-2 *IFT*-cube: neural activity as a function of tone burst maximum intensity *I*, tone burst frequency *F* and time *T* after tone burst onset. Explanation in the text. 7-3, 7-4, 7-5 Random *FT*-scans upon bilateral (I + C), ipsilateral (I) and contralateral (C) stimulation in the text. Note the difference between the duration of the tone bursts (80 ms) and the time interval between tone bursts onsets (250 ms) in this case

4. Discussion

As reported in the methods section Nembutal anaesthesia was used in the surgical procedure, maintained in the CN-preparation and replaced by Flaxedil relaxation in the MGB-preparation. It is known that cats take a long time (up to 8 hours) to recover from Nembutal anaesthesia, which has been demonstrated to have a profound effect on neuronal responsiveness (Webster and Aitkin, 1971; Kallert, 1974): less activity and less suppression are to be found. This has been shown to occur already at the level of the dorsal cochlear nucleus (Evans and Nelson, 1973). All MGB-units, except for three of them, were recorded well beyond the period of 8 hours after Nembutal anaesthesia was stopped. In the case of unit 90-3 (Fig. 6) about 6 hours had expired. A comparison of results for the MGB-units does not give rise to the idea that neural responsiveness was strongly influenced by the anaesthesia.

From the present results it is clear that complex natural acoustic stimuli are effective in evoking single unit responses at low and higher levels in the cat's auditory system.

At the level of the cochlear nucleus the responses to natural sounds can be understood fairly well from the responses to tone bursts if the spectrotemporal properties of the natural sound stimuli are taken into account. At the level of the medial geniculate body the responses to different types of stimuli are much more complex. Responses to natural sounds do not appear to be simply predictable from responses to tonal stimuli. Apparently at this level new information can



Diagram 1. Construction scheme for realisation of a stimulus ensemble inspired by the Acoustic Biotope of the cat

be gained by using natural stimuli, apart from technical stimuli.

One may wonder whether an experimental program, in which natural sounds with relatively short duration (in the present case about 2s) are repetitively presented in order to compile PST-histograms, is the most optimal way to investigate neural responses to natural stimuli. Apart from the time consuming nature of such an approach when many different sounds are to be presented, several aspects of the stimulus, like novelty, context and the like, are strongly affected by it. Although, probably, in the present experiments the latter will not play a very important role, this might well be the case when recording from awake, behaving animals. A stimulus ensemble which takes into account these considerations as well as those mentioned in the introduction has been constructed and used: the Acoustic Biotope (Aertsen et al., 1979). A further elaboration of this ensemble can be found in the Appendix.

The analysis of data obtained with the Acoustic Biotope as a stimulus ensemble must be able to

1) decide whether any relation between the stimulus ensemble and the neural activity exists (Aertsen et al., 1979), and

2) if so, determine the characteristics of the stimulus response relation.

From 2) it follows that the analysis should comprise a detailed description of the stimulus ensemble. This description must involve the physical characteristics (spectro-temporal properties) as well as biological characteristics (ethological and ecological observations) of the sounds. This analysis thus bears much relation to the notions of "information bearing elements" (Suga, 1973) and "bioacoustic dimensions" (Scheich, 1977). The analysis of the stimulus-response relation can be performed according to methods inspired by nonlinear system theory (Johannesma, 1972; Grashuis, 1974). A report on the results from this analysis for an ethologically based stimulus ensemble, which takes into account both first and second order aspects of the stimulus-response relation, is in preparation (Aertsen et al., in prep.).

Appendix

The Acoustic Biotope as a Stimulus Ensemble

The natural environment in which an animal lives, as far as functionally related to the animal, is defined as its *biotope* or *habitat*. It encompasses both the biotic and abiotic components of this environment (Dahl, 1921; Odum, 1971). Important characteristics are: geographical location, geophysical properties, climate regime, variety and numbers of occurring plants and animals The total of sensory aspects of the biotope forms the *Sensory Biotope*, which consists of different components, according to the different sensory modalities of the animal. The ensemble of sound signals from the biotope which can be perceived by the animal forms the *Acoustic Biotope*. It is clear that the Acoustic Biotope, and consequently the stimulus ensemble inspired by it, will vary for different animal species. Both the degree of complexity and the extent of the stimulus ensemble depend on the biotope of the animal as well as on its behavioural repertoire.

Elements from the Acoustic Biotope can be characterized according to different types of properties. Some of them are:

- ecological properties: sounds which have to do with geophysics, with animals, plants, with climate (wind, rain) etc.

- ethological properties. relation to behavioural context, for instance collection of food, avoidance of enemies, reproduction, social behaviour (Collias, 1959; Marler, 1961; Winter et al., 1966; Winter and Funkenstein, 1973).

physical properties: spectro-temporal characteristics of the sound signals: frequency bandwidth, FM, AM etc.

- statistical properties probability of occurrence of elements and combinations of elements.

A scheme for the composition of a wide ensemble of natural sound stimuli, inspired by the Acoustic Biotope of the cat, is shown in Diagram 1. This scheme takes into account the characteristics of sound signals as mentioned.

A stimulus ensemble based on this scheme has been constructed and used (Aertsen et al., 1979). The natural sounds used in the present investigation are a small selection from this ensemble. Note that technical sounds are included The purpose of this is to be able to relate neural responses to natural stimuli and these commonly used technical stimuli

Acknowledgements This investigation forms part of the research program of the Workgroup Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (The Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (Z W O) Discussions with Jan van Gisbergen and Ton Vendrik helped to shape the ideas The expertise of Jan Boezeman, Koos Braks, Jan Bruijns and Hans Krijt were a great help in the experiments and data analysis

References

- Aertsen, A M H J, Smolders, J W T, Johannesma, P I M Neural representation of the Acoustic Biotope On the existence of stimulus-event relations for sensory neurons Biol Cybernetics 32, 175–185 (1979)
- Aertsen, A M H J, Olders, J H J, Johannesma, P I M Neural representation of the Acoustic Biotope Spectro-temporal analysis of auditory neuron responses to tonal and natural stimuli in the grassfrog (Rana temporaria L) (in preparation)
- Berman, A L The brainstem of the cat Madison, Milwaukee, London Univ of Wisconsin Press 1968
- Boer, E de, Kuiper, P Triggered correlation IEEE Trans Biomed Lng BME-15, 169-179 (1968)
- Boer, E de, Jongh, H R de On cochlear encoding Potentialities and limitations of the reverse-correlation technique J Acoust Soc Am 63, 115-135 (1978)
- Bullock, T H (cd) Recognition of complex acoustic signals Life Sciences Res Rep, Vol 5 Berlin, Heidelberg, New York Springer 1977
- Collias, N.E. An ecological and functional classification of animal sounds In Animal sounds and communication, pp 368-391 Lanyon, W.E., Tavolga, W.N. (eds.) Am Inst Biol Sci., Washington 1959
- Dahl, F Grundlagen einer okologischen Tiergeographie, Teil 1 Jena Gustav Fischer 1921
- Evans, E F, Nelson, P G The responses of single neurones in the cochlear nucleus of the cat as a function of their location and the anaesthetic state Exp Brain Res 17, 402–427 (1973)
- Evans, E F Neural processes for the detection of acoustic patterns and for sound localization In The neurosciences, third study program, pp 131-145 Schmidt, FO, Worden, FG (eds) Cambridge MIT Press 1974
- Evans, E F Cochlear nerve and cochlear nucleus In Handbook of sensory physiology, Vol V/2 Auditory system, pp 1–108 Keidel, W D, Neff, W D (eds.) Berlin, Heidelberg, New York Springer 1975
- Gabor, D Theory of communication J IFF (I ondon), pt III, 93, 429-457 (1946)
- Gersuni, G V, Vartanian, I A Time dependent features of adequate sound stimuli and the functional organization of central auditory neurons. In Basic mechanisms in hearing, pp 623–673 Møller, A R (cd.) New York, London Academic Press 1973
- Gisbergen, J A M van, Grashuis, J L, Johannesma, P I M, Vendrik, A J H Spectral and temporal characteristics of activation and suppression of units in the cochlear nuclei of the anaesthetized cat Exp Brain Res 23, 367–386 (1975)
- Grashuis, J L The pre-event stimulus ensemble An analysis of the stimulus-response relation for complex stimuli applied to auditory neurons Ph D Thesis, Nijmegen 1974
- Johannesma, PIM The pre-response stimulus ensemble of neurons in the cochlear nucleus In Proc of the IPO Symp on Hearing Theory, pp 58-69 Cardozo, BL (ed) Eindhoven 1972

- Kallert, S Telemetrische Mikroelektrodenuntersuchungen am Corpus geniculatum mediale der wachen Katze Habilitationsschrift, Erlangen-Nurnberg 1974
- Kiang, NY-S, Watanabe, T, Thomas, EC, Clark, LF Discharge patterns of single fibers in the cat's auditory nerve Cambridge, Mass MIT Press 1965
- Kiang, NY-S, Moxon, FC Tails of tuning curves of auditory nerve fibres J Acoust Soc Am 55, 620–630 (1974)
- Marler, P The logical analysis of animal communication J Theoret Biol 1, 295 317 (1961)
- Møller, A R Coding of amplitude modulated sounds in the cochlear nucleus of the rat In Basic mechanisms in hearing, pp 593 619 Møller, A R (ed) New York, London Academic Press 1973
- Møller A R Coding of time-varying sounds in the cochlear nucleus Audiology 17, 446–468 (1978)
- Osen, K K Cytoarchitecture of the cochlear nuclei in the cat J Comp Neurol 136, 453-484 (1969a)
- Osen, K K The intrinsic organization of the cochlear nuclei in the cat Acta Oto-Laryng OL 67 352-359 (1969b)
- Odum, EP Fundamentals of ecology Philadelphia Saunders 1971
- Potter, R K, Kopp, G A, Green, H C Visible speech New York Van Nostrand 1947
- Scheich, H Central processing of complex sounds and feature analysis In Recognition of complex acoustic signals, pp 161– 182 Bullock, T H (ed) Life Sciences Res Rep, Vol 5 Berlin, Heidelberg, New York Springer 1977
- Sebeok, TA Animal communication Bloomington, London Indiana Univ Press 1968
- Singleton, R C, Poulter, T C Spectral analysis of the call of the male killer whale IEEE Trans Audio and Electroacoust AU-15, 104-113 (1967)
- Suga, N Analysis of information-bearing elements in complex sounds by auditory neurons of bats Audiology 11, 58-72 (1972)
- Suga, N Feature extraction in the auditory system of bats In Basic mechanisms in hearing, pp 675–742 Møller, A R (ed) New York, London Academic Press 1973
- Suga, N Specialization of the auditory system for reception and processing of species-specific sounds Fed Proc 37, 2342-2354 (1978)
- Watanabe, T Fundamental study of the neural mechanism in cats subserving the feature extraction process of complex sounds Jpn J Physiol 22, 569-583 (1972)
- Webster, W R, Aitkin, L M Evoked potential and single unit studies of neural mechanisms underlying the effects of repetitive stimulation in the auditory pathway Electroencephalogr Clin Neurophysiol 31, 581-592 (1971)
- Winter, P, Ploog, D, Latta, J Vocal repertoire of the squirrel monkey (Saimiri sciureus), its analysis and significance Exp Brain Res 1, 359-384 (1966)
- Winter, P, Funkenstein, H H The effect of species specific vocalisation on the discharge of auditory cortical cells in the awake squirrel monkey (Saimiri sciurcus) Exp Brain Res 18, 489 504 (1973)
- Worden, FG, Galambos, R (eds) Auditory processing of biologically significant sounds Neurosci Res Program Bull 10, Brookline Mass Neurosciences Research Program 1972

Received June 18, 1979

A M H J Aertsen

Department of Medical Physics and Biophysics University of Nijmegen

Geert Grooteplein Noord 21

Numegen

The Netherlands

CHAPTER 4

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

I. Characterization of Tonal and Natural Stimuli

Aertsen, A.M.H.J. and Johannesma, P.I.M. Biol. Cybernetics <u>38</u>, 223-234 (1980)

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

I. Characterization of Tonal and Natural Stimuli

A. M. H. J. Aertsen and P. I. M. Johannesma

Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Abstract. In order to investigate the spectro-temporal selectivity of single auditory neurons in the auditory midbrain of the grassfrog under different stimulus conditions a stimulus ensemble consisting of tonal and natural stimuli has been constructed. The selection of characteristics of the tonal stimuli was inspired by the general properties of vocalizations of the male grassfrog. The natural stimuli contained a variety of vocalizations of various anuran species and other sounds, taken from the natural environment of the grassfrog (i.e. its biotope). The main point of interest in the present paper is the spectro-temporal characterization of acoustic stimuli. For narrowband signals a parametric description in terms of instantaneous amplitude and instantaneous frequency appears to be attractive. A more general approach, also covering wideband signals, is a description by means of the dynamic power spectrum. Both types of analysis are described. Mathematically they can be based on the theory of analytic signals. The basic principles of this theory are summarized.

1. Introduction

The neural representation of the acoustic environment is investigated predominantly by the study of single neuron behaviour. Different types of stimuli (e.g. tones, clicks, noise) are presented to the animal and the pattern of action potentials, generated by the neuron, is studied as a function of one or more parameters of the stimulus.

Widely used methods to describe the single neuron behaviour are the tuning curves (Kiang et al., 1965), PST-histograms (Kiang et al., 1965), and the reverse correlation function (De Boer and De Jongh, 1978). Spectral selectivity, amplitude sensitivity and the time constants involved are the characteristics of prime importance. On the other hand ethological investigations have revealed the behavioural importance of spectrotemporal properties of (species-specific) vocalizations. The investigation of the underlying neural mechanism has received considerable attention (e.g. Worden and Galambos, 1972; Bullock, 1977).

The question to what extent the single neuron response to one type of stimuli, say natural sounds, can be understood from the response to another type of stimuli, say tones, has often been considered on a rather qualitative level (e.g. Smolders et al., 1979). The present investigation was undertaken to study this question in more detail. This paper (Part I) will be concerned with a description of the stimulus ensembles which were used: tonal stimuli and natural stimuli. In a companion paper (Part II, Aertsen et al., 1980) we introduce procedures to analyse the spectro-temporal properties of stimuli which influence the neural firing pattern: the neuron's spectro-temporal receptive field. A description of the procedures will be given, as well as the results of their application to a number of single unit responses to an ensemble of tonal stimuli. In a forthcoming paper (Part III) we will use one of these procedures to investigate the spectro-temporal receptive fields for natural stimuli.

The animal under investigation is the grassfrog, Rana temporaria L., which is quite common in Western Europe. For several anuran species both the spectral (e.g. Capranica, 1965; Gerhardt, 1974) and temporal (e.g. Van Gelder et al., 1978; Gerhardt, 1978) properties of species-specific vocalizations have been shown to be of behavioural significance. These findings, amongst others, suggest that the approach to investigate the frog's spectro-temporal receptive field properties under different stimulus conditions may provide information regarding the neural basis of behaviour in the natural acoustic environment.

A necessary prerequisite for this approach is a detailed description of the spectro-temporal properties



Fig. 1a and b. The "B-call" of the male grassfrog (Rana temporaria L.). a Time course of vocalization on different time scales, indicated by horizontal bars. Vertical bars refer to identical, arbitrary amplitude values. Note the quasiperiodic, pulsating character of the call. b Sonogram (dynamic amplitude spectrum) of vocalization in different representations, showing the spectro-temporal structure of the call: a dominant frequency band around 500 Hz and less intensive bands around 1100 Hz and 1500 Hz. Frequency resolution 39 1 Hz, time resolution 12.8 ms

of the stimulus ensemble. This analysis can be implemented in different ways, depending on the type of stimuli (narrowband vs. wideband). A description of the procedures which have been used to represent the stimuli of interest, tonal and natural sounds, is the main point of interest in the present paper. A comparison is made between on the one hand a parametric description in terms of temporal envelope and instantaneous frequency, which can be applied to narrowband stimuli, and on the other hand a spectro-temporal representation in the form of a dynamic power spectrum, applicable to an arbitrary signal.

2. The Tonal Stimulus Ensemble

2.1. Gamma-Tones

The selection of tonal stimuli, used in the present investigation, was inspired by the vocal repertoire of the male grassfrog. A typical example of a vocalization is shown in Fig. 1a. In general these vocalizations can be considered as quasi-periodic sequences of short similar sound elements (Van Gelder et al., 1978). Spectral analysis of the calls reveals a dominant frequency band in the range of 300–600 Hz, with smaller contributions from frequency bands in the ranges of 1000–1400 Hz and 1400–1900 Hz (Brzoska et al., 1977). These figures provide only a rough indication; they vary for the different types of calls. The characteristics of female calls as well as more complex patterns appearing in some male calls (e.g. frequency modulation) are not considered here. Significant correlations between temperature and call duration, number of sound elements in the calls as well as the repetition rate of sound elements have been demonstrated (Schneider, 1973; Van Gelder et al., 1978).

The male vocalization shown in Fig. 1a is classified as a "B-call" on the base of its physical characteristics: duration of the call and the number of sound elements (Van Gelder et al., 1978). The ethological context of this call leads to its classification as a "mating call" (Schneider, 1973; Brzoska, 1977) or a "chorus call" (Van Gelder et al., 1978). The dynamic spectrum or sonogram of this call, shown in Fig. 1b in two representations, shows the dominant frequency band around 500 Hz, with additional higher frequency bands at about 1100 Hz and 1500 Hz, which are due mainly to the "transient" onset of the sound elements.

An approximative formal description of a single sound element from the vocalizations is given by the amplitude modulated waveform

$$e(t) = Am(t)\cos(\omega t + \phi), \qquad (2.1)$$

where A = amplitude factor; $m(t) = \text{amplitude modu$ $lator}$; $\omega = \text{angular carrier frequency}$; $\phi = \text{phase angle}$.

For the amplitude modulator m(t) we propose

$$m(t) = c \left(\frac{t}{\beta}\right)^{\gamma - 1} \exp\left(\frac{-t}{\beta}\right) \quad t \ge 0 \quad \beta > 0$$
$$= 0 \qquad t < 0 \quad \gamma \ge 1, \qquad (2.2)$$

where β = duration parameter; γ = form parameter; $c = \left(\frac{2^{2\gamma-1}}{\beta\Gamma(2\gamma-1)}\right)^{1/2}$ normalization constant, to give $m^2(t)$ unit area; $\Gamma(z)$ = Gamma function (e.g. Abramowitz and Stegun, 1972).

The form m(t) appears both as the integrand in the definition of the Gamma function $\Gamma(\gamma)$ and as the density function of the Gamma distribution (e.g. Cox and Lewis, 1966), therefore we propose to use for (2.1) the term "Gamma-tone" or " γ -tone". The choice of the particular waveform (2.1) with m(t) specified by (2.2) was inspired, amongst others, by its use in a parametric description of impulse responses for single units in the cat's auditory nerve (De Boer, 1973) and cochlear nucleus (Johannesma, 1972; Grashuis, 1974), as obtained by the "reverse correlation" method (De Boer

and De Jongh, 1978). The introduction of this type of waveform goes back to Flanagan (1965) who used it in modelling the impulse response for the human basilar membrane.

Numerical values of the parameters which provide a reasonable approximation for the calls of the male grassfrog are

$$\beta = 1.45 \text{ ms}$$

$$\gamma = 3$$

$$\omega/2\pi = 500 \text{ Hz}$$

$$\phi = -\pi/2.$$
(2.3)

The stimulus ensemble in the present investigation consisted of sequences of y-tones with constant duration parameter β and form parameter γ [see (2.3)] and different values of frequency ω and amplitude factor A. The γ -tone sequences are described in detail in Sect. 4. A segment of such a sequence is shown in Fig. 2a.

2.2. Complex Representation of Signals

In the analysis of single unit responses to the γ -tone ensemble we will use a description of the stimulus ensemble in terms of nonlinear functionals of the signal x(t): the instantaneous amplitude or temporal envelope a(t), the instantaneous phase $\phi(t)$ and the instantaneous

angular frequency $\omega(t) \left[\text{or the instantaneous frequency} f(t) = \frac{\omega(t)}{2\pi} \right]$. These functionals can be defined elegantly by means of the concept of the analytic signal (Gabor, 1946; Ville, 1948). Using this convention a signal representation in terms of instantaneous amplitude a(t) and phase $\phi(t)$ bears a one-to-one relation to a representation in terms of signal values x(t). A signal representation in terms of instantaneous amplitude a(t) and frequency $\omega(t) = 2\pi f(t)$ determines the signal apart from its absolute phase. In the case of a tonal stimulus ensemble the amplitude-frequency representation provides us with a framework intuitively suited

to the analysis of the auditory system, where amplitude sensitivity and frequency selectivity are considered as important characteristic parameters (e.g. Kiang et al., 1965).

The analytic signal $\xi(t)$ associated with the real signal x(t) is constructed by adding an imaginary component $i\tilde{x}(t)$ to the given signal x(t), where i = 1/-1. As a consequence the analytic signal is a complex valued function of time, defined by

$$\xi(t) = x(t) + i\tilde{x}(t). \qquad (2.4)$$



Fig. 2a and b. Segment from a sequence of γ -tones a Signal x(t) as a function of time **b** Upper trace: instantaneous amplitude a(t). Lower trace instantaneous frequency $f(t) = \frac{1}{2\pi}\omega(t)$. Further explanation in fext

The real function $\tilde{x}(t)$, associated with x(t), is defined as the Cauchy principle value of the integral

$$\tilde{x}(t) = (\mathscr{H}x)(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} ds \frac{x(s)}{t-s}.$$
(2.5)

This linear integral transform, in fact the convolution of x(t) and the integral kernel $\frac{1}{\pi t}$ is known as the Hilbert transform (Thitchmarsh, 1937). There is a discrepancy in the sign of the definition of $\tilde{x}(t)$ between Gabor on the one hand and Ville, Thitchmarsh on the other hand. In (2.5) we conform to the latter.

The analytic signal (2.4) is also referred to as the pre-envelope (Dugundji, 1958). The function $\tilde{x}(t)$ is referred to as the signal in quadrature to x(t): x and \tilde{x} are mutually orthogonal.

The reciprocal relation holds

$$x(t) = (\mathscr{H}^{-1}\tilde{x})(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} ds \frac{\tilde{x}(s)}{t-s}$$
(2.6)

from which it follows that

$$\tilde{\tilde{x}}(t) = -x(t) \tag{2.7a}$$

or, equivalently,

$$(i\mathscr{H})^2 = 1, \tag{2.7b}$$

where 1 denotes the identity operator. For the harmonic oscillation

$$x(t) = A_0 \cos(\omega_0 t + \phi_0) \tag{2.8a}$$

the quadrature signal $\tilde{x}(t)$ is

$$\tilde{x}(t) = A_0 \sin(\omega_0 t + \phi_0) \tag{2.8b}$$

and the analytic signal is

$$\xi(t) = A_0 \exp i(\omega_0 t + \phi_0). \qquad (2.8c)$$

In the frequency domain the following relation holds

$$\hat{\xi}(\omega) = (1 + \operatorname{sgn}(\omega))\,\hat{x}(\omega)\,, \qquad (2.9)$$

where

$$sgn(\omega) = +1 \quad \omega > 0$$

$$0 \quad \omega = 0$$

$$-1 \quad \omega < 0 \quad (2.10)$$

and $\hat{x}(\omega)$ denotes the Fourier transform of x(t).

For negative frequencies the spectrum of the analytic signal $\xi(t)$ equals zero, for positive frequencies it has the same form as the spectrum of the original signal x(t).

Since the analytic signal $\xi(t)$ is a complex function it can be written in the form

$$\xi(t) = a(t) \exp i\phi(t) = \exp(\alpha(t) + i\phi(t)), \qquad (2.11)$$

where the instantaneous amplitude or temporal envelope is

$$a(t) = |\xi(t)| = (\xi^*(t)\,\xi(t))^{1/2}, \qquad (2.12)$$

the logarithm of the instantaneous amplitude is

$$\alpha(t) = \ln a(t) = \operatorname{Re} \ln \xi(t) = \frac{1}{2} \ln(\xi^*(t) \xi(t))$$
(2.13)

and the instantaneous phase is

$$\phi(t) = \arg \xi(t) = \operatorname{Im} \ln \xi(t) = \frac{1}{2i} \ln(\xi^{*-1}(t)\xi(t)). \quad (2.14)$$

In these equations $\xi^*(t)$ denotes the complex conjugate of $\xi(t)$. The *instantaneous angular frequency* $\omega(t)$ is defined as the time derivative of the instantaneous phase

$$\omega(t) = \frac{d}{dt}\phi(t) = \operatorname{Im}\frac{d}{dt}\ln\xi(t).$$
(2.15)

Analoguously the relative amplitude modulation $\beta(t)$ is given by the time derivative of the logarithm of the instantaneous amplitude

$$\beta(t) = \frac{d}{dt}\alpha(t) = \operatorname{Re}\frac{d}{dt}\ln\xi(t). \qquad (2.16)$$

The temporal intensity (density) I(t) is defined as

$$I(t) = a^{2}(t) = \xi^{*}(t)\,\xi(t), \qquad (2.17)$$

the spectral intensity (density) is defined as

 $J(\omega) = \hat{\xi}^*(\omega)\,\hat{\xi}(\omega)\,. \tag{2.18}$

In this terminology we conform to Middleton (1960).

Application of these definitions to the example of the harmonic oscillation (2.8) leads to

$$x(t) = A_0 \cos(\omega_0 t + \phi_0)$$

$$a(t) = A_0 \qquad \phi(t) = \omega_0 t + \phi_0$$

$$\alpha(t) = \ln A_0 \qquad \omega(t) = \omega_0 \qquad (2.19)$$

$$\beta(t) = 0$$

$$I(t) = A_0^2.$$

We conclude that for the harmonic oscillation the formal generalized definitions lead to expected results.

2.3. Parametric Representation of the Gamma-Tone Ensemble

In this section the notions as introduced in Sect. 2.2 will be applied to the γ -tone ensemble. This approach is simplified by a product theorem for Hilbert transforms (Tricomi, 1951; Bedrosian, 1963; Cain, 1973), which states

$$(\mathscr{H}g_1g_2)(t) = g_1(t)(\mathscr{H}g_2)(\tau)$$
(2.20)

provided that at least one of the following two conditions is fulfilled:

1. the spectral intensity of $g_2(t)$ is confined to a frequency range which is located entirely above that for $g_1(t)$,

2. both $g_1(t)$ and $g_2(t)$ are analytic signals.

Amplitude modulated tones can be represented by

$$x(t) = Am(t)\cos(\omega_0 t + \phi_0). \qquad (2.21)$$

If the carrier frequency ω_0 is larger than the maximum frequency present in m(t), condition 1 is fulfilled and as a consequence the analytic signal is given by

$$\xi(t) = Am(t) \exp i(\omega_0 t + \phi_0). \qquad (2.22)$$

If this is not strictly the case, but still ω_0 is large with respect to the bandwidth of m(t), (2.22) may be used as an approximation with a correction term which decreases in magnitude as ω_0 becomes higher (Lerner, 1960; Rubin and Di Franco, 1963). For narrowband AM-signals (2.22) approximates fairly well the analytic signal associated with (2.21). This result can be generalized to include slow phase- or frequency modulation (Lerner, 1960; Rihaczek, 1966).

From (2.22) it follows that the instantaneous amplitude a(t) and angular frequency $\omega(t)$ are given by

$$a(t) = Am(t)$$

$$\omega(t) = \omega_0.$$
(2.23)

The γ -tone ensemble used in the present study is described by

$$x(t) = \sum_{k=0}^{K-1} A_k U_T(t-kT) m(t-kT) \cos(\omega_k(t-kT) + \phi),$$
(2.24)

where K = total number of γ -tones in the stimulus ensemble; $A_k = \text{amplitude}$ factor of k-th γ -tone; $U_T(t)$ = rectangular time window; $U_T(t) = 1, 0 \le t \le T$; $U_T(t)$ = 0, elsewhere; m(t)...amplitude modulator as specified by (2.2); $\omega_k = \text{carrier}$ frequency of k-th γ -tone; T = timeinterval between onsets of γ -tones.

It can be verified that, to a good approximation, this ensemble can be considered as an ensemble of narrowband signals for all parameter values involved (cf. Sect. 4.1 and Appendix 1). As a consequence the *instantaneous amplitude* or *temporal envelope* a(t) is given approximately by

$$a(t) = \sum_{k=0}^{K-1} A_k U_T(t-kT) m(t-kT)$$
(2.25)

and the instantaneous frequency $\omega(t)$ can be approximated by

$$\omega(t) = \sum_{k=0}^{K-1} \omega_k U_T(t-kT).$$
(2.26)

These two parameters, both a function of time, are shown for a segment of the γ -tone ensemble in Fig. 2b. The formal definitions lead to results which agree to our intuitive notions. However, the intuitive definitions of instantaneous amplitude and frequency only apply to signals of a simple tonal character, whereas these formally generalized definitions of amplitude and frequency [(2.12) resp. (2.15)] are applicable to any type of signal.

The parametric analysis described in this paper is based on the concepts of instantaneous frequency and instantaneous amplitude, defined as functionals of the analytic signal. The Hilbert transform, and as a consequence the analytic signal and all functions derived from it, are functionals of the real signal x(t), where both the past and future of x(t) have to be evaluated [see (2.5)]. Especially the definition of "instantaneous frequency", because of this "non-causal" nature, has given rise to considerable debate in the literature (e.g. Mandel, 1974; Gupta, 1975, and references therein). These particular definitions have been used here because of their mathematical elegance and because they have proven to be useful in signal and system theory, both linear and nonlinear (e.g. Rihaczek, 1968; Deutsch, 1969). The notion of instantaneous frequency $\omega(t)$ on one hand and Fourier-frequency ω on the other hand should not be confused. The instantaneous frequency $\omega(t)$ is a time dependent functional of a signal, whereas the spectral variable ω is the independent variable in the domain, complementary to the temporal domain, associated with it by means of the Fourier transform.

The natural environment in which an animal lives, as far as functionally related to the animal, is defined as its habitat or biotope (Dahl, 1921). It encompasses both the biotic and abiotic environment. The Sensory Biotope may be defined as the total of sensory aspects of the biotope. It consists of several components, each one corresponding to a specific sensory modality of the animal under investigation. The acoustical component is of special interest here: the Acoustic Biotope (Aertsen et al., 1979; Smolders et al., 1979). The Acoustic Biotope contains different kinds of sounds, each with an associated probability of occurrence. In general it will not be possible to give an explicit mathematical expression for the probability density of the Acoustic Biotope, as can be done for a technically based stimulus ensemble like Gaussian noise. For any particular realization of the Acoustic Biotope, however, it is possible to evaluate numerically the relative frequency of occurrence of the various sounds. Such a realization may be considered as a wide selection of natural sounds, the probabilistic structure of which more or less truly portrays the Acoustic Biotope. The Acoustic Biotope thus is considered as a probabilistic generator of stimulus ensembles as particular realizations, comparable to the Gaussian probability density function and the particular noise signals generated by it.

For the present investigation a realization of a wide stimulus ensemble, inspired by ethological and ecological observations was made. It consisted of a selection of sounds, recorded from the natural environment of the animal under investigation, i.e. the grassfrog. The realization was composed according to the scheme as shown in Diagram 1.

It consisted of 32 different sound segments, each one with a duration of about 8s. The figures in Diagram 1 specify the numbers of sounds taken from different categories. Next to the "A-, B-, and C-calls" from the *male grassfrog* (typology according to Van Gelder et al., 1978), the stimulus ensemble contained such various sounds as the vocalizations of several other anuran species [e.g. the green frog (Rana esculenta L.), the moorfrog (Rana arvalis Nilsson), the European treefrog (Hyla arborea L.) and the common toad (Bufo bufo L.)], sounds from predators (e.g. the heron, the gull), prey (the fly, the cricket) and other animals (e.g. the crow) as well as sounds from abiotic origin (e.g. water, wind, rain). The 32 segments of sound were concatenated in a random order on analog tape, thus constituting a stimulus ensemble lasting about 4 min, its bandwidth being 5 kHz. In the following this realization will be denoted briefly as the "acoustic biotope". Different sounds from this stimulus ensemble are shown in Fig. 3.



Diagram 1. Construction scheme for the realization of a stimulus ensemble inspired by the Acoustic Biotope of the grassfrog (Rana temporaria L.)



Fig. 3a–f. Different segments of sound, 1 s each, from the natural stimulus ensemble, the acoustic biotope, of the grassfrog (Rana temporaria L). Vertical scale is identical in all cases a Initial part of a C-call of the male grassfrog (total duration approx 19s) b Chorus of grassfrog vocalizations c Vocalizations of the moorfrog (Rana arvalis Nilsson) d Vocalizations of gulls and chorus of green frogs (Rana esculenta L.) against noisy background (wind) e Vocalizations of heron and other birds f Sound of a rippling brook

The construction scheme, shown in Diagram 1, takes into account several characteristics of the Acoustic Biotope, albeit in a rather rude way. Ethological and ecological considerations have led to the various categories as they appear in the scheme (Smolders et al., 1979). The probabilistic structure of the Acoustic Biotope is reflected in the numbers of sounds taken from each category. In the present context, however, special attention should be given to the physical structure of the signals in the stimulus ensemble, more notably their spectro-temporal intensity density.

Formally one might proceed along the lines as sketched in Sect. 2, resulting in a parametric description of natural sounds in terms of amplitude and frequency as functions of time. This approach, however, is most fruitful and at the same time agrees to our intuitive notions in the case of narrowband signals. In the case of a wideband signal the parametric approach lacks the intuitive appeal and results in parameters, the behaviour of which as functions of time comes close in complexity to the behaviour of the signal itself. A more general spectro-temporal description of signals is given by the Dynamic Power Spectrum (e.g. Flanagan, 1965) or spectrogram (e.g. Koenig et al., 1946; Potter et al., 1947). The dynamic power spectrum P(f, t) of a signal gives the intensity as a function of time for the different frequency components of the signal. It can be measured by means of a set of bandpass filters and envelope detectors. Appendix 2 gives a description of the mathematical background and the principles of operation of an instrument, built for this purpose, the Dynamic Spectrum Analyser (DSA). In the present investigation the filters used were 1/3 octave filters. As a consequence the dynamic spectrum is a set of functions of time representing the instantaneous intensity of the signal in a set of neighbouring 1/3 octave bands. Figure 4 may serve as an example: it shows the dynamic power spectrum of a *B*-call of the grassfrog as measured by the DSA (cf. Fig. 1).

The DSA also offers the possibility of triggered averaging of dynamic power spectra. The temporal relation between the trigger and the signal of interest can be varied between completely post- and completely pre-trigger position. This possibility of determining the average spectro-temporal characteristics of an ensemble of signals will be used in the investigation of the stimulus-event relation of auditory neurons (cf. the accompanying paper, Part II, Aertsen et al., 1980 and a forthcoming paper, Part III).

The parametric description of signals (cf. Sect. 2) and the description by means of the dynamic power spectrum can be formally connected by introduction of a general *Complex Spectro-Temporal Intensity Density* (CoSTID) of signals, defined on the analytic signal (Rihaczek, 1968; Johannesma and Aertsen, 1979).

4. Practical Realization of the Stimulus Ensembles

4.1. The Tonal Stimulus Ensemble

The tonal stimulus ensemble consisted of sequences of γ -tones with constant duration parameter β and form parameter γ [see (2.3)] and different values of frequency ω and amplitude factor A. The phase ϕ was either set to $-\pi/2$ or was not controlled. The frequency range was 4 octaves: either from 125 2000 Hz or from 250-4000 Hz, depending on the frequency characteristics of the neuron; values in this range were taken at 255 postitions, equidistant on log frequency scale (64 intervals per octave). The amplitude factor resolution was 127 values at equal intervals ΔA (on linear scale) and minimum value ΔA , thus covering a range of about 42 dB, apart from the dynamic range within the individual y-tones. The different values of ω and A were presented in a pseudo-random order to avoid possible systematic effects (e.g. hysteresis) of monotonous variation of these parameters (details of the pseudo-random sequences are described in the companion paper, Part II). The complete sequence consists of $255 \times 127 = 32,385$ y-tones. The interval between onsets of successive y-tones was 16 ms. Although a y-tone as function of time theoretically stretches to plus infinity, the modulator m(t), with β and γ as specified in (2.3), has decreased to less than 0.5% of its maximum value after 16 ms. This enables a truncation of the y-tone after 16 ms without much



Fig. 4. Different representations of the Dynamic Power Spectrum of the "B-call" of the male grassfrog (cf. Fig. 1), determined by the Dynamic Spectrum Analyser (DSA), using a set of 1/3 octave filters

distortion. The result is a sequence of γ -tones with at any time only one tonal component present (cf. Fig. 2).

The tonal stimuli were generated by a Programmable Stimulus Generator (PSG), containing the following main components: the signal source was a programmable frequency synthesizer (Rockland 5100), the output of which was amplitude modulated by 2 serial 12-bits multiplying DA-converters (Hybrid Systems DAC 316-12); stimulus control was performed by a PDP 11/10. The generated waveform was lowpass filtered (2 cascaded -24 dB/oct Bessel filters, each -3 dB at 5 kHz) before it was amplified and presented to the animal.

4.2. The Natural Stimulus Ensemble

Most of the recordings of frog and toad vocalizations were made in the vicinity of Nijmegen and were obtained from the Department of Animal Ecology at the University of Nijmegen, the Netherlands. Details of the recording techniques were described by Van Gelder et al. (1978). The remainder of the natural sounds constituting the stimulus ensemble, was also recorded in the vicinity of Nijmegen (recorder Uher 4200, tape speed 19 cm/s, microphones Sennheiser MD 211 N). A stimulus tape containing a number of these sounds was composed according to the scheme in Diagram 1. This tape was played back to the animal during the experiment from an AM-recorder (Otari MX 5050 QXH, tape speed 38 cm/s). Before amplification and presentation to the grassfrog the signal was lowpass filtered according to the same specifications as the tonal stimulus ensemble.

4.3. Application

The stimulus ensembles as described here, together with the signal representations associated with them, have been used in an investigation into the spectro-temporal properties of the stimulus-event relation of single auditory neurons in the auditory midbrain area of the grassfrog. The spectro-temporal intensity density of the particular stimuli which precede the occurrence of action potentials, the *Pre-Event Stimulus Ensemble* (PESE), is compared to the intensity density of the overall stimulus ensemble. The result of this comparison leads to the "spectro-temporal receptive field" of an auditory neuron. The results of this investigation, together with a discussion on the feasibility of the approach as given here, are described in separate papers. The companion paper (Part II, Aertsen et al., 1980) is concerned with the analysis of neural responses to the tonal stimulus ensemble. The response to the natural stimulus ensemble will be the subject of a forthcoming paper (Part III).

Appendix 1

Some Properties of the γ -Tones

A1.1. The Amplitude Modulation. The amplitude modulator of the γ -tone is given by [see (2.2)]

$$m(t) = c \left(\frac{t}{\beta}\right)^{\gamma - 1} \exp\left(-\frac{t}{\beta}\right) \quad t \ge 0 \quad \beta > 0$$

$$= 0 \qquad t < 0 \quad \gamma \ge 1.$$
(A1.1)

In this expression c is a normalization constant, to give $m^{2}(t)$ unit area:

$$c = \left(\frac{2^{2\gamma - 1}}{\beta \Gamma(2\gamma - 1)}\right)^{1/2},$$
 (A1.2)

where $\Gamma(x)$ denotes the Gamma function (Abramowitz and Stegun, 1972).

The parameter β influences the *duration* of m(t), it can be considered as a time scaling factor. The parameter γ controls the *form* of m(t). The larger γ , the more symmetrical m(t): for $\gamma = 1$ m(t) describes pure exponential decay, for $\gamma \rightarrow \infty$ m(t) approaches a Gaussian form. The temporal intensity $I(t) = m^2(t)$ is represented R

by the same function, with $\beta_I = \frac{\beta}{2}$ and $\gamma_I = 2\gamma - 1$.

The spectrum $\hat{m}(\omega)$ of m(t) is

$$\hat{m}(\omega) = \frac{c\beta\Gamma(\gamma)}{(1+i\beta\omega)^{\gamma}}.$$
(A1.3)

The spectral intensity density $J(\omega)$ is given by

$$J(\omega) = \hat{m}^*(\omega)\hat{m}(\omega) = \frac{c^2\beta^2\Gamma(\gamma)}{(1+\beta^2\omega^2)^{\gamma}}.$$
 (A1.4)

Several characteristics of the modulator m(t) can now be calculated, e.g. the maximum value and the first and second order moment, both in the time and frequency domain (Gabor, 1946; Grashuis, 1974). The maximum value is:

œ

in the time domain: $m_{\max} = c(\gamma - 1)^{\gamma - 1} e^{-(\gamma - 1)}$ at $t = \beta(\gamma - 1)$

in the frequency domain: $|\hat{m}|_{\max} = c\beta\Gamma(\gamma)$ at $\omega = 0$.

The mean value μ_t of the "epoch" t of m(t) is given by

$$\mu_{\rm r} = \frac{\int\limits_{-\infty}^{\infty} dt m(t) t m(t)}{\int\limits_{-\infty}^{\infty} dt m(t) m(t)} = \frac{\beta}{2} (2\gamma - 1). \tag{A1.6}$$

The standard deviation σ_t , which describes the temporal width or duration of m(t), is given by

$$\sigma_t = \left(\frac{\int_{-\infty}^{\infty} dt \, m(t) \, (t-\mu_t)^2 \, m(t)}{\int_{-\infty}^{\infty} dt \, m(t) \, m(t)}\right)^{1/2} = \frac{\beta}{2} (2\gamma - 1)^{1/2} \,. \quad (A1.7)$$

Similar definitions can be given in the frequency domain. The mean value μ_{ω} of the angular frequency is given by

$$\mu_{\omega} = \frac{\int_{-\infty}^{\infty} d\omega \hat{m}^{*}(\omega) \, \omega \hat{m}(\omega)}{\int_{-\infty}^{\infty} d\omega \, \hat{m}^{*}(\omega) \, \hat{m}(\omega)} = 0.$$
(A1.8)

The standard deviation σ_{ω} , describing the spectral width is given by

$$\sigma_{\omega} = \left(\frac{\int\limits_{-\infty}^{\infty} d\omega \, \hat{m}^{*}(\omega) \, (\omega - \mu_{\omega})^{2} \, \hat{m}(\omega)}{\int\limits_{-\infty}^{\infty} d\omega \, \hat{m}^{*}(\omega) \, \hat{m}(\omega)}\right)^{1/2} = \frac{(2\gamma - 3)^{-1/2}}{\beta}$$

$$(\gamma > \frac{3}{2})$$
(A1.9)

or, equivalently,

$$\sigma_f = \frac{\sigma_\omega}{2\pi} = \frac{(2\gamma - 3)^{-1/2}}{2\pi\beta}.$$
 (A1.10)

Combining the measures for width in time and frequency domain we obtain the spectro-temporal resolution of the signal or its "uncertainty product" (Gabor, 1946)

$$\Delta = \sigma_{\omega}\sigma_{t} = \frac{1}{2} \left(\frac{2\gamma - 1}{2\gamma - 3}\right)^{1/2} \qquad (\gamma > \frac{3}{2})$$
(A1.11)

which, for a Gaussian waveform $(\gamma \rightarrow \infty)$, reaches a minimum of $\frac{1}{2}$. The parameters of the γ -modulator



Fig. 5. Functional diagram of Dynamic Spectrum Analyser (DSA). The various control functions have been omitted for reasons of clarity

1

used in the present investigation are $\beta = 1.45$ ms and $\gamma = 3$ [see (2.3)]. This gives numerical values for the maximum:

- in the time domain at t = 2.90 ms

- in the frequency domain at $\omega = 0$ rad/s (A1.12)

for the moments:

- in the time domain:
$$\mu_t \simeq 3.63 \text{ ms},$$

 $\sigma_t \simeq 1.62 \text{ ms}$
- in the frequency domain: $\mu_{\omega} = 0 \text{ rad/s},$
 $\sigma_{\omega} \simeq 398.2 \text{ rad/s}$
or: $\mu_f = 0 \text{ Hz},$
 $\sigma_f \simeq 63.4 \text{ Hz}$ (A1.13)

and for the uncertainty product:

$$\Delta \simeq 0.645, \tag{A1.14}$$

which is only slightly above the theoretical limit of $\frac{1}{2}$.

A1.2. The y-tone. The y-tone e(t) is defined by [see (2.1)]

$$e(t) = Am(t)\cos(\omega_{y}t + \phi)$$
(A1.15)

with the spectrum

•••

$$\hat{e}(\omega) = \frac{1}{2}A(\hat{m}(\omega - \omega_{\gamma}) + \hat{m}(\omega + \omega_{\gamma})).$$
(A1.16)

The Hilbert transform of e(t) is

(A1.17) $e(t) = Am(t)\sin(\omega_v t + \phi)$

provided that the highest frequency component in m(t)is below ω_{γ} [see (2.20)]. This condition is not fulfilled by the spectrum $\hat{m}(\omega)$, however it is useful to regard (A1.17) as an approximation with an error determined by the spectral tail of m(t) (Rubin and Di Franco, 1963). The integral

$$F(v) = \frac{\int\limits_{-v}^{v} d\omega \,\hat{m}^*(\omega) \,\hat{m}(\omega)}{\int\limits_{-\infty}^{\infty} d\omega \,\hat{m}^*(\omega) \,\hat{m}(\omega)}$$
(A1.18)

gives the *fraction* of spectral intensity in the interval (-v, v). Substitution of (A1.3), using y=3, leads to (Gradshteyn and Ryzhik, 1965)

$$F(v) = \frac{1}{2\pi} c^2 \beta \Gamma^2(\gamma) \left(\frac{u}{4(1+u^2)^2} + \frac{3u}{8(1+u^2)} + \frac{3}{8} \operatorname{arctg} u \right)_{u=-\beta_v}^{u=-\beta_v}.$$
(A1.19)

Numerical evaluation of F(v) gives

$$F(\sigma_{\omega}) = 0.75$$

$$F(2\sigma_{\omega}) = 0.95$$

$$F(3\sigma_{\omega}) = 0.99.$$

(A1.20)

From the results (A1.13) and (A1.20) we conclude that for the y-tones as used in the present investigation, with f_{y} in the range of 125 Hz to 4000 Hz, the analytic signal representation is, to a fair approximation, given by

$$\xi(t) \simeq Am(t) \exp i(\omega_{y}t + \phi) \tag{A1.21}$$

which gives, for the instantaneous amplitude or temporal envelope

$$a(t) \simeq Am(t) \tag{A1.22}$$

and for the instantaneous frequency

$$\omega(t) \simeq \omega_{y}. \tag{A1.23}$$

Obviously the approximations (A1.21-23) are better for higher carrier frequency ω_{y} .

Appendix 2

Dynamic Spectrum Analyser (DSA)

A2.1. Spectro-Temporal Analysis. A spectro-temporal analysis of sound can be made by measuring the Dynamic Power Spectrum P(f,t), also referred to as short-time power spectrum (Fano, 1950; Schroeder and Atal, 1962; Flanagan, 1965; Gambardella, 1971). The dynamic power spectral analysis has been implemented in hardware, thereby operating in real time, on the principle of simultaneous, multiple bandpass filtering and envelope detection (Hoeks and Krijt, 1976). Figure 5 shows a schematic diagram of the Dynamic Spectrum Analyser (DSA).



Fig. 6. Dynamic Power Spectrum of an impulse, as determined by the DSA, using a set of 1/3 octave filters. Note the frequency dependent time shift in the spectrogram

The frequency analysis is performed by a set of 1/3 octave filters (HP8065A), each one followed by a Hilbert transform section, producing two signals with a phase difference of $\pi/2$ rad. These two signals correspond to the real and imaginary part of the analytic bandfiltered signal [cf. (2.4)]. Summation of the squares of both signals gives the temporal intensity of the bandfiltered signal [cf. (2.17)]. The result of this section is P(f,t) at $f = \mu_{f,l}$ (l=1, L), where $\mu_{f,l}$ is the center frequency of the l-th bandpass filter. An example is shown in Fig. 4: the dynamic power spectrum of a "B-call" of the grassfrog.

The filter section consists of 44 1/3 octave filters between 2 Hz and 50 kHz, with center frequencies equidistant on log scale. From this range a frequency interval of maximally 8 octaves can be selected. The overall behaviour of each filter is that of a 6-th order Chebychev filter with a ripple in the passband region of less than 0.5 dB. For each of the filters the effective bandwidth σ_f [cf. (A1.10)] equals about 0.082 μ_f , the effective duration σ_i of the impulse response [cf. (A1.7)] equals about $1.52/\mu_f$, as a consequence the uncertainty product $\Delta = 2\pi\sigma_f \sigma_f$ [cf. (A1.11)] for each of the filters is 0.78. The attenuation outside the passband region is 20 dB at 0.79 μ_f and 1.26 μ_f , 50 dB at $\mu_f/2$ and $2\mu_f$ and amounts to 70 dB at $\mu_f/4$ and $4\mu_f$. Due to the phase characteristics of the bandpass filters and the Hilbert transformers the filtered signals have a delay, depending on the filter center frequency μ_{f} . This effect is demonstrated by the dynamic power spectrum of an impulse, shown in Fig. 6. The delay caused by the bandpass filter is the group delay μ_t of the filter impulse response [cf. (A1.6)], where for each of the filters μ_t is inversely proportional to the center frequency: $\mu_{t}\mu_{f} \simeq 3.45$. The frequency dependent delay of the Hilbert transformers added to μ_t gives a total delay μ'_t for which it holds $\mu'_t \mu_f \simeq 3.8$. Obviously this delay

should be taken into account in the interpretation of results, especially the estimation of latencies etc.

The differing absolute bandwidth of the 1/3 octave filters causes the spectral intensity density to be transformed: a 1/f-spectrum turns "white" and a white spectrum would look like an "*f*-spectrum". An intensity correction option, consisting of filter-frequency dependent amplifiers combined with the Hilbert transformers, is available. Further spectral manipulation is possible by means of the variable gain of each bandfilter, covering a range of 40 dB.

Both the set of "real" and the set of "imaginary" filtered signals from the Hilbert transformers are multiplexed, using 2 simultaneously operating analog time multiplexers. Both multiplexers are followed by 9-bits AD-converters (maximum rate 100 ksamples/s). The differential time jitter of sampling moments is less than 10 ns. Multiplexing and sampling can be performed in 2 modes: 1. sequential sampling in which the center frequency of the highest filter selected determines the sampling rate (Shannon), which is then used to sample all filter outputs and 2. proportional sampling in which the sampling rate for the 3 filters in each separate octave is determined by the upper filter frequency from that octave [compare the "logons" introduced by Gabor (1946)]. In the latter sampling mode the total number of samples is strongly reduced. The outputs from the Hilbert sections are AC-coupled before the multiplexer to eliminate possible DC-offsets, causing undesired cross products in the squarers. Up to this level of analysis the dynamic range is 54 dB.

The operations of squaring and summation are performed digitally, using 18 bits resolution, and result in 18-bits digitized temporal intensities of the bandfiltered signals.

A2.2. Averaging of Dynamic Power Spectra. In the DSA the dynamic power spectra of an ensemble of signals can be averaged. In the present investigation the ensemble of interest is the ensemble of signals *prior* to the neural events (the PESE, see the companion paper, Part II).

In order to have access to the dynamic power spectrum prior to an event, the data from the "sum of squares"-units are continuously stored into a delay line memory (4096 18-bits words). The maximum temporal interval to be contained in this memory is determined by the selection of filters and the sampling mode. If necessary this interval can be reduced by powers of 2. Upon a trigger the contents of the delay-line memory is retrieved and averaged with the contents of the main memory (4096 18-bits words). Data transfer time is 7.0–7.5 μ s/word. The retrieval of data from the delay line memory is performed asynchronously with the storage of data. Triggering can be done in internal,

manual and external mode, the latter both on negative and positive slope of the event. The temporal relation between the trigger and the dynamic power spectrum to be transferred is set by the temporal trigger mode. The 5 possibilities are 0%, 25%, 50%, 75%, and 100% pre-event dynamic spectrum, combined with the complementary fraction of the post-event dynamic spectrum During transfer of data to main memory triggering is inhibited. The resulting "dead time" equals the maximum of 1 the post-trigger time interval and 2. the time taken to transfer a complete dynamic power spectrum to the main memory (about 29 5 ms for full memory). This "dead time" may result in a missing of triggers and a consecutive departure from the true ensemble average to an amount determined by the rate of (neural) events and the total number of samples per dynamic spectrum.

Averaging is performed on a running base: the weight factor of individual spectra is continuously adjusted to the number of averages already executed Two running average algorithms are implemented. In the present investigation we have used throughout

$$q_n = \frac{q_{n-1}}{2^{k_n - k_{n-1}}} + \frac{p_n}{2^{k_n}},$$
 (A2 1)

where q_n , q_{n-1} =running average of *n* resp n-1 dynamic power spectra, p_n =dynamic power spectrum of *n*-th ensemble element, k_n =integer number, with $2^{k_n-1} < n \le 2^{k_n}$

In this algorithm q_n is a true average, however, it has correct scaling only for $n=2^k$, k=0, 1, 2, . The other algorithm

$$q'_{n} = q'_{n-1} + \frac{p_{n} - p_{n-1}}{2^{k_{n}}}$$
(A2 2)

has correct amplitude scaling but does not provide a true average. Averaged data from the main memory can be displayed on an oscilloscope. Intensity as function of time and filter frequency using a quasi-3D display and/or intensity modulation Read-out of the main memory is also possible to XY-recorder and computer. In the latter case a restricted portion of the main memory, containing the various selected control functions and preset values of the DSA, is transferred as well.

Acknowledgements This investigation forms part of the research program of the Workgroup Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (The Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (Z WO) The authors wish to thank Jos Eggermont and Ton Vendrik for helpful discussions on this manuscript Pieter Flst, Arnold Hoeks, Hans Krijt, and Jan Lammers made essential contributions to the development of the Dynamic Spectrum Analyser and actually realized it Tape recordings of vocalizations of various species of anurans were kindly provided by Jan van Gelder (Department of Animal Ecology, University of Nijmegen)

References

- Abramowitz, M, Stegun, IA (eds) Handbook of mathematical functions New York Dover 1972
- Aertsen, A M H J, Smolders, J W T, Johannesma, P I M Neural representation of the Acoustic Biotope on the existence of stimulus-event relations for sensory neurons Biol Cybernetics 32, 175-185 (1979)
- Aertsen, A M H J, Johannesma, P I M, Hermes, D J Spectrotemporal receptive fields of auditory neurons in the grassfrog II Analysis of the stimulus-event relation for tonal stimuli Biol Cybernetics 38, 235 248 (1980)
- Bedrosian, E A product theorem for Hilbert transforms Proc IEEE 51, 868 869 (1963)
- Boer, E de On the principle of specific coding J Dyn Syst Meas Control 95G, 265 273 (1973)
- Boer, E de, Jongh, H R de On cochlear encoding potentialities and limitations of the reverse correlation technique J Acoust Soc Am 63, 115-135 (1978)
- Brzoska, J., Walkowiak, W., Schneider, H. Acoustic communication in the grassfrog (Rana temporaria L) calls, auditory thresholds, and behavioral responses J. Comp. Physiol. 118, 173-186 (1977).
- Bullock, T H (ed) Recognition of complex acoustic signals (Life Sciences Res Rep, Vol 5) Berlin, Heidelberg, New York Springer 1977
- Cain, G D Hilbert transform relations for products Proc IEEE 61, 663-664 (1973)
- Capranica, R R 1 he evoked vocal response of the bullfrog a study of communication by sound Cambridge, MA MIT Press 1965
- Cox, D R, Lewis P A W The statistical analysis of series of events London Methuen 1966
- Dahl, F Grundlagen einer okologischen Tiergeographie, Teil I Jena Gustav Fischer 1921
- Dugundji, J Envelopes and pre-envelopes of real waveforms IRE Trans Inf Theory IT-4, 53-57 (1958)
- Fano, R M Short-time autocorrelation functions and power spectra J Acoust Soc Am 22, 546 550 (1950)
- Flanagan, J L Speech analysis, synthesis, and perception Berlin, Heidelberg, New York Springer 1965
- Gabor, D Theory of communication J IEE (London) Part III 93, 429–457 (1946)
- Gambardella, G A contribution to the theory of short-time spectral analysis with nonuniform bandwidth filters IEEE Trans Circuit Theory CT-18, 455-460 (1971)
- Gelder, JJ van, Evers, PMG, Maagnus, GJM Calling and associated behaviour of the common frog, Rana temporaria, during breeding activity J Anim Ecol 47, 667 676 (1978)
- Gerhardt, H C The significance of some spectral features in mating call recognition in the green treefrog (Hyla cinerea) J Exp Biol 61, 229 241 (1974)
- Gerhardt, H C Mating call recognition in the green treefrog (Hyla cinerea) the significance of some fine-temporal properties J Exp Biol **74**, 59 73 (1978)
- Gradshteyn, IS, Ryzhik, IM Table of integrals, series, and products New York, London Academic Press 1965
- Grashuis, J L The pre-event stimulus ensemble an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons Ph D Thesis, Nijmegen 1974
- Gupta, M S Definition of instantaneous frequency and frequency measurability Am J Phys 43, 1087-1088 (1975)
- Hoeks, APG, Krijt, HJ Hardware processing of non-stationary signals Internal report, Dept of Medical Physics and Biophysics, University of Nijmegen 1976
- Johannesma, P I M The pre-response stimulus ensemble of neurons in the cochlear nucleus In Proc of the IPO Symp on Hearing Theory, pp 58-69 Cardozo, B L (ed) Eindhoven 1972

- Johannesma, PIM, Aertsen, AMHJ Neural image of sound in the grassfrog In Hearing mechanisms and speech, pp 79–86 Creutzfeld, O, Schleich, H, Schreiner, Chr (eds) Exp Brain Res [Suppl 11] (1979)
- Kiang, N Y -S, Watanabe, T, Thomas, F C, Clark, L F Discharge patterns of single fibers in the cat's auditory nerve Cambridge, MA MIT Press 1965
- Koenig, W, Dunn, HK, Lacy, LY The sound spectrograph J Acoust Soc Am 18, 19–49 (1946)
- Lerner, R M A matched filter detection system for complicated Doppler shifted signals IRE Trans Inf Theory IT-6, 373-385 (1960)
- Mandel, L Interpretation of instantaneous frequencies Am J Phys 42, 840–846 (1974)
- Middleton, D An introduction to statistical communication theory New York, Toronto, London McGraw-Hill 1960
- Potter, R K, Kopp, G A, Green, H C Visible speech New York Van Nostrand 1947
- Rihaczek, AW Hilbert transforms and the complex representation of real signals Proc IEEE 54, 434-435 (1966)
- Rihaczek, AW Signal energy distribution in time and frequency IEEE Trans Inf Theory IT-14, 369-374 (1968)
- Rubin, WL, DiFranco, JV Analytic representation of wideband radio frequency signals J Franklin Inst 275, 197-204 (1963)
- Schneider, H Die Paarungsruße einheimischer Ranidae (Anura, Amphibia) Bonn Zool Beitr 24, 51-61 (1973)

- Schroeder, M R, Atal B S Generalized short-time power spectra and auto-correlation functions J Acoust Soc. Am 34, 1679-1683 (1962)
- Smolders, JWT, Acrtsen, AMHJ, Johannesma, PIM Neural representation of the Acoustic Biotope a comparison of the response of auditory neurons to tonal and natural stimuli in the cat Biol Cybernetics 35, 11–20 (1979)
- Thitchmarsh, EC Introduction to the theory of Fourier integrals London Oxford University Press 1937
- Tricomi, FG On the finite Hilbert transformation Q J Math Oxford 2, 199-211 (1951)
- Ville, J Theorie et applications de la notion de signal analytique Cables Transm 2, 61-74 (1948)
- Worden, FG, Galambos, R (eds) Auditory processing of biologically significant sounds (Neurosci Res Prog Bull, Vol 10) Brookline, MA Neurosciences Research Program 1972

Received May 11, 1980

A M H J Aertsen Department of Medical Physics and Biophysics University of Nijmegen Geert Grooteplein Noord 21 NL-6525 EZ Nijmegen The Netherlands CHAPTER 5

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

II. Analysis of the Stimulus-Event Relation for Tonal Stimuli

Aertsen, A.M.H.J., Johannesma, P.I.M. and Hermes, D.J. Biol. Cybernetics <u>38</u>, 235-248 (1980)

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

II. Analysis of the Stimulus-Event Relation for Tonal Stimuli

A. M. H. J. Aertsen, P. I. M. Johannesma, and D. J. Hermes

Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Abstract. The nature of the stimulus-response relation for single auditory neurons is reflected in the properties of the Pre-Event Stimulus Ensemble: the ensemble of stimuli, preceding the occurrence of an action potential (neural event). This paper describes methods to analyse the spectro-temporal properties of this ensemble. These methods are based on the analytic signal representation of acoustic signals and functionals derived from it: the instantaneous amplitude and instantaneous frequency and the dynamic power spectrum. The procedures have been applied to a number of extra-cellular single unit recordings from the grassfrog, Rana temporaria L., recorded during presentation of an ensemble of tonal stimuli. The outcome of this analysis describes the spectro-temporal receptive field of the neuron under the present stimulus conditions. The procedure, based on the dynamic power spectrum is applicable to an *arbitrary* stimulus ensemble, thus allowing a comparison of the spectrotemporal receptive fields for different types of stimuli.

1. Introduction

The investigation of stimulus-response relations of auditory neurons, in general, is performed on the base of a low-dimensional type of analysis. One aspect of the neural activity (e.g. firing rate) is related to one or more parameters (e.g. frequency or intensity) of the stimulus. Experiments with tone burst stimuli play a dominant role. Well known examples of this approach are the *tuning curve* and the PST-*histogram* (e.g. Kiang et al., 1965), which characterize the *spectral* resp. the *temporal* behaviour of the neuron. The question whether the response of a neuron to different types of stimuli (e.g. noise, natural sounds) can be understood on the base of these characteristics has been answered positively for peripheral neurons (Evans, 1975). For more central neurons the picture evidently is more complex (Worden and Galambos, 1972; Bullock, 1977; Smolders et al., 1979). Frequency selectivity on itself, in general, proves to be an inadequate characteristic of neural function. The same holds for simple temporal characteristics.

A commonly used way to characterize sound, especially more complex sound, is the spectrogram or sonogram (e.g. Potter et al., 1947; Singleton and Poulter, 1967): the intensity density of the signal is given as a function of both *frequency* and *time*. Inspired by this observation procedures were developed to analyse the spectro-temporal selectivity of auditory neurons for a variety of acoustic stimuli. The present paper is concerned with the methodical aspects of this approach. The spectro-temporal intensity density of the particular stimuli which precede the occurrence of action potentials is evaluated against the intensity density of the overall stimulus ensemble. This analysis, basically a crosscorrelation-type of approach, leads to the "spectro-temporal receptive field" of an auditory neuron. The approach is based on the spectrotemporal characterization of sound, described in the accompanying paper (Part I, Aertsen and Johannesma, 1980).

Spectro-temporal receptive fields upon presentation of an ensemble of tonal stimuli were determined for single neurons in the auditory midbrain area (torus semicircularis) of the grassfrog (*Rana temporaria* L.). Findings in several anuran species (e.g. Capranica, 1965; Gerhardt, 1974, 1978; Van Gelder et al., 1978) indicate that both spectral and temporal aspects of species-specific vocalizations are behaviourally significant. These findings suggest that the combined spectro-temporal approach may provide information about the neural basis of behaviour in the natural acoustic environment. Two types of spectro-temporal analysis have been used: a *parametric* description (*instantaneous frequency* and *amplitude*) and a description by means of the *dynamic power spectrum*. The first one is especially suited for narrowband stimuli, the second one is more general and can be used for an arbitrary stimulus ensemble. The latter one will also be used in the investigation of spectro-temporal receptive fields upon stimulation with natural stimuli. This will be the subject of a forthcoming paper (Part III).

2. Methods

Preparation and Recording

Extracellular single unit recordings were made in the torus semicircularis of adult grassfrogs from Ireland Before the preparation the frog was anaesthetized in a 01% solution of MS-222 (ethyl-*m*-aminobenzoate methanesulphonate). After the disappearance of the hindleg reflex upon pinching between the toes the concentration was lowered to 005%. After removal of a portion of the skin on the head, a hole was drilled into the region of the skull overlying the tectum opticum. The dura mater was removed. The wound was covered with paraffin oil to prevent it from drying

During the experiments the frog, wrapped up in a thin layer of moist cotton gauze, was fixated into a small holder and was moistened continuously The purpose of the latter was to prevent the frog from drying and to maintain the skin respiration. The moistening fluid was a 0.025-0.05% solution of MS-222, this in order to maintain anaesthesia at a level at which corneal reflex returned but no other movements occurred.

Experiments were carried out at a temperature of 14–16 °C. A microelectrode was advanced in the brain by a remotely controlled hydraulic microdrive (Trent Wells) Two types of microelectrodes were used teflon coated, Insl-X insulated tungsten electrodes (typical impedance 1–3 M Ω at 1000 Hz) and glass micropipettes, filled with woods metal-indium alloy, with platinum tip (Dowben and Rose, 1953) (typical impedance 300–800 k Ω at 1000 Hz) Recordings from both types of electrodes did not show strikingly different results. Single unit recording times ranged from 5 min up to 2 h

Stimulus Presentation and Data Registration

Experiments were performed with the frog in an electrically shielded acoustic chamber (IAC 1202A) The stimulus ensemble consisted of sequences of short tonal elements, the Gamma-tones (γ -tones). The characteristics of this tonal ensemble and the technical procedures of generation are described in detail in the companion paper (Part I, Aertsen and Johannesma, 1980). A segment of a sequence of γ -tones is shown in Fig 1

The stimulus ensemble was generated by a Programmable Stimulus Generator (PSG), controlled by a PDP 11/10 The signal was amplified (Luxman 507X) and the sound was delivered by two electrodynamic microphones (Sennheiser MD 211N), which were coupled to the frog's tympanic membranes by means of closed cavities, filled with damping material to reduce resonances. The sound pressure level was measured using half inch condensor microphones (Bruel and Kjaer 4134), connected to the acoustic couplers. The frequency response of the sound system, as measured in situ, was flat within 10 dB for frequencies between 100 and 3000 Hz, with considerable loss for frequencies exceeding 3000 Hz (some 24 dB/oct). Identical stimuli were presented to both ears in all cases. The average sound intensity was in the range of 60–80 dB SPL

Neural activity was amplified (Grass P16), action potentials were transformed into unitary pulses by means of a level discriminator and fed to a Data Acquisition System (DAS), built around a PDP 11/45 The timing of both the PSG and the DAS was controlled



Fig. 1a and b. Segment from a sequence of γ -tones a Signal x(t) as a function of time. b Upper trace instantaneous amplitude a(t). Lower

trace instantaneous frequency $f(t) = \frac{1}{2\pi}\omega(t)$

by the same external clock, providing a time resolution of neural events, relative to the stimulus, of $10 \,\mu s$.

Spectro-temporal analysis of the stimulus-event relation was performed by means of software (AFT-cube, Average IFT, Sect. 3.2) and hardware (Dynamic Spectrum Analyser; Aertsen and Johannesma, 1980). For the purpose of the latter type of analysis the DAS could be operated in an "inverse" mode, providing unitary pulses in real time from filed spike trains, synchronized with the stimulus from the PSG

3. Spectro-Temporal Analysis of the Stimulus-Event Relation

3.1. The Pre-Event Stimulus Ensemble (PESE)

Extracellular single unit recordings from auditory neurons during presentation of an acoustic *Stimulus Ensemble* (SE) result in sequences of action potentials. Because of the all or none character of the action potentials the single unit activity can be modelled as a series of neural events, a stochastic point process:

$$z(t) = \sum_{n=1}^{N} \delta(t - t_n) \quad (0 \le t \le T)$$
(3.1)

with $\delta(t) = \text{Dirac}$ delta function; $\{t_n\} = \text{moments}$ of occurrence of action potentials, recorded during presentation of the Stimulus Ensemble; T = duration of the SE; N = number of action potentials.

The question which stimulus properties determine the neural events can be investigated by means of an analysis of the properties of the stimulus-event (s-e)relation. The nature of the *s*-*e* relation is reflected in the characteristics of the *Pre-Event Stimulus Ensemble* (PESE), defined as the ensemble of stimuli which precede a neural event (Johannesma, 1972; Johannesma et al., 1973; Grashuis, 1974):

PESE = {
$$x_n(\tau) = x(t_n - \tau); n = 1, N; \tau \ge 0$$
}. (3.2)

The concept of the PESE is illustrated in Fig. 2. By definition the PESE forms a subset of the SE. A comparison of the characteristics of the PESE as compared to those of the SE provides a description of the function of the unit in the neural representation of acoustic stimuli.

3.2. Parametric Analysis of the PESE: AFT-cube and Average IFT

The concepts of the parametric signal representation, described in the accompanying paper (Part I), and the Pre-Event Stimulus Ensemble (Sect. 3.1) will now be combined for the analysis of the single neuron response to a γ -tone ensemble. Apart from a phase constant each stimulus preceding a neural event is fully described by its instantaneous amplitude $a(\tau)$ and instantaneous frequency $f(\tau)$. Therefore each element from the PESE can be geometrically depicted by a trajectory in a 2-dimensional (a, f)-space (i.e. a parameter space) or, equivalently, by a curve in a 3-dimensional (a, f, τ) -space. The latter representation will be used in the following.

The PESE corresponds to the set of pre-event curves in (a, f, τ) -space. The functional properties of the neuron, generating the events, are reflected in this set of PESE-curves as compared to the a priori set of SE-curves. A graphical representation of the parametric description of the PESE (and likewise of the SE) is



Fig. 2. Definition of the Pre-Event Stimulus Ensemble (PESE)

given by the AFT-cube, shown in Fig. 3. Sampled replicas of PESE-curves are mapped into the AFTcube, the axes of which denote, respectively, instantaneous amplitude $a(\tau)$ (vertical axis), instantaneous frequency $f(\tau)$ (horizontal axis, log scale), and time τ before the occurrence of action potentials. Values of $a(\tau)$ have been normalized with respect to the maximum amplitude occurring in the γ -tone ensemble. The AFT-cube is made up of a number of (f, τ) -cross sections taken at equidistant values of a. Each cross section contains a number of dots which represent the projections of portions of sampled (a, f, τ) -curves within a small range Λa around the selected value of a. Throughout this paper Δa was chosen 0.1, the sample interval of (PE)SE-curves was 1.01 ms.

The AFT-cube shown in Fig. 3a represents the PESE for the torus semicircularis unit 26-6. This result should be confronted with the characteristics of the SE.



Fig. 3a and b. AFT-cubes parametric representation of tonal elements from the Pre-Event Stimulus Ensemble of the torus semicircularis unit 26-6 (a) and the Stimulus Ensemble (b) The parameters of interest are instantaneous amplitude a, instantaneous frequency f and time before the event τ The cube's axes correspond to these parameters. The number of elements N from the PESE resp. the SE is indicated in the figures. Further explanation in text



Fig. 4a–c. Cross sections in parameter space of the Pre-Event Stimulus Ensemble of the torus semicircularis unit 26-6 (cf. Fig. 3a). a Instantaneous amplitude vs. instantaneous frequency: (a, f). b Instantaneous amplitude vs. time before neural event: (a, τ) . c Instantaneous frequency vs. time before neural event: (f, τ) For each of the cross sections the numerical values of the parameter window and the number of elements from the PESE having values within this window are indicated. In all three cases the complete PESE was subjected to the analysis. The location in parameter space of the various windows (as indicated by arrows) causes the difference between the numbers of elements within the various cross sections. Further explanation in text

These are depicted in the AFT-cube in Fig. 3b. In this case a random sequence of events which bears no specific relation to the γ -tone ensemble supplied the triggering moments. This random sequence of events was made up of the moments of positive going level crossings of a Gaussian wideband noise. From the results in Fig. 3 we observe for unit 26-6 a relative abundancy of dots in a specific region of the AFTcube. More detailed information about the characteristics of this specific region can be obtained from different types of cross sections in (a, f, τ) -space. Some examples for unit 26-6 are shown in Fig. 4. An (a, f)cross section (Fig. 4a) gives the amplitude-frequency selectivity of the unit at a certain time τ before the spike; dimensionally it may be compared to a "tuning curve". An (a, τ) -cross section (Fig. 4b) gives the temporal amplitude sensitivity at a certain frequency, whereas an (f, τ) -cross section (Fig. 4c) shows the spectro-temporal selectivity at a certain amplitude. The periodic structure of the (a, τ) -cross section (Fig. 4b), also appearing in the AFT-cube (Fig. 3a), is a combined effect of the periodic structure of the stimulus ensemble (every 16 ms one γ -tone) and the neuron's firing characteristics. From Fig. 4b we note that the PESE did not contain any low amplitude tones with the preferred frequency at around 23 ms before the event, which points to an amplitude threshold mechanism. Furthermore we note from Fig. 4b the relative low frequency of occurrence of higher amplitude tones at around 39 ms before the event as compared to the more uniform density at around 7 and 55 ms before the event, which reflects a post-activation suppression phenomenon. From Fig. 4c we note the frequency dependence of the latency. The small regularities along the frequency axis (Figs. 3a and 4c) do not reflect a true neuron property, but are due to the interaction of statistical properties of the y-tone ensemble and the neural response properties (cf. Discussion and Appendix).

For tonal stimuli both the SE and the PESE lead to a collection of points in (a, f, τ) -space, visualized by the AFT-cube. If the stimulus ensemble is rich enough the number of points per unit volume in (a, f, τ) -space $r(a, f, \tau)$ forms a reliable estimate of the density $\varrho(a, f, \tau)$ of stimuli in parameter space. The chosen resolution of a, f, and τ (i.e. the dimensions of the unit volume in parameter space) and the number of events N, used to compile the AFT-cube determines the precision Λr of the estimate. After proper normalization the density $\varrho(a, f, \tau)$ may be interpreted as a probability density function: the probability per unit extension in parameter space to find particular (a, f, τ) combinations in the stimulus ensemble.

Similarly the AFT-cube of the PESE leads to the histogram $r_e(a, f, \tau)$ as an estimate of the pre-event



Fig. 5a and b. Average IFT average intensity I as a function of frequency f and time before the event τ of tonal elements from the Pre-Fvent Stimulus Ensemble of unit 26-6 (a) and the Stimulus Ensemble (b) (cf, Figs 3 and 4) The number of elements N from the PESE resp the SE considered in the averaging procedure is indicated in the figures. The intensity scales in a and b are in terms of arbitrary, identical units. Further explanation in text

probability density in parameter space $\varrho_e(a, f, \tau)$ The reliability of this estimate is determined by the a priori precision Δr . The probabilistic interpretation of results from the parametric analysis of the SE and PESE is inspired by the ensemble theory as used in statistical physics (e.g. Tolman, 1938), which is rendered feasible by the stochastic nature of the mechanism underlying neural firing

On the base of the densities $\varrho(a, f, \tau)$ and $\varrho_e(a, f, \tau)$ the expected value of any quantity $q(a, f, \tau)$ over the SE respectively the PESE can be calculated In the study of the auditory system the intensity of a signal as a function of both frequency and time is of interest. The expected value of the spectro-temporal intensity of the sound in the PESE is given by

$$E(I_e(f,\tau)) = \int da \, a^2 \varrho_e(a,f,\tau) \,. \tag{33}$$

This function can be estimated by the corresponding average over the PESE

$$\langle I_e(f,\tau) \rangle = \frac{\sum_{a} a^2 r_e(a, f, \tau)}{\sum_{a} r_e(a, f, \tau)}.$$
(3.4)

The integration in (3 3) and the summation in (3 4) extend over all amplitude values present in the PESE. Similar expressions hold for the a priori expected value over the SE

The average spectro-temporal intensity of sound in the PESE and the SE for an ensemble of γ -tones is shown for unit 26-6 in Fig 5

Ordinate values in Fig. 5a and b refer to identical, arbitrary units The result of this type of analysis will be referred to as *Average* IFT. The Average IFT of unit 26-6 is shown in Fig. 6a in another representation, the grey density coding, where intensity is represented by means of a number of discrete grey-levels

Regarding Figs. 5 and 6a the following observations can be made

I. Figure 5b shows that for the SE $\langle I(f,\tau) \rangle$ is a constant value I_0 within statistical variations

II Figure 5a and 6a indicate that for the PESE on the base of values of $\langle I_{\varrho}(f,\tau) \rangle$ four spectro-temporal domains can be discerned

1 The average intensity is relatively high in the domain $(f, \tau) \in (300-750 \text{ Hz}, 15-30 \text{ ms})$

2 The average intensity is relatively low (but not zero) in the domain $(f, \tau) \in (300-750 \text{ Hz}, 30-50 \text{ ms})$

3 The average intensity is relatively low (down to zero) in the domains $(f, \tau) \in (125 \ 300 \text{ Hz}, 15-30 \text{ ms})$ and $(f, \tau) \in (750-2000 \text{ Hz}, 15-30 \text{ ms})$

4 The average intensity is, apart from the regularities already noted, similar to the a priori expected value for all other (f, τ) -combinations.

These observations are summarized schematically in Fig 6b

The experimental observations lead to the following conclusions regarding unit 26-6 (see Fig 6c)

1 There is an increased probability of neural firing upon adequate stimulation with tones: activation A with a simple frequency selectivity and timing preference around $(f, \tau) = (520 \text{ Hz}, 22 \text{ ms})$

2. Once the neuron has fired to an adequate tonal stimulus, there is a decreased probability of firing to the same type of stimulus *suppression S*, which may be due to refractoriness, adaptation and/or post-excitatory inhibition. The suppression may last up to





Fig. 6a-c. Average IFT of tonal elements from the Pre-Event Stimulus Ensemble of unit 26-6. a Grey density representation of result as shown in Fig. 5a. Note that here the frequency- and time-axes have been interchanged with respect to Fig. 5a. b Schematic description of Figs. 5a and 6a. The figures in Fig. 6b refer to the text. c Schematic interpretation of Figs. 5a and 6a; spectro-temporal receptive field, consisting of activation (A) and supression (S). Further explanation in text

20 ms. Due to the definition of τ as time before the neural event this suppression domain is preceding the activation domain in Fig. 6c.

3. There cannot be made precise conclusions regarding the spectro-temporal domain 3 (possible lateral effects), due to the absence of spontaneous activity, the specific spectro-temporal sclectivity of the neuron and the single tone character of the SE. At any rate, in this domain a possible activation is not dominant over a possible suppression.

4. The neuron shows a "memory" which, including the latency of some 15 ms, is in the order of 50 ms.

Summarizing, the average spectro-temporal intensity of the PESE, the particular selection from the SE as it was made by the neuron, deviates significantly from what could be expected a priori in a specific domain in frequency and time: the spectro-temporal receptive field of the auditory neuron as determined by an ensemble of γ -tones on the base of a parametric analysis. The receptive field for unit 26-6, made up of an activation domain and a suppression domain, is shown schematically in Fig. 6c.

It should be noted that neither the (a, f)-space nor the (a, f, τ) -space fulfils the requirements of a vector space (Lang, 1973). As a consequence simple operations on the trajectories in parameter space may result in trajectories corresponding to signals which do not belong to the original ensemble. As a simple example the addition or averaging of two tonal stimuli does not normally result in a tonal stimulus. Likewise the Average IFT of the PESE based on the use of tonal stimuli represents sound but not necessarily a tonal sound. In general the Average IFT (like the AFT-cube) represents the degree of optimality of sounds for the neuron (optimal in the sense of probability of spike generation). The sounds have been optimized with respect to the frequency and amplitude of the separate components, the optimum phase relations of the components still have to be determined.

3.3. Average Dynamic Power Spectrum of the PESE

A meaningful interpretation of the AFT-cube and Average IFT, as introduced in the preceding section, is restricted to a stimulus ensemble of narrowband sounds, where instantaneous amplitude and frequency, apart from being well defined, are intuitively understandable. Although formal problems not necessarily have to arise, this approach is less fruitful for wideband sounds like noise, clicks, tonal complexes and most natural sounds.

A more general spectro-temporal description of signals is given by the Dynamic Power Spectrum (e.g. Flanagan, 1965) or spectrogram (e.g. Koenig et al., 1946; Potter et al., 1947). The dynamic power spectrum P(f,t) of a signal gives the intensity as a function of time for the different frequency components of the signal. It can be measured by a Dynamic Spectrum Analyser (DSA). The mathematical background and principles of operation of this instrument are described in the accompanying paper (Part I, Aertsen and Johannesma, 1980). The DSA determines the dynamic power spectrum of a signal in the form of a set of functions of time, representing the instantaneous intensity of the signal in a set of neighbouring 1/3 octave bands.

In general the dynamic spectrum P(f, t), which is a nonlinear functional defined on an arbitrary signal, describes the spectro-temporal structure of that signal.



Fig. 7a and b. Spectro-temporal receptive field of torus semicircularis unit 26-6, determined by an ensemble of γ -tones. a Different representations of the Average Pre-Event Dynamic Power Spectrum, determined by the DSA. The number of elements N_1 presented to the DSA and the number of elements N_2 actually considered in averaging ($N_2 < N_1$) are indicated in the figure b Average IFT (cf. Figs. 5a and 6a) Note that the frequency- and time-axes have been interchanged with respect to Fig. 5a

In this approach the dynamic spectrum is considered to be an important characteristic of an acoustic stimulus. This leads to the hypothesis that the dynamic spectrum may differ distinctly for signals belonging to the PESE as compared to signals from the SE. A possible way to investigate this hypothesis is the computation of the Average Dynamic Power Spectrum of the PESE:

$$\langle P_e(f,\tau) \rangle = \frac{1}{N} \sum_{n=1}^{N} P_n(f,\tau),$$
(3.5)

where $P_n(f, \tau)$ is the dynamic spectrum of the *n*-th element of the PESE. This procedure of spike triggered averaging of pre-event dynamic spectra is implemented in the DSA (cf. Part I). The experimental results for the PESE of unit 26-6, using again the γ -tone stimulus ensemble, is given in Fig. 7a. For comparison the results of the Average IFT for this neuron, using the same γ -tones, is shown in Fig. 7b. The latter result is identical to the one shown in Fig. 5a; to facilitate a comparison the frequency- and time-axes have been interchanged.

Due to limitations of the DSA ("dead time" upon triggering, in this case about 32.3 ms; cf. Part I) not all of the elements of the PESE have been considered in averaging. The numbers involved are indicated in Fig. 7a. The average dynamic power spectrum of the SE (not shown here), like the Average IFT, gives an approximately constant value for all (f, τ) combinations.

From this result we conclude that the Average Dynamic Power Spectrum of the PESE for unit 26-6

indeed differs distinctly from the a priori expected result i.e. the Average Dynamic Power Spectrum of the SE. For narrowband signals, remaining within a given 1/3 octave band the dynamic power spectrum reduces to the parametric IFT-representation. Therefore it is not surprising that the spectro-temporal receptive field for γ -tones as determined by the DSA (Fig. 7a), apart from obvious differences in resolution, agrees well with the receptive field determined on the basis of a parametric description (Figs. 6a and 7b). The distortion, caused by the frequency dependent delay in the DSA (cf. Part I), however, should be taken into account in the estimation of time constants like latency.

In the frame of the dynamic power spectrum analysis we have defined and constructed a spectrotemporal analysis of the PESE for a neuron which, contrary to the parametric approach, can be applied for an arbitrary choice of the stimulus ensemble. In particular this procedure enables to determine the spectro-temporal receptive field of a neuron in a (pseudo) natural acoustic environment (Aertsen and Johannesma, 1979). This leads to the question whether the spectro-temporal receptive field is dependent on the type of stimulus ensemble and, if so, whether normalizing procedures can be developed, to account for this. In view of the dimensions of the analysis, i.e. frequency and time, a natural first choice would be to correct the Average Dynamic Power Spectrum of the PESE for the overall spectro-temporal structure of the SE. The spectro-temporal receptive fields of auditory neurons for a natural stimulus ensemble will be investigated in more detail in the forthcoming paper (Part III).



Fig. 8a-c. Spectro-temporal analysis of torus semicircularis unit 25-5, using an ensemble of γ -tones. **a** AFT-cube. **b** Average IFT in different representations. **c** Average Pre-Event Dynamic Power Spectrum in different representations. The numbers of elements from the PESE, taken into account in the various procedures, are indicated in the figures. Further explanation in text



Fig. 9. Spectro-temporal analysis of torus semicircularis unit 26-8, using an ensemble of γ -tones. Details as in Fig. 8

The analysis as described in this paper was applied to 29 torus semicircularis recordings, obtained from 11 grassfrogs. This leads to a variety of results, four examples of which are shown in Figs. 8 11.

In all examples shown we present the AFT-cube (a), the Average IFT in two representations (b) and the average dynamic power spectrum, likewise in two representations (c). Note the different frequency and time scales in the various cases. The values of the average measures (b and c) have been scaled to the individual extrema. The "dead-time" in the DSAanalysis (cf. Part I) in these examples amounted to about 32.3 ms (Fig. 8c), resp. 64.6 ms (Figs. 9c, 10c, and 11c). None of the neurons did show any spontaneous activity. The units 25-5 (Fig. 8) and 26-8 (Fig. 9) represent neurons with relatively high firing rate under stimulus conditions; on the average 2.96 resp. 2.82 spikes/s, whereas the units 38-1 (Fig. 10) and 26-4 (Fig. 11) showed a much lower activity: on the average 0.05 resp. 0.12 spikes/s. In the latter two cases the activity decreased considerably during the first few seconds of presentation of the y-tone ensemble.

In all 29 cases it was possible to distinguish an activation type of spectro-temporal receptive field. In a number of cases the receptive field also showed a more (e.g. Figs. 8 and 9) or less (e.g. Fig. 10) clear suppression domain preceding the activation domain. This shows up most clearly in the grey density representations.

The limited number of recordings does not allow to draw any inferences regarding a possible classification of torus semicircularis neurons into subpopulations.

A closer inspection of the examples shows the following: Unit 25-5 (Fig. 8) is responsive to tones with frequency within a band of about 1 octave around 1350 Hz. The latency, defined as the interval between the neural event and the moment in time of the maximum value of the Average IFT, is about 20 ms. This particular definition, as arbitrary as any other one, was adopted because it can still be used when the result contains more noise (e.g. Fig. 11). From Fig. 8a we note that the latency in firing to individual tones is frequency dependent, with shortest latency for a frequency of about 1350 Hz and gradually increasing with increasing or decreasing frequency. The quite sharply defined activation type receptive field is preceded by a suppression domain with similar frequency characteristics, which does not extend in time to more than 50 ms before the action potential (Fig. 8b and c).

Unit 26-8 (Fig. 9) shows a wide frequency selective band for activation at low frequencies (125-300 Hz)without a pronounced best frequency. The lower frequency bound of the γ -tone ensemble (125 Hz) obviously is still within this unit's receptive field. The latency, which in this case is hardly frequency dependent (Fig. 9a), is about 33 ms. The sharply defined activation domain is preceded by a suppression domain at the same low frequencies which comprise the activation domain, lasting up to about 100 ms before the event (Fig. 9b and c). Note that the suppression domain becomes narrower in frequency with increasing time before spike. Especially in this case (Fig. 9c) the temporal distortion, caused by the frequencydependent delay in the DSA (cf. Part I) is very obvious.

Unit 38-1 (Fig. 10) shows a less sharply defined activation domain with best frequency around 1900 Hz, bandwidth 0.5-1 octave. The latency equals about 55-60 ms. The activation domain is preceded by a suppression domain with the same frequency characteristics, extending in time to as much as about 125 ms before the event (Fig. 10b and c). In this case we could not exclude the possibility that the recording also contained some spikes from one or more other neurons.

Unit 26-4 (Fig. 11) shows a vague activation type receptive field for frequencies around 400 Hz, with a latency of about 40–45 ms. This neuron does not show any sign of a suppression domain in the receptive field. The receptive field, although vague, is reproducible under repeated stimulation.

The location and shape of the activation domain is rather varied as is clear from the examples. In general it was not possible to predict the existence of a suppression domain from the information about the activation domain. For instance a unit was found which looked very similar to unit 26-8 (Fig. 9), except that it did not show any sign of suppression preceding the activation domain.

As expected for narrowband stimuli (Sect. 3.3) the Average Pre-Event Dynamic Power Spectra (c) agree with the results of the Average IFT (b).

As already noted the regular fine structure along the frequency axis (e.g. Fig. 9) is not solely attributable to the properties of the neuron. This effect is due to the interaction of the statistical characteristics of the γ -tone sequence and the neuron's receptive field properties (cf. Discussion and Appendix). For the same reason additional evidence had to ensure that the suppression domain in the receptive field truly represents a neural property. This evidence could be obtained from the receptive field properties deduced from stimulation with a γ -tone sequence with frequency values shifted over one octave, and statistical properties shifted correspondingly. In all cases shown and in most of the remaining cases this additional evidence ensured the neural origin of the suppression domain.

5. Discussion

In the present paper we have described methods to investigate the nature of the stimulus-event relation for single auditory neurons. The characteristics of the *Pre-Event Stimulus Ensemble* (PESE) as compared to those



Fig. 10a–c. Spectro-temporal analysis of torus semicircularis unit 38-1, using an ensemble of γ -tones Details as in Fig 8

UNIT 26-4



Fig. 11a–c. Spectro-temporal analysis of torus semicircularis unit 26-4, using an ensemble of γ -tones Details as in Fig. 8

of the total Stimulus Ensemble (SE) indicate which stimulus properties determine the occurrence of action potentials. In this investigation we have focused on a spectro-temporal analysis of the PESE. Two approaches for this analysis have been described. The first one is based on a parametric description of sound in terms of amplitude and frequency as functions of time. A parametric representation of the PESE has been given: the AFT-cube. Ouadratic averaging over different amplitude values gives the average intensity of stimuli in the PESE as a function of frequency and time before the neural event: Average IFT. A useful application of the parametric approach is restricted to narrowband stimuli. An approach, applicable to an arbitrary ensemble of stimuli is based on the description of sound by the Dynamic Power Spectrum. In this approach the introduction of averaging procedures leads to the Average Pre-Event Dynamic Power Spectrum. In order to evaluate these methods, they have been applied to the event sequences recorded from single neurons in the torus semicircularis of the grassfrog, stimulated with an ensemble of narrowband stimuli: the y-tones.

Results and Restrictions

As described in the methods section (Sect. 2) recordings were made from anaesthetized animals (MS-222). This may have affected the firing characteristics of neurons in the torus semicircularis. It is known for some anuran species that already at the level of the auditory nerve barbiturate anaesthesia has a depressant effect on neural activity (Capranica, 1976). Neurons in our recordings did not show any spontaneous activity, whereas most of the cells recorded from the torus semicircularis of curarized green treefrogs (Hyla cinerea) were spontaneously active, albeit in general at very low rates (Feng, 1975; quoted by Capranica, 1976). Results of experiments, currently in progress, using a lower level of MS-222 anaesthesia indicate a richer and more complex firing behaviour of neurons in the torus semicircularis.

The stimulus ensemble used in the present study consisted of single y-tones. Apart from the obvious limitations imposed by the use of one particular envelope waveform, the single-tone character combined with the absence of spontaneous activity does not provide information about possible lateral suppression. Moreover the composite spectral pattern of vocalizations (cf. Fig. 1 in Part I) is not represented in the y-tones. Ethological investigations have demonstrated the behavioural significance of these spectral features for some anuran species (e.g. Capranica, 1965; Gerhardt, 1974). These findings, combined with the demonstrated non-linearities in the peripheral auditory system of anurans (Capranica and Moffat, 1979), point to the importance of extending the single-tone ensemble to a multi-tone ensemble.

In order to avoid systematic effects of monotonous parameter variation the y-tones have been presented as a pseudo-random sequence. The statistical properties of this sequence, however, are not ideal in the sense that the different order serial correlations show particular regularities (cf. Appendix). For instance the probability of having a tone with frequency f_2 , given a tone with frequency f_1 immediately preceding it, is not uniform, the same holds for higher order combinations. This is reflected in the spectro-temporal receptive field of more complex neurons, determined by presentation of such a sequence of tones. The spectrotemporal selectivity of a neuron, composed of an activation and a suppression domain, separated in time, effectively results in a neural analysis of serial correlations in the γ -tone sequence. This affects the composition of the PESE for that neuron. The regular fine structure along the frequency axis in the AFTcube and Average IFT as well as the dips in the activation areas of e.g. units 26-6 (Figs. 3 and 5a) and 26-8 (Fig. 9) are caused by this interaction of stimulus properties and receptive field properties. These findings once more point to the important influence of statistical properties of a stimulus ensemble, based on pseudo-random algorithms (e.g. Swerup, 1978; Eckhorn and Pöpel, 1979). For the neurons as mentioned here a monotonous parameter variation would have caused even more drastic effects. A general solution to this type of problems is not available, particular solutions will be dictated by a specific statement of the question of interest and experimental restrictions as for instance limited recording time.

Recordings from the torus semicircularis did not always result in (quasi-)stationary sequences of neural events. In a number of cases (e.g. the units 26-4 and 38-1 in Sect. 4) the firing rate after a few seconds of stimulation decreased to less than 5% of the initial firing rate. This "long-term adaptation" appeared to be reproducible, provided that a long enough silent period had elapsed. Adaptation is not measured by application of our analysis procedures to the total recording of neural activity. Application of the analysis to segments of these recordings suggests an increased spectro-temporal selectivity in the adapted stage as compared to the initial stage. Apparently for a number of neurons, using this tonal stimulus ensemble, the preevent stimulus ensemble is inhomogeneous. Obviously a further investigation into this adaptation phenomenon demands more specific experiments and analysis. The results presented in this paper bear to the integral recording in all cases. The neurons, which show this "long-term adaptation", have spectro-temporal receptive fields which are, under the present stimulus conditions, more vaguely defined than for the "nonadapting" neurons. Moreover they generally show long latencies (\sim 40–70 ms) as compared to the "nonadapting" neurons ($\sim 10-35$ ms). These findings point

to a possible sub-population of more complex neurons in the torus semicircularis with long integration time and/or of higher order.

Since in all experiments stimuli were presented binaurally it is not possible to draw any conclusion regarding binaural interactions (Feng, 1975; quoted by Capranica, 1976).

Frequency and Time

The goal of the analytic procedures presented here is a spectro-temporal representation of the stimulus-event relation instead of separate spectral and temporal descriptions. This choice was influenced mainly by two arguments. The first argument concerns the neuronal characteristics to be encountered. In general there are no a priori arguments to expect that the spectrotemporal selectivity of auditory neurons is separable in two independent dimensions of frequency and time. This implies that analytic procedures which aim at a general applicability have to refrain from an a priori separation of the relevant variables. Separability should be an outcome of the analysis rather than an assumption. For instance the cases of "cf/fm-sensitive" neurons (e.g. Suga et al., 1978) and neurons with composite intensity, frequency and time selectivity (e.g. Miller et al., 1974) are examples where an a priori separation may give rise to misleading conclusions. The second argument concerns the stimulus ensemble. In order to find out to what extent the stimulus-event relation is invariant with respect to the type of stimuli presented (e.g. tones, noise, vocalizations), it is necessary that in all cases the s-e relation is analysed in terms of the same variables by means of preferably identical or at least comparable methods. Complex sounds such as vocalizations and noise possess an intricate spectro-temporal structure (see for example the numerous sonograms in Potter et al., 1947). This implies that the use of a spectro-temporal representation of simple tonal stimuli facilitates a comparison to results for more complex stimuli. The arguments given here lead to analytic procedures which are applicable to a wide class of stimulus ensembles and a wide class of neural characteristics within the spectro-temporal domain. A complementary property of these general procedures is that they provide more global information on the stimulus-event relation. Once these global properties have been found more specific hypotheses can be tested using more specific stimuli and more specific analytic procedures.

An a posteriori inspection of results for a number of neurons in the torus using the present analytic procedures does not reveal a very specific spectrotemporal coupling in the receptive fields, apart from a frequency dependence of the latency in some cases. Recent experiments in our laboratory, however, indicate that, at least for low frequency units, the picture is more complex. The AFT-cube provides a parametric representation of the Pre-Event Stimulus Ensemble. This type of result is closely related to a dot display analysis of frequency and/or intensity scans (Nelson et al., 1966; Van Gisbergen et al., 1975). From the AFT-cube and its different cross sections it is possible to extract neural characteristics like best frequency, latency, amplitude sensitivity, bandwidth etc. Several of these parameters may turn out to be a function of one or more of the three variables in the AFT-cube (e.g. latency as a function of frequency, bandwidth as a function of time).

The AFT-cube in this context provides the most complete representation of the s-e relation, while both the Average IFT and the Average Pre-Event Dynamic Power Spectrum provide an average characteristic of the PESE, thereby necessarily discarding information present in the AFT-cube. These two average measures describe the average distribution of intensity over frequency and time of stimuli which preceded a neural event. This interpretation of these measures is expressed in the term "spectro-temporal receptive field". In this context the concept of "receptive field" of a neuron should be interpreted in an abstract sense as the collection of characteristics of stimuli which affect the neural firing behaviour (e.g. Sejnowski, 1976). This definition is a generalization and a projection into the stimulus domain of the original definitions, which were expressed in terms of the receptor surface (Adrian et al., 1931; Hartline, 1940; quoted by Grüsser and Grüsser-Cornehls, 1973). In the auditory literature the concept has been used with reference to the spatial properties of adequate acoustic stimuli (Wickelgren, 1971; Knudsen and Konishi, 1978). In the present study the dimensions of interest are abstract parameters of the stimulus, namely frequency and time, which is expressed in the term spectro-temporal receptive field. The receptive field may consist of different subdivisions: activation and/or suppression, their position in parameter space is given by the (f, τ) -coordinates of the extrema of the average intensity density (or, alternatively, of the mean). The "receptive field size" can be defined by the product $2\pi\sigma_t\sigma_t$ of the intensity density over the relevant area(s), which, from the point of view of the animal, is analogous to the "uncertainty product" as defined for signals (Gabor, 1946). For the examples shown the receptive field subdivisions were fairly large (5-25) as compared to the "size" of the y-tones (0.65) (cf. Part I) and, likewise, the main frequency components of the vocalization elements. We did not find double activation areas, neither was their frequency extent large enough to encompass different frequency bands in the vocalizations. It is striking to note the similarity of time constants in the composite receptive fields (activation and suppression) and the temporal composition of the vocalizations (sound elements, separated by silent periods, cf. Fig. 1 in Part I).



Fig. 12a-d Scatter diagrams of serial correlation of numbers r in pseudo random sequences, based on the original "HP algorithm '(a and b) and the permutation-algorithm (c and d) a and c refer to serial correlation with lag 1, b and d refer to lag 2 The number of bits n in the shift register in this case is 8, which results in a sequence of 255 numbers Further explanation in text

Both the Average IFT and the Average Dynamic Spectrum are cross correlation functions of the neural activity and functionals of the stimulus In this sense they are related to a more general type of analysis of nonlinear systems, the Wiener approach (e g Lee and Schetzen, 1965) At this level of analysis the correspondence is of an intuitive type The Average IFT and the Average Dynamic Power Spectrum can be formally connected by introduction of a general Complex Spectro-Temporal Intensity Density function of signals (CoSTID), defined on the analytic signal (Rihaczek, 1968, Johannesma and Aertsen, 1979) This underlying function can be tied up with the second order Wiener kernel

The Average IFT technique can provide a detailed description of the spectro-temporal receptive field, the resolution is determined by the spectro-temporal resolution of the stimulus ensemble The Average Dynamic Power Spectrum gives a more global measure of the receptive field, the resolution is determined by the filter characteristics of the spectrum analyser The Average IFT is applicable for an ensemble of *narrowband* stimuli, the dynamic spectrum analysis can be applied to an *arbitrary* stimulus ensemble

In the present investigation we have studied the responses from single torus semicircularis neurons to γ -tones. The waveform of the γ -tones was chosen in an attempt to mimic some of the characteristics of the male

grassfrog's vocalizations It is tempting to interpret the describing parameters of the γ -waveform as "bioacoustic dimensions" of the vocal repertoire of the male grassfrog Variation of the parameters – in this case A and ω – in neurophysiological experiments may then be interpreted as an investigation into the possible "focal zônes" at the single neuron level (Scheich, 1977) The receptive field properties of torus semicircular neurons for a wide ensemble of natural stimuli (vocalizations, environmental sounds), using the techniques as described in this paper, will be the subject of the forthcoming paper (Part III)

Appendix

The Pseudo-Random Sequences

The pseudo-random modulation of the amplitude factor A and the carrier frequency ω_{γ} in the sequence of γ -tones (cf Part I) was generated by a maximum length series algorithm as implemented in the HP3722A noise generator. The contents of the *n* bits shift register with feedbacks, used in the closed loop condition for generating pseudo-random binary sequences, was interpreted as a sequence of *n* bits integer numbers. This results in a sequence of $2^n - 1$ different values, being a permutation of the natural numbers in the range $(0, 2^n)$. The amplitude sequence consisted of 127 values (n=7), the frequency sequence contained 255 values (n=8). These values can be converted to values in the case of ω_{γ} . Thereby the amplitude factor and the carrier frequency can be considered as random variables of a lattice type (Papoulis, 1965) with a uniform probability on the lattice points.

In order to reduce the strong serial correlations between succeeding values, the sequence was permuted to a new one by taking as succeeding elements those from the original sequence with order numbers 1, n+1, 2n+1, etc, with cyclic continuation of the original sequence. Thereby the contents of the shift register has been totally "refreshed" for each new element of the sequence. This procedure reduces the serial correlation in the sequence to a large extent, as is demonstrated for the case n=8 by the scatter diagrams of the serial correlations with lags 1 and 2 in Fig 12

It is evident, however, that the restrictions of short pseudorandom sequences cannot be fully overcome Some degree of structure in the sequence remains Note for instance the non-uniform distribution of points in the scatter diagram with lag 1 (Fig 12c), with rather large holes, also along the diagonal and the regular fine structure in the scatter diagram with lag 2 (Fig 12d) These observations have to be taken into account in the interpretation of results from experiments using these pseudo-random sequences

Acknowledgements This investigation forms part of the research program of the Workgroup Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (The Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (ZWO) The authors wish to thank Jos Eggermont and Ton Vendrik for helpful discussions on this manuscript The expertise of Koos Braks, Jan Bruijns, Wim van Deelen, and Jos Olders was a great help in the experiments and data analysis

References

Adrian, E D, Cattell, M, Hoagland, H Sensory discharges in single cutaneous nerve fibers J Physiol (London) 72, 377-391 (1931)

Aertsen, A M H J, Johannesma, P I M Spectro-temporal analysis of auditory neurons in the grassfrog In Hearing mechanisms and speech, pp 87–93 Creutzfeld, O, Scheich, H, Schreiner, Chr (cds) Exp Brain Res [Suppl II] (1979)

- Aertsen, A M H J, Johannesma, P I M Spectro-temporal receptive fields of auditory neurons in the grassfrog I Characterization of tonal and natural stimuli Biol Cybernetics 38, 223-234 (1980)
- Bullock, T H (ed) Recognition of complex acoustic signals (Life Sciences Res Rep, Vol 5) Berlin, Heidelberg, New York Springer 1977
- Capranica, R R The evoked vocal response of the bullfrog a study of communication by sound Cambridge, MA MIT Press 1965
- Capranica, R R Morphology and physiology of the auditory system In Frog neurobiology, a handbook, pp 551-575 Llinas, R, Precht, W (eds.) Berlin, Heidelberg, New York Springer 1976
- Capranica, R R, Moffat, A J M Nonlinear properties of the peripheral auditory system In Comparative studies of hearing in vertebrates Berlin, Heidelberg, New York Springer 1979
- Dowben, R M, Rose, J E A metal-filled microclectrode Science 118, 22-24 (1953)
- Eckhorn, R, Popel, B Generation of Gaussian noise with improved quasi-white properties Biol Cybernetics 32, 243-248 (1979)
- Evans, F F Cochlear nerve and cochlear nucleus In Handbook of sensory physiology, Vol V/2 Auditory system, pp 1-108 Keidel, W D, Neff, W D (eds.) Berlin, Heidelberg, New York Springer 1975
- Feng, A S Sound localization in anurans an electrophysiological and behavioral study Ph D Thesis, Cornell Univ, Ithaca, New York 1975
- Flanagan, J L Speech analysis, synthesis, and perception Berlin, Heidelberg, New York Springer 1965
- Gabor D Theory of communication, J IEF (London) Part III 93, 429 457 (1946)
- Gelder, JJ van, Evers, PMG, Maagnus, GJM Calling and associated behaviour of the common frog, Rana temporaria, during breeding activity J Anim Ecol 47, 667–676 (1978)
- Gerhardt, H C The significance of some spectral features in mating call recognition in the green treefrog (Hyla cinerea) J Exp Biol 61, 229–241 (1974)
- Gerhardt, H C Mating call recognition in the green treefrog (Hyla cinerea) the significance of some fine-temporal properties J Exp Biol 74, 59-73 (1978)
- Gisbergen, J A M van, Grashuis, J L, Johannesma, P I M, Vendrik, A J H Spectral and temporal characteristics of activation and suppression of units in the cochlear nuclei of the anaesthetized cat Fxp Brain Res 23, 367–386 (1975)
- Grashuis, J L The pre-event stimulus ensemble an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons Ph D Thesis, Nijmegen 1974
- Grusser, O-J, Grusser-Cornehls, U Neuronal mechanisms of visual movement perception and some psychophysical and behavioral correlations In Handbook of sensory physiology, Vol VII/3A Central processing of visual information, pp 333 429 Jung, R (ed) Berlin, Heidelberg, New York Springer 1973
- Hartline, H K The receptive fields of the optic nerve fibers Am J Physiol 130, 690-699 (1940)
- Johannesma, P I M The pre-response stimulus ensemble of neurons in the cochlear nucleus In Proc of the IPO Symp on Hearing Theory, pp 58-69 Cardozo, B L (ed) Eindhoven 1972
- Johannesma, PIM, Gisbergen, JAM van, Grashuis, JL, Vendrik, AJH Forward and backward analysis of stimulus response relations of single cells in the cochlear nucleus In Proc IV International Biophysics Congress (IUPAB), Vol 3, pp 508– 523 Moscow USSR Academy of Sciences 1973

- Johannesma, PIM, Aertsen, AMHJ Neural image of sound in the grassfrog In Hearing mechanisms and speech, pp 79 86 Creutzfeld, O, Scheich, H, Schreiner, Chr (eds) Exp Brain Res [Suppl II] (1979)
- Kiang, NY-S, Watanabe, T, Thomas, EC, Clark, I F Discharge patterns of single fibers in the cat's auditory nerve Cambridge, MA MIT Press 1965
- Knudsen E I, Konishi, M Center-surround organization of auditory receptive fields in the owl Science 202, 778-780 (1978)
- Koenig, W, Dunn, HK, Lacy, LY The sound spectrograph J Acoust Soc Am 18, 19-49 (1946)
- Lang, S Linear algebra Reading, MA Addison-Wesley 1973
- Lee, Y W, Schetzen, M Measurement of the Wiener kernels of a nonlinear system by crosscorrelation Int J Control 2, 237 254 (1965)
- Miller, J M, Beaton, R D, O'Connor, T, Pfingst, B L Response pattern complexity of auditory cells in the cortex of unanesthetized monkeys Brain Res 69, 101-113 (1974)
- Nelson, PG, Erulkar, SD, Bryan, JS Responses of units of the inferior colliculus to time-varying acoustic stimuli J Neurophysiol 29, 834–860 (1966)
- Papoulis, A Probability, random variables and stochastic processes Tokyo McGraw-Hill Kogakusha 1965
- Potter, RK, Kopp, GA, Green, HC Visible speech New York Van Nostrand 1947
- Rihaczek, AW Signal energy distribution in time and frequency IFFE Trans Inf Theory IT-14, 369-374 (1968)
- Scheich, H Central processing of complex sounds and feature analysis In Recognition of complex acoustic signals, pp 161– 182 Bullock, T H (ed) (Life Sciences Res Rep, Vol 5) Berlin, Heidelberg, New York Springer 1977
- Sejnowski, T J On global properties of neuronal interaction Biol Cybernetics 22, 85-95 (1976)
- Singleton, R C, Poulter, T C Spectral analysis of the call of the male killer whale IEEE Trans Audio Electroacoust AU-15, 104-113 (1967)
- Smolders, JWT, Aertsen, AMHJ, Johannesma, PIM Neural representation of the acoustic biotope a comparison of the response of auditory neurons to tonal and natural stimuli in the cat Biol Cybernetics 35, 11-20 (1979)
- Suga, N, O'Neill, W E, Watanabe, T Cortical neurons sensitive to combinations of information bearing elements of biosonar signals in the mustache bat Science 200, 778-781 (1978)
- Swerup, C On the choice of noise for the analysis of the peripheral auditory system Biol Cybernetics 29, 97 104 (1978)
- Tolman, R C The principles of statistical mechanics London Oxford University Press 1938
- Wickelgren, B.C. Superior colliculus some receptive field properties of bimodally responsive cells Science 173, 69–71 (1971)
- Worden, FG, Galambos, R (eds) Auditory processing of biologically significant sounds (Neurosci Res Prog Bull, Vol 10) Brookline, MA Neurosciences Research Program 1972

Received May 11, 1980

A M H J Aertsen Department of Medical Physics and Biophysics University of Nijmegen Geert Grooteplein Noord 21 NL-6525 FZ Nijmegen The Netherlands

####
CHAPTER 6

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

III. Analysis of the Stimulus-Event Relation for Natural Stimuli

Aertsen, A.M.H.J., Olders, J.H.J. and Johannesma, P.I.M. Biol. Cybernetics <u>39</u>, 195-209 (1981)

Spectro-Temporal Receptive Fields of Auditory Neurons in the Grassfrog

III. Analysis of the Stimulus-Event Relation for Natural Stimuli

A. M. H. J. Aertsen, J. H. J. Olders, and P. I. M. Johannesma

Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Abstract. The stimulus-event relation of single units in the auditory midbrain area, the torus semicircularis, of the anaesthetized grassfrog (Rana temporaria L.) during stimulation with a wide ensemble of natural stimuli, was analysed using first and second order statistical analysis techniques. The average stimulus preceding the occurrence of action potentials, in general, did not prove to give very informative results. The second order procedure consisted in the determination of the average dynamic power spectrum of the pre-event stimuli, following procedures as described elsewhere (Aertsen and Johannesma, 1980; Aertsen et al., 1980). The outcome of this analysis was filtered with the overall power spectrum of the complete stimulus ensemble in order to correct for its non-uniform spectral composition. The "stimulus-filtered" average pre-event dynamic spectrum gives a first indication of the "spectro-temporal receptive field" of a neuron under natural stimulus conditions. Results for a limited number of recordings are presented and, globally, compared to the outcome of an analogous analysis of experiments with tonal stimuli.

1. Introduction

The nature of the stimulus-response or stimulus-event relation of single auditory neurons under natural stimulus conditions has mostly been investigated on a rather qualitative level (e.g. Worden and Galambos, 1972; Bullock, 1977; Smolders et al., 1979). Poststimulus time histograms (PSTH) are determined for (species-specific) vocalizations and their properties are compared to the spectral tuning and/or temporal characteristics like latency, determined with technical stimuli, e.g. tones. This approach has proven to be successful in peripheral parts of the auditory system, however, it has met considerable problems in reconciling the results in more central parts of the brain.

The present paper is the third one in a series, dealing with the spectro-temporal analysis of the response properties of auditory neurons under various stimulus conditions. This analysis basically follows a statistical approach, inspired by procedures which have been used for noise stimuli (De Boer and Kuyper, 1968; Johannesma, 1972, 1980; De Boer and DeJongh, 1978). In the first paper (Part I: Aertsen and Johannesma, 1980) methods were described to give a spectro-temporal characterization of sounds, both narrowband (e.g. tones) and wideband (e.g. natural sounds). The application of these methods to the analysis of the stimulus-event relation using a tonal stimulus ensemble was the subject of the second paper (Part II: Aertsen et al., 1980). The present paper (Part III) is concerned with a first attempt to study the single neuron response to natural sounds, using the procedures as developed in the Papers I and II. The stimulus ensemble consisted of a wide selection of sounds as they occur in the natural environment, i.e. the biotope, of the animal under investigation, the grassfrog (Rana temporaria L.). The stimulus sequence had a duration of approximately 4 min and contained such various sounds as species-specific vocalizations, vocalizations from other frogs and toads, sounds from other animals (prey and predators) and sound from abiotic origin (e.g. wind and rain). The composition and characteristics of this natural stimulus ensemble. the Acoustic Biotope, have been described in detail in Part I.

The properties of the stimulus-event relation have been investigated by analysing the ensemble of stimuli, preceding the occurrence of action potentials, the *Pre-Event Stimulus Ensemble* (PESE; Johannesma, 1972). A comparison of the characteristics of the complete *Stimulus Ensemble* (SE) and this subensemble, the PESE, provides information about the selecting mechanism of the neuron and leads to a functional description of its role in the neural representation of the



Fig. 1a and b. The "B-call" of the male grassfrog (Rana temporaria L.). a Time course of vocalization on different time scales, indicated by horizontal bars. Vertical bars refer to identical, arbitrary amplitude values. b Sonogram (dynamic amplitude spectrum) of vocalization in different representations. Frequency resolution 39.1 Hz, time resolution 12.8 ms

sensory environment. The properties of the PESE can be investigated by an analysis of the ensemble average of various relevant stimulus functionals. In the present investigation this cross correlation approach to the study of the stimulus-event relation has been applied for first and second order signal characteristics. The first order analysis leads to the average pre-event stimulus, in fact the "revcor"-procedure, first introduced by De Boer and Kuyper (1968). The second order properties of the PESE have been studied by analysing the spectro-temporal structure of pre-event stimuli, leading to the average pre-event dynamic power spectrum (Part II, Aertsen et al., 1980). The outcome of the second order analysis has to be normalized for the spectro-temporal properties of the stimulus ensemble in order to arrive at a correct representation of the neural spectro-temporal selectivity, the spectrotemporal receptive field (cf. Part II). As a first, incomplete, attempt the average pre-event dynamic spectra have been "corrected" for the overall spectral composition of the acoustic biotope: "stimulus filtering".

In the present paper the procedures will be described, furthermore the concepts will be demonstrated by application to a number of extra-cellular single unit recordings from the grassfrog's auditory midbrain area, the torus semicircularis, recorded during presentation of the natural stimulus ensemble. The results of the spectro-temporal analysis will be, globally, compared to the findings from tonal experiments, analysed in the very same way (cf. Part II). Such a comparison should provide an indication to what extent the receptive field is a *stimulus-invariant* neural characteristic.

A short, preliminary report on the approach as followed in this paper and the preceding ones (Parts I and II) has been published elsewhere (Aertsen and Johannesma, 1979).

2. Methods

Preparation and Recording

Extracellular single unit recordings were made in the torus semicircularis of adult grassfrogs from Ireland. The methods used for the preparation of the animal and the recording of single unit activity have been described in detail in Part II. During the experiments the animal was maintained at a level of anaesthesia (MS-222) at which corneal reflex returned but no other movements occurred.

Stimulus Presentation and Data Registration

Experiments were performed with the frog in an electrically shielded acoustic chamber (IAC 1202A). The stimulus ensemble consisted of a selection of various sounds as they occur in the natural environment of the animal under investigation, i.e. the grassfrog. The stimulus sequence had a duration of approximately 4 min and contained, amongst others, vocalizations of various species of frogs and toads, vocalizations of predators and prey as well as sounds of abiotic origin (e.g. wind and rain). The characteristics of this natural stimulus ensemble, the *Acoustic Biotope*, and the technical procedures of sound recording and composition of the stimulus tape have been described in detail in Part I. An example of one of the stimulus segments, a "*B*-call" of the male grassfrog (typology according to Van Gelder et al., 1978) is given in Fig. 1.

The stimulus tape was played back to the animal from an AM-recorder (Otari MX 5050QXH, tape speed 38 cm/s). After lowpass filtering (2 cascaded $-24 \, dB/oct$ Bessel filters, each $-3 \, dB$ at 5 kHz) the signal was amplified (Luxman 507X) and presented to the animal by two electro-dynamic microphones (Sennheiser MD 211 N), coupled to the frog's tympanic membranes by means of closed cavities. The frequency response of the sound system, as measured in situ, was flat within 10 dB for frequencies between 100 and 3000 Hz, with considerable loss for frequencies exceeding 3000 Hz (some 24 dB/oct). Identical stimuli were presented to both ears in all cases. The average sound intensity, measured in the acoustic couplers by half inch condensor microphones (Bruel and Kjacr 4143) was in the range of 60–80 dB SPL.

Neural activity was amplified (Grass P16) and recorded on the stimulus delivering tape recorder. It was transformed into unitary trigger pulses (neural events) by means of a level discriminator. The analysis of the stimulus-event relation was performed by means of a hardware signal averager (Biomac 1000, Datalab) and a real-time dynamic spectrum analyser (DSA; Part I). The results of both types of analysis could be read out and transferred to the laboratory's PDP 11/45 for further elaboration. Furthermore in a number of cases the neural events, together with a synchronizing clock signal (20 kHz) from the stimulus recorder, were fed to a Data Acquisition System (DAS), built around a PDP 11/45. These data were used for evaluation of the cross coincidence histograms (Sect. 3.1).

In these experiments we also recorded the neural responses to an ensemble of *tonal* stimuli. Procedures for the analysis of the stimulus-event relation as well as the results of those experiments have been described in Part II

3. The Stimulus-Event Relation

Extra-cellular recording of single unit activity during presentation of an ensemble of stimuli, e.g. a long sequence of natural sounds, results in a sequence of action potentials which can be modelled as a series of neural events:

$$z(t) = \sum_{n=1}^{N} \delta(t - t_n) \tag{1}$$

with $\delta(t)$ = Dirac delta function, $\{t_n\}$ = moments of occurrence of action potentials, N = number of action potentials during presentation of stimulus ensemble (duration T).

In general, repetition of the experiment will not lead to an identical neural activity z(t). Therefore z(t)has to be considered as a particular realization of a stochastic point process.

A functional description of the neuron with respect to its role in the neural representation of sensory stimuli can be given by an analysis of the *stimulus-event relation*. This analysis can be separated logically into 2 components:

1. Does a stimulus-event relation *exist*, i.e. are the statistics of the stochastic point process governed in some way by the stimulus ensemble or is it an autonomous process.

2. If a stimulus-event relation is shown to exist, what is the *nature* of it, i.e. which stimulus properties affect the point process and in which way can their effect be described.

3.1. Existence: Event-Event Correlation

The existence of the s-e relation was investigated by presenting the natural stimulus ensemble twice. The resulting two sequences of neural events $z_1(t)$ and $z_2(t)$ were tested for similarity by evaluation of the Cross Coincidence Function, based on the crosscorrelation function $R_{z_1z_2}(\sigma)$ of the event sequences $z_1(t)$ and $z_2(t)$: event-event correlation (Aertsen et al., 1979)

$$R_{z_1 z_2}(\sigma) = \frac{1}{T} \int_0^T dt \, z_1(t) z_2(t+\sigma) \,. \tag{2}$$

A relatively high number of coinciding events near $\sigma = 0$ indicates that the event sequences are synchronized more than would be expected for independent realizations of an autonomous point process. This synchronization is attributed to stimulus effects: the more distinct the cross coincidence function, the more

strict the coupling between the stimulus ensemble and the neural activity. An example of this analysis for two torus semicircularis neurons is shown in Fig. 2. Each column corresponds to one neuron. The dot displays represent the neural activity during consecutive presentations of the stimulus ensemble, $z_1(t)$ and $z_2(t)$. The form of the cross coincidence function is represented by $N_{1,2}(\sigma)$: the number of coinciding events within a time bin Δ for varying time shift σ . The arrow in the graph of $N_{1,2}(\sigma)$ represents the expectation value $E(N_{1,2})=N_1N_2\frac{\Delta}{T}$ for uncorrelated stationary sequences, it may serve as a reference. Note the highly different values of the binwidth Δ and the time scale σ for the two neurons.

Both units show a distinct peak around $\sigma = 0$, which indicates that indeed the neural activity is influenced reproducibly by presentation of the stimulus ensemble. The width of the peak, measured at half the maximum value, however, differs significantly: about 4 ms for unit 26-6 (Fig. 2, left) versus as much as 250 ms for unit 28-3 (Fig. 2, right). Furthermore we note the regular fine structure in the coincidence histogram for unit 26-6: smaller peaks at time intervals of about 35 ms. This is connected to the rather stereotyped response of this neuron (cf. the dot displays in Fig. 2) to the regular, pulsatile grassfrog vocalizations in the stimulus ensemble (cf. Fig. 1). This fine structure is altogether absent in the firing pattern of unit 28-3; the form of $N_{1,2}(\sigma)$ does not change much when analysed with a smaller binwidth Δ . These findings indicate that, whereas the overall activity N of the neurons differs not more than a factor of about 2.5, there is a dramatic difference in the temporal precision of the coupling between stimulus ensemble and neural activity. This difference is expected to be reflected in the nature of the s-e relation, which will be investigated in the following sections. On the other hand, the form of the coincidence histogram provides a test for the adequateness of a quantitative description of the s-e relation. The two examples shown in Fig. 2 more or less portray the extremes which were encountered in a relatively small population of torus semicircularis neurons which were investigated. A more rigorous statistical approach to the question discussed here has recently been described by Pedersen (1980).

3.2. Nature : Stimulus-Event Correlation

The analysis of the neural representation of natural stimuli is nearly always performed on the base of the Peri- (or Post-) Stimulus Time Histogram (PSTH) (e.g. Kiang and Moxon, 1974; Worden and Galambos, 1974; Bullock, 1977; Smolders et al., 1979): a natural



Fig. 2. Existence of stimulus-event relation. Dot displays and cross coincidence histograms of neural activity recorded from two neurons in the torus semicircularis of the grassfrog during subsequent presentations of the acoustic biotope. Left column: unit 26-6, right column: unit 28-3

stimulus, necessarily of short duration (~ 1 s), is presented repetitively and the associated neural activity is summated in the form of a histogram to yield an estimate of the *event density* n(t) under stimulus conditions. Conspicuous features in the PSTH, i.e. significant deviations from an a priori expected level, are related to the spectro-temporal properties of the stimulus. A necessary prerequisite for this approach is full *controllability* of the stimulus.

This requirement, however, excludes the possibility to analyse the neural activity recorded from normally behaving animals in a natural (acoustic) environment. Therefore we will follow another approach in which the more modest requirement of observability of the stimulus has to be fulfilled. This approach, which makes use of correlation functions of stimulus and neural activity, is based on the concept of the *Pre-Event Stimulus Ensemble* (PESE) (Johannesma, 1972, 1980; Grashuis, 1974; Aertsen et al., 1980). The PESE is defined as the ensemble of stimuli $x_n(\tau) = x(t_n - \tau)$ which precede a neural event $\delta(t-t_n)$ and, by definition, forms a subset of the *Stimulus Ensemble* (SE).



Fig. 3. Definition of the Pre-Event Stimulus Ensemble (PESE)

The concept of the PESE is illustrated in Fig. 3. A comparison of the characteristics of the PESE of a neuron as compared to those of the SE provides a description of the function of that neuron in the neural representation of acoustic stimuli. This approach is rendered feasible by the following observations (Johannesma, 1972):

1. The information on the stimulus-event relation, contained in the PSTH n(t) is also present in the different order cross correlation functions of stimulus x(t) and neural event sequence z(t), or, equivalently, in the different order moment functions of the PESE.

2. The definition of the PESE as well as the determination of its characteristics by means of the moment functions do not require the stimulus to be repetitive. In fact, complex non-repetitive stimuli may be more adequate because of the potential richness of the SE, and, consequently, the PESE. This enables the use of a complex stimulus ensemble like the acoustic biotope for natural stimuli and Gaussian noise in the case of technical stimuli.

3. Since the stimulus does not have to be repetitive a correlation-type of approach does not require full controllability of the stimulus, the observability of the stimulus is sufficient.

These observations indicate that if the experimental situation does not allow the stimulus to be fully controllable, the formation and analysis of the PESE, also for natural stimuli may be an attractive alternative. In the present investigation we have applied this approach and studied the first and second order characteristics of the $s \ e$ relation for a wide natural stimulus ensemble: the acoustic biotope (Aertsen and Johannesma, 1980) by means of *stimulus-event* correlation.

4. The Average Stimulus Preceding a Neural Event

A natural first choice to study the PESE is to determine its average value. The first order moment function or average value $r_e(\tau)$ of the PESE is given by

$$r_e(\tau) = \langle x_e(\tau) \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n(\tau)$$
(3)

with τ = time *before* the neural event ($\tau > 0$), $x_n(\tau) = n$ -th element from the PESE, N = number of action potentials recorded during presentation of the SE (duration T).

This type of analysis was first introduced into sensory neurophysiology as the "revcor" procedure (De Boer and Kuyper, 1968) and used to study the stimulus-event relation for auditory nerve fibres (e.g. De Boer and De Jongh, 1978) and cochlear nucleus neurons (e.g. Grashuis, 1974) for Gaussian white noise. It has been applied in visual electrophysiology as well (e.g. Marmarelis and Marmarelis, 1978). It has been shown that $r_e(\tau)$, apart from a scaling factor, equals the first order crosscorrelation function of stimulus x(t)and event density n(t) (Johannesma, 1972).

In the present investigation the average stimulus preceding an event has been determined by means of a hardware signal averager with an additional delay line to account for the necessary shift ("triggered correlation", De Boer and Kuyper, 1968). The outcome of this procedure for the torus semicircularis unit 26-6 is



Fig. 4a–c. Average pre-event stimulus $r_e(\tau)$ of the torus semicircularis unit 26-6 during first **a** and second **b** presentation of the acoustic biotope, vs. the a priori expected result for this stimulus ensemble **c**, determined by using a random sequence of events. The vertical scale in all three figures refers to arbitrary, identical units. The numbers N_1 and N_2 in the figures indicate the number of events presented to the signal averager (N_1), respectively the number of events actually considered in averaging (N_2)

shown in Fig. 4. In order to test the reproducibility of this result, the acoustic biotope was presented twice and the resulting averages of both recordings are given separately (Fig. 4a and b).

These results should be compared to the a priori expected outcome for the acoustic biotope. The latter is shown in Fig. 4c, using the same amplitude scale as in Fig. 4a and b, to allow for a proper comparison. This result was obtained using a random sequence of events which bears no relation to the SE and which consisted of the moments of positive going high-level crossings of Gaussian wideband noise. Events occur-

ring during the processing sweep of the signal averager are ignored. Due to this "dead time" only N_2 of the total number of elements N_1 in the PESE ($N_2 \leq N_1$) are taken into account in averaging. The numbers N_1, N_2 are indicated in Fig. 4. A comparison of the results in Fig. 4 shows that the average pre-event stimulus of unit 26-6 (Fig. 4a and b) contains oscillations and amplitude variations, which are not present to that extent in the result for random events (Fig. 4c). The most prominent component is an oscillatory waveform with a duration of about 10-15 ms and its maximum amplitude some 15 ms prior to the neural event. The dominant frequency of the waveform is in the order of 500 Hz. This waveform is also present at other values of τ (including $\tau < 0$), now with smaller amplitude and less distinct, as though smeared out. This repetition should be attributed to the quasi-periodic structure of the grassfrog vocalizations in the natural stimulus ensemble (cf. Fig. 1). The gross properties of the average stimulus preceding a neural event in this case appear to be quite reproducible upon repeated stimulation, the reproducibility of the detailed waveform, however, is not very impressive. The latter is not very surprising since phase-lock in this frequency region is hardly expected to be found at this level of the frog's auditory system [cf. the time-locked response to trains of clicks which at the auditory nerve level of the bullfrog already disappeared for a repetition rate of 200 Hz (Frishkopf and Goldstein, 1963)]. Already the coincidence histogram for this unit (cf. Fig. 2, left) indicated a temporal precision of neural firing of as much as 4 ms.

A tentative conclusion from the results in Figs. 2 and 4 might be that unit 26-6 is mainly sensitive for stimulus frequencies around 500 Hz, whereas temporally it might be coupled to a slowly varying parameter of the stimulus, like the envelope or intensity, with a latency of 15-20 ms. This interpretation is preliminary, however, for several reasons. First, the missing of triggers (more that half in this case) may lead to erronous results (e.g. Wilson and Evans, 1975). Secondly, the stimulus ensemble is certainly not like a Gaussian white noise. It is highly structured in time and frequency and the 500 Hz component from the species-specific vocalizations is abundantly present and moreover in a very specific temporal organization (cf. Fig. 1). Third, higher frequencies for which the neuron might be sensitive may have been lost altogether in the averaging procedure as described, due to a poor quality of phase-lock.

For the moment the relevance of first order averaging of the PESE for a natural stimulus ensemble, while using the experimental techniques as described in Sect. 2 (e.g. anaesthesia, simple hardware averaging) awaits further evidence. A possible way to overcome some of the problems indicated here is described in the following section.

5. Spectro-Temporal Averaging of the Pre-Event Stimulus Ensemble

The next obvious step, after a first order analysis of the PESE, would be to proceed with a second order analysis, i.e. to study the second order moment function of the PESE $r_e(\sigma, \tau)$, defined by

$$r_e(\sigma,\tau) = \langle x_e(\sigma) x_e(\tau) \rangle = \frac{1}{N} \sum_{n=1}^N x_n(\sigma) x_n(\tau).$$
(4)

This function is also referred to as the time-dependent autocorrelation function of the PESE. Analogously to the first order average it proves to be proportional to the second order crosscorrelation function of stimulus and neural activity. In the case of a Gaussian white noise stimulus this approach leads to the second order Wiener kernel of the nonlinear system describing the transformation of the stimulus x(t) to the neural activity z(t) (Lee and Schetzen, 1965; Johannesma, 1972). This approach, however, is not intuitively fitted to an understanding of auditory information processing. We prefer an alternative approach, based on a second order representation of sound as a function of both frequency and time: the Dynamic Power Spectrum $P(\omega, \tau)$, also referred to as short-time power spectrum or spectrogram (e.g. Potter et al., 1947; Flanagan, 1965). The dynamic power spectrum $P(\omega, \tau)$ and the time-dependent autocorrelation $r(\sigma, \tau)$ can be formally related by an appropriate Fourier transform (Fano, 1950; Schroeder and Atal, 1962). On the other hand this approach is a generalization of the parametric description of sound in terms of instantaneous frequency and amplitude, which is applicable in the case of narrowband signals (Aertsen and Johannesma, 1980).

We have investigated the spectro-temporal structure of the PESE by measuring the ensemble average of dynamic power spectra, associated with the individual pre-event stimuli: the Average Pre-Event Dynamic Power Spectrum $p_e(\omega, \tau)$, given by

$$p_e(\omega,\tau) = \langle P_e(\omega,\tau) \rangle = \frac{1}{N} \sum_{n=1}^{N} P_n(\omega,\tau), \qquad (5)$$

where $P_n(\omega, \tau)$ indicates the dynamic power spectrum of the *n*-th pre-event stimulus $x_n(\tau)$. The dynamic spectrum analysis has been implemented in hardware in order to enable the analysis to be real time: the *Dynamic Spectrum Analyser* (DSA). Basically it consists of a set of $\frac{1}{3}$ octave filters followed by envelope detectors, the result is the intensity as a function of time for the different frequency components of the



Fig. 5a-c. Average pre-event dynamic power spectrum $p_e(f, \tau)$ of the torus semicircularis unit 26-6 during first **a** and second **b** presentation of the acoustic biotope, vs the a priori expected result for this stimulus ensemble **c**. Results are presented in two representations. 3-D (upper graph) and grey-coding (lower graph) The intensity scale in all three figures refers to arbitrary, identical units The numbers N_1 and N_2 in the figures indicate the number of events presented to the DSA (N_1) , respectively the number of events actually considered in averaging (N_2)

signal. The DSA also offers the facility of spiketriggered averaging of individual spectrograms. A description of the mathematical background and principles of operation is given in Part I (Aertsen and Johannesma, 1980). It should be noted here that the phase characteristics of the bandpass filters in the DSA give rise to a frequency dependent delay in the dynamic spectrum (cf. Fig. 6 in Part I). Obviously this delay should be taken into account when estimating temporal characteristics, like latency. Numerical values throughout the present paper have been corrected for this effect.

The average dynamic power spectrum of pre-event stimuli of unit 26-6 is shown in Fig. 5a in two ways: in a 3-D form and a grey-scale representation. The degree of reproducibility of this result is indicated by Fig. 5b, which was obtained upon repeated stimulation with the natural stimulus ensemble. The a priori expected outcome for this stimulus ensemble is shown in Fig. 5c, which was determined using a random sequence of events. All three averages were scaled to the same, arbitrary intensity scale in order to allow a proper comparison. Like in the first order analysis (Sect. 4) not all events were taken into account in the averaging procedure, due to the "dead time" in the analysis equipment, which equals 29.5 ms in all cases

OVERALL STIMULUS SPECTRUM



Fig. 6. Average spectral intensity density J(f) of the acoustic biotope, determined with the DSA (resolution: $\frac{1}{3}$ octaves), using a random sequence of events and additional averaging over time τ

considered in the present paper. The actual numbers N_1, N_2 ($N_1 \ge N_2$) are indicated in the figures. It is clear that, due to the better instrument specifications far fewer events are lost in this type of analysis than in the first order analysis.

From the results in Fig. 5 we learn that the average pre-event dynamic power spectrum for unit 26-6, using a natural stimulus ensemble, obtains conspicuously high values in the frequency range of about 350-600 Hz, some 20 ms prior to the neural event, as compared to the a priori expected dynamic spectrum. The duration of this elevation is about 15 ms and it is preceded as well as followed by remarkably low values in the dynamic spectrum, over an interval of about the same size. For the very same reason as indicated with relation to Fig. 4 the same elevation appears again for other values of τ , with gradually decreasing amplitude, moreover becoming less distinct. This result, again, points to a rather simple activation type frequency selectivity for unit 26-6, with a best frequency around 400-500 Hz and a latency in the order of 20 ms.

The outcome of this analysis, however, and as a consequence the interpretation of its results, is influenced strongly by the spectro-temporal structure of the stimulus ensemble, as was noted already in the case of first order averaging. The average pre-event dynamic power spectrum should be "normalized" with respect to the spectro-temporal characteristics of the stimulus ensemble in order to arrive at a measure of the spectro-temporal selectivity of the neuron: its "Spectro-Temporal Receptive Field" (Aertsen et al., 1980). In the case of a non-white Gaussian stimulus ensemble this normalization can be obtained simply by a spectral weighting in the bifrequency domain ("postwhitening", e.g. Lee and Schetzen, 1965). In the case of a highly structured stimulus ensemble like the acoustic biotope an appropriate normalization procedure has not yet been formally derived. For the moment we will follow a more intuitive approach, which obviously

cannot lay any claim to formal justification. The primary goal is to compensate for the non-uniform spectral intensity density of the acoustic biotope. To this end an estimate was determined of the overall spectral composition of the stimulus ensemble. In order to enable its use in connection to the results as in Fig. 5, this estimate was determined using the DSA with the settings as in Fig. 5. In order to suppress the variance in the spectral estimate due to the temporal structure of the stimulus ensemble, averaging was performed with a random sequence of as many as about 7500 events. In order to avoid strong dead time effects in the analysis this sequence had a time span of 15 times that of the biotope, with an event rate comparable to that in Fig. 5c, i.e. about 500 events per complete stimulus presentation. The resulting dynamic power spectrum did still contain some temporal variations, this was eliminated by averaging over time τ . The resulting overall stimulus intensity spectrum $J(\omega)$, with a frequency resolution of $\frac{1}{3}$ octaves, shown in Fig. 6, is obviously not "white". The most dominant peak (around 500 Hz) is caused mainly by the grassfrog's vocal repertoire. The low frequency peak (below some 250 Hz) mainly reflects background noise, like wind and water, whereas the high frequency contribution (above some 1250 Hz) mainly originates from vocalizations of other animals. In order to compensate for this non-uniform spectral intensity density of the stimulus ensemble, the average pre-event dynamic power spectrum $p_{e}(\omega, \tau)$ was divided by the overall stimulus spectrum $J(\omega)$ to yield the "stimulus*filtered*" average pre-event dynamic spectrum $q_{e}(\omega, \tau)$:

$$q_e(\omega,\tau) = \frac{p_e(\omega,\tau)}{J(\omega)}.$$
(6)

The outcome of this procedure, applied to the results in Fig. 5 is shown in Fig. 7. The small peaks in the high frequency region (upper 3 filters) of the spectrograms merely reflect numerical instabilities due to the division by relatively small numbers (cf. Fig. 6). The "whitening"-effect of the "stimulus-filtering" can be noted from a comparison of the "random"-results in Figs. 5c and 7c. The main effect on the "neural"-results of unit 26-6 (Figs. 7a, b vs. 5a, b) is a broadening of the frequency range of the domain of high intensity values, without altering the peak frequency value very much. This implies that, also after a correction has been made for the relative abundancy of stimulus intensity in the spectral region around 500 Hz, the average pre-event dynamic spectrum of unit 26-6 maintains the relatively high intensity values in that region. This clearly points to a genuine activation type frequency selectivity for frequencies around 500 Hz. The tuning of unit 26-6 furthermore appears to be simple and quite broad.



Fig. 7a-c. Average pre-event dynamic spectrum divided by stimulus spectrum. Spectro-temporal analysis of the stimulus-event relation of the torus semicircularis unit 26-6, corrected for the spectral composition of the acoustic biotope ("stimulus-filtering"): average pre-event dynamic power spectrum (Fig. 5), divided by the overall stimulus spectrum (Fig. 6)

Obviously the "stimulus-filtering" procedure has not been able to account for stimulus artefacts in the temporal structure of the average dynamic spectrum. This problem remains to be solved, it escapes from the intuitive type of reasoning which was applied here. It even remains to be seen whether a formally derived normalization procedure allows for an uncoupling of spectral and temporal normalization. We expect that a proper normalization will rather turn out to be a combined spectro-temporal normalization, because of the specific spectro-temporal coupling present in the acoustic biotope. Still, we maintain that some type of "stimulus-filtering" as described in (6) should be contained in an appropriate normalization, which is why the results in Fig. 7 should be regarded as a first step towards a naturally determined spectro-temporal receptive field. An interpretation of the results in Fig. 7 would be: activation in a frequency region around 500 Hz with a latency of some 20 ms. Noteworthy is furthermore the distinct temporal boundary of the activation domain, which covers an interval of approximately 15 ms and which is separated from the other elevation domains by "silent" intervals of about the same size. When connected with the stereotyped pulsatile structure of the male grassfrog vocalizations in the stimulus ensemble (note the time scale in Fig. 1) this indicates a rather strict temporal coupling of the neural firing process to the acoustic stimulus. This

finding is in accordance with the results of the coincidence analysis for this unit (Fig. 2). Unit 26-6 has also been investigated with a sequence of short tone pips. This revealed a composite spectro-temporal receptive field for tonal stimuli, containing both an activation and a suppression domain (cf. Figs. 6 and 7 in Part II). The characteristics of the "tonal" and "natural" activation domains match quite well. The suppression for tonal stimuli covers the same frequency range as the activation domain, whereas in time the suppression domain precedes the activation, i.e. it covers larger values of τ in the spectrogram. This points to a "postactivation-suppression" phenomenon : once the neuron has fired to an adequate tonal stimulus, the probability of firing to the same type of stimulus immediately following it (in this case: within some 20 ms) decreases considerably. It is tempting to connect this suppression phenomenon, revealed with tonal stimuli, with the "silent" intervals in Fig. 7 and the temporal structure of the grassfrog vocalizations (short sound elements, separated by silent intervals of about the same size).

Summarizing, the approach described here to analyse the average dynamic spectrum of pre-event stimuli may well be a possible way to investigate the spectrotemporal structure of the single neuron sensitivity domain for acoustic stimulation, i.e. its spectrotemporal receptive field, under natural stimulus conditions. Results of this analysis, however, have to be normalized with respect to the characteristics of the stimulus ensemble used. The "stimulus filtering" as described may be regarded as an intuitive first step towards this goal.

6. Results

The analysis as described in this paper was applied to 24 single unit recordings in the torus semicircularis from 11 grassfrogs. In all cases but one the natural stimulus ensemble was presented twice, in all cases a tonal analysis was performed as well.

For all neurons the *existence* of a stimulus-event relation could be established, whether simply by mere visual inspection of the dot displays or by use of the cross coincidence histogram. Unit 28-3 (Fig. 2, right) was the least clear in this respect.

All recordings were subjected to the first and second order analysis of the PESE as described in this paper. The results for 8 units will be presented as an illustrative sample. The average pre-event stimulus is shown in Fig. 8a-h, together with the result for a random sequence of events (Fig. 8i). The average dynamic power spectrum of pre-event stimuli of these units together with the result for random events is shown in Figs. 9 (3-D representation) and 10 (grey vague (Fig. 9f) or even very vague (Fig. 9g and h). The



Fig. 8a-i. Average pre-event stimulus $r_e(\tau)$ of eight torus semicircularis units a-h, determined with the acoustic biotope, vs. the a priori expected result for this stimulus ensemble i. The vertical scale in all figures refers to arbitrary, identical units. Further details as in Fig. 4

coding). Finally, Figs. 11 (3-D) and 12 (grey) contain the "stimulus filtered" average pre-event dynamic power spectrum for the 8 neurons and the random event sequence. In each figure the results were scaled using the same amplitude respectively intensity scale, to allow for a comparison. The numbers of events, incorporated in averaging (N_1, N_2) are indicated in the figures.

The analysis of the average pre-event stimuli (Fig. 8) in general does not lead to very informative results. The reproducibility upon repeated stimulation appears to be rather low, i.e. in general definitely less reproducible than in the case of unit 26-6 (Fig. 4). Consequently the interpretation of these results remains difficult for the time being, be it merely for the reasons already mentioned in Sect. 4.

The average pre-event dynamic power spectrum (Figs. 9 and 10) gives rise to a variety of results. In all cases one or more (f, τ) -domains can be distinguished with notably higher intensity values. These domains vary from being quite distinct (e.g. Fig. 9b, d, and e) to

elevations cover various frequency regions, also temporally the results differ to quite a degree. The reproducibility of the results upon repeated presentation of the stimulus ensemble, in general, is rather high, be it that the variance increases with decreasing number of neural events. For an interpretation in terms of spectro-temporal sensitivity of the neurons these original averages do not satisfy for reasons explained in Sect. 5. A better impression can be gained from an observation of the "stimulus filtered" average preevent dynamic power spectra (Figs. 11 and 12), weighed for the overall spectral composition of the stimulus ensemble. When compared to the original averages, the effect of "stimulus filtering" varies from only minor (e.g. Fig. 9a and e) to quite drastic (e.g. Fig. 9b and c). Quite clear activation domains appear for high frequencies (e.g. Fig. 9a, c, and e) as well as low frequencies (e.g. Fig. 9b and d), together with units with only a very vague spectral preference (Fig. 9f-h). Latency varies from short (e.g. Fig. 9a, b, and e: about 20 ms) to longer (e.g. Fig. 9c and f: about 60 ms), very long (Fig. 9g: about 100 ms) or even no clear latency (Fig. 9h). In this context it is interesting to note that unit 28-3 with the most vague activation domain (Fig. 9h) corresponds to the very broad coincidence histogram shown in Fig. 2 (right).

Although the "stimulus-filtered" average pre-event dynamic power spectra do not really represent the spectro-temporal receptive fields for natural stimuli, it still is worthwhile to compare the present findings with the main characteristics of the receptive fields of these units, as they were determined using an ensemble of tonal stimuli (Part II: Aertsen et al., 1980). Findings from the latter experiments are summarized in Table 1, more detailed results for the units in Fig. 7 and in Figs. 11 and 12a-d are given in Part II. Most of the information in Table 1 is self-explanatory, the last column deserves some explanation. The firing rate of neurons in the torus to a long ($\sim 8 \text{ min}$) sequence of short (16 ms) tone pips, immediately following each other, appeared to behave in, grossly spoken, two different ways: either it remained constant during the whole sequence ("stationary" units: +) or, after a few seconds of stimulation, it decreased strongly to stay at a very low level or, after some time ($\sim 1 \min$) recovered slightly but never retained the initial level ("long-term adapting" units: -). The characteristics of the tonal receptive fields appeared to be coupled to this phenomenon (cf. Table 1 and Part II). The comparison of these results will receive further attention in the Discussion.

7. Discussion

In the present paper procedures were described to analyse the stimulus-event relation of auditory neu-

N=95-82

N=72-68

N=143-141

1



Fig. 9a-i. Average pre-event dynamic power spectrum $p_e(f, \tau)$ of eight torus semicircularis units a-h, determined with the acoustic biotope, vs the a priori expected result for this stimulus ensemble \mathbf{i} The intensity scale in all figures refers to arbitrary, identical units Further details as in Fig 5





Fig. 10. Grey-scale representation of the results in Fig 9



Fig. 12. Grey-scale representation of the results in Fig 11

Unit	Receptive field type (activation A and, or suppression S)	Best frequency (Hz)	Spectral width (octaves)	Latency (ms)	Temporal width (ms)		Response type
					Activation	Suppression	Listationary (+) or adapting (-)]
26-6	<i>A. S</i>	520	1	20	15	20	+
25-5	A, S	1350	1	20	15	20	+
26-8	A, S	≤ 300	≥15	30	15	60	+
38-1	A, S	1900	0.5-1	60	30	60	_
26-4	A	400	1	40	30	_	_
34-2	Л	1140	05	15	15	_	+
62-2	A	≥2000	≥05	70	30		_
29-3	Α	200-300	1 1 5	120	50	_	-
28-3	_		—	—	—	—	-

Table 1. Characteristics of the stimulus-event relation of nine torus semi-circularis neurons, determined with an ensemble of tonal stimuli (Acrtsen et al., 1980). Numerical values are given in round numbers. Further explanation in text

rons, using an ensemble of natural stimuli. Basically these methods consist of the determination of different types of average measures, defined on the ensemble of stimuli, which precede the occurrence of action potentials. In order to illustrate the concepts, these procedures were applied to a limited number of extracellular single unit recordings from the torus semicircularis of the grassfrog.

Various points concerning the experimental conditions have already been discussed elsewhere (Aertsen et al., 1980), among which the possible effects of anaesthesia (MS-222) and the binaural presentation of stimuli.

The approach as described is based on the concept of the Pre-Event Stimulus Ensemble (PESE), the special selection from the stimulus ensemble, made by the neuron by means of the neural events it fires. The properties of this selection are thought to characterize the neuron's behaviour in the representation of acoustic stimuli. A characterization of this sub-ensemble by means of various average measures results in a global description of the stimulus-event relation, a property which is common to any statistical approach. In fact the various averages represent as many projections of the pre-event stimulus ensemble, each projection emphasizing one or more special aspects of the nature of the stimulus-event relation. Formally the determination of these average measures is identical to the calculation of the cross correlation functions of the neural activity and various functionals of the stimulus. These cross correlation functions play a central role in the Wiener-Lee approach to the theory of non-linear systems (Lee and Schetzen, 1965; Marmarelis and Marmarelis, 1978). Accordingly the procedures as outlined in the present paper and preceding ones (Parts I and II) may be interpreted as elements of a probabilistic formulation of the non-linear analysis of neural systems (Johannesma, 1980).

First Order Averaging of the PESE

Results in the present study of first order averaging of the PESE in general were not very informative. Occasionally the average pre-event stimulus showed a distinct waveform (e.g. Figs. 4, 8b, and 8e). The reproducibility of the results was low. In a number of cases the waveform provided an indication of a possible activation-type frequency selective mechanism. The few units which did produce a distinct waveform in the average pre-event stimulus do not necessarily have to show phase-lock (Lavine, 1971) to the stimulus waveform. The results of the dynamic spectrum analysis (Figs. 5, 9, and 10) indicate that in a number of cases the neural firing behaviour is coupled temporally to the intensity of certain frequency components. This phenomenon, in combination with the stereotyped waveforms of various vocalizations in the natural stimulus ensemble may well give rise to a "virtual" phase-lock in the response. The general correspondence in magnitude between the average pre-event stimuli (Figs. 4 and 8) and the average pre-event dynamic spectra (Figs. 5, 9, and 10) seems to support this possibility. The presence of genuine phase-lock to the stimulus waveform can be investigated by additional stimulation with a sign reversed copy of the stimulus ensemble and a detailed inspection of the form of the cross coincidence histograms of the resulting recordings of neural activity around $\sigma = 0$. It is interesting to note that this same extension of the stimulus ensemble has been suggested in another context: the possibility to improve the statistical properties of pseudo-random noise sequences (Swerup, 1978).

Summarizing: given the limited amount of experimental evidence the results on the whole so far, together with the problems already mentioned in Sect. 4, suggest that for a natural stimulus ensemble the procedure of first order averaging of the PESE of neurons at this level of the auditory system of the grassfrog, under the present experimental conditions, leads to results which are not simply interpretable given our present-day understanding of the auditory nervous system

Spectro-Temporal Averaging of the PESE

The second order analysis of the PESE was performed with respect to the variables of frequency and time These dimensions of sound have traditionally played an important role in the investigation of the auditory system, furthermore they have been shown to be of behavioural significance in several anuran species (e.g. Capranica, 1965, Gerhardt, 1974, 1978, Van Gelder, 1978). Moreover the analysis aims at a combined spectro-temporal representation of the stimulus-event relation instead of separate spectral and temporal descriptions. The arguments for this choice have been discussed in Part II Recent findings indicate that indeed the spectro-temporal receptive fields of units in the grassfrog's auditory midbrain, determined under various stimulus conditions, in general cannot be factorized into separate spectral and temporal characteristics (in preparation)

A representation of sound in terms of frequency and time, which is applicable to any type of signal, is provided by the dynamic power spectrum as implemented in the DSA (cf. Part I). The DSA, like any sonograph, however, shows several shortcomings The spectral resolution, and the temporal resolution coupled to it, have to be selected a priori These characteristics of the frequency analysing mechanism of the DSA appear in the spectro-temporal signal description Phase relations between different frequency components are lost, due to the quadratic operation, inherent to any sonograph. This implies that it is impossible to make an a posteriori trade-off between spectral and temporal resolution The particular selection of filters, used in the present investigation $(\frac{1}{3}$ octave bands), is an arbitrary one, the method of "constant percentage bandwidth" analysis, however, is quite common in auditory research. The combined spectro-temporal resolution of the filters in the DSA, the "uncertainty product" (Gabor, 1946), equals 0.78, which is only slightly above the theoretical limit of 0.5. This is evidently much smaller than the "size" of the activation domains which were found for the torus semicircular units. This suggests that, in view of a combined spectro-temporal analysis, the specifications of the DSA, under the present conditions, are quite satisfactory

Quite another technical limitation of the DSA is formed by the "dead-time" in averaging (here 295 ms, Part I), which results in a missing of neural triggers Thanks to the relatively low firing rates of neurons in the torus semicircularis this did not impose very severe limitations (cf. the numbers N_1 and N_2 , indicated in the Figs. 5, 9, and 10) In cases of higher event rates, however, one should be aware of possible systematic effects A solution to this problem is purely a technical matter, a simple solution might be a repeated analysis of the same data combined with the marking of events, already used in averaging

Several of the spectro-temporal shortcomings of the dynamic power spectrum analysis are the consequence of its nature an instrument-mediated projection of an underlying spectro-temporal intensity density of signals. It can be shown that any DSA can be expressed as a linear functional of an underlying Spectro-Temporal Complex Intensity Density (CoSTID) (Rihaczek, 1968, Johannesma and Aertsen, 1979) This CoSTID as such has no physical meaning but various important signal characteristics can easily be derived from it. A spectro-temporal analysis of the PESE by means of the CoSTID provides a close formal link to the more conventional type of second order cross correlation analysis (e.g. Lee and Schetzen, 1965). This more formal approach to the spectrotemporal description of sound may very well prove to be indispensable in the derivation of adequate stimulus-normalization procedures, necessary to attain a correct representation of the neural spectro-temporal receptive field, which takes into account the a priori knowledge of the characteristics of the stimulus ensemble (Aertsen and Johannesma, in preparation) Preliminary results of simulation studies, following that approach and using as a stimulus ensemble various mixtures of natural stimuli and Gaussian wideband noise, appear to be promising The "stimulus-filtering" as described in the present paper also appears to play an important role in that more formal approach, be it that other aspects (i.e. the autocorrelation structure) of the stimulus ensemble have to be taken into account as well. At the present level of analysis it can be stated that the "stimulusfiltered" average pre-event dynamic spectrum provides a useful first step towards the determination of the neural spectro-temporal receptive field.

Stimulus-Invariance of Receptive Field

A comparison of the present findings for an ensemble of natural stimuli to the outcome of an analysis using tonal stimuli provides an indication to what extent the spectro-temporal receptive field of a neuron still depends on the stimulus ensemble used, after the necessary corrections for stimulus properties have been made The question of interest is, whether the receptive field is possibly completely or partially *stimulusinvariant*. This evidently is a necessary prerequisite if one hopes to reconcile the results for different types of stimuli. In general there is, however, no obvious a priori reason why this stimulus-invariance should hold for such various stimuli as tones, natural sounds or noise. Consequently, the stimulus-invariance should be investigated, rather than be presupposed. This is the main argument for an analysis of the stimulus-event relation for various stimulus ensembles, mutually as remote as noise, tones and natural sounds, where in each case identical, or at least comparable, procedures are followed. At the same time this approach implies that in all cases the stimulus ensemble should be wide enough to encompass all relevant aspects of the particular projection of the *s*-*e* relation one is interested in. In the present case this implies that the natural stimulus ensemble should be constructed as wide as possible in the sense that all sorts of spectro-temporal intensity patterns occur. An ensemble, restricted to species-specific vocalizations will evidently be less adequate in this respect, other natural sounds have to be included as well, possibly together with synthetical variations, to form an acoustic (pseudo-)biotope.

A comparison of the present findings with natural stimuli and the results of the same type of analysis of tonal experiments (cf. Table 1 and Part II) leads to the following observations. The torus semicircularis units which showed a "stationary" response to a sequence of short tone pips (denoted by "+" in Table 1), had a tonal receptive field which, in general, was well defined, relatively small-sized, had short latency and could consist of an activation domain with or without a preceding suppression domain. The units which showed a "nonstationary" response to tones ("long-term adaptation", denoted by "-" in Table 1) had receptive fields which were vaguer, larger (especially in the temporal dimension), had longer latency, and contained an activation domain. Sometimes a preceding suppression domain could be distinguished, but not always was the receptive field clear enough to decide on this. Occasionally "nonstationary" units were found e.g. unit 28-3 for which hardly any sign of a tonal receptive field could be found at all, although the cross coincidence analysis clearly pointed to the existence of a stimulus-event relation. It appears that the results of natural stimulation of the "stationary" units, on the whole, agree quite well with the tonal findings for these units. For the "adapting" neurons the acoustic biotope, in general, appeared to be relatively more effective in eliciting spikes than the tonal sequences, probably due to its more complex character (e.g. non-stationarity, spectral composition). Their naturally determined spectro-temporal selectivity was less distinct than for the stationary units, but in general contained as much as or even more structure than their tonally determined receptive fields did. Especially the temporal

spread was very large. Unit 28-3 (Figs. 11a and 12a), for instance, only showed a clear frequency selectivity (activation around 250 Hz) after the time-window of the analysis was enlarged to several hundreds of ms. This result is in accordance with the large temporal uncertainty in the cross coincidence histogram for this unit (Fig. 2, right). Even at these large values of the time-window the tonal results did not give rise to any spectro-temporal selectivity.

Summarizing: for the "stationary" units the results of tonal and natural analysis agree quite well, the "nonstationary" units are evidently more complex and, at the present level of analysis, a proper comparison of the spectro-temporal selectivity under different stimulus conditions cannot really be made in a satisfactory way. For these latter units this comparison awaits the development of more adequate stimulusnormalization procedures. Only then the outcome truly reflects stimulus dependence of the receptive fields.

A possible way to evaluate the adequateness of the spectro-temporal selectivity of a neuron, remaining within the same stimulus ensemble, might be found by making use of the cross coincidence function (Sect. 3.1). The receptive field of a neuron can be translated, by making use of a probabilistic formulation, into a predictor of neural activity, given a certain spectrointensity pattern temporal of the stimulus (Johannesma, 1980). This predictor might be used to generate a theoretically expected event density upon stimulation with the acoustic biotope. The cross correlation function of this expected event density with an actual recording of that neuron may be compared to the experimental cross coincidence histogram of that same neuron under identical stimulus conditions. This comparison may provide an indication to what degree the receptive field analysis forms an adequate description of the stimulus-event relation for that particular stimulus ensemble, or whether important characteristics may have been overlooked or may be hidden in higher order properties of the PESE.

Acknowledgements. This investigation forms part of the research program of the Workgroup Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (The Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (Z.W.O.).

The authors wish to thank Jos Eggermont and Ton Vendrik for helpful discussions on this manuscript and Koos Braks, Jan Bruijns, Wim van Deelen, and Dik Hermes for expert assistance and/or advice in the experiments and data analysis.

References

Aertsen, A.M.H.J., Smolders, J.W.T., Johannesma, P.I.M.: Neural representation of the acoustic biotope: on the existence of stimulus-event relations for sensory neurons. Biol. Cybern. 32, 175-185 (1979)

- Aertsen, A.M.H.J., Johannesma, P.I.M.: Spectro-temporal analysis of auditory neurons in the grassfrog. In Hearing mechanisms and speech, pp 87-93 Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.), Exp Brain Res. [Suppl. II] (1979)
- Aertsen, A M.H.J., Johannesma, P.I.M : Spectro-temporal receptive fields of auditory neurons in the grassfrog I. Characterization of tonal and natural stimuli. Biol. Cybern. 38, 223–234 (1980)
- Aertsen, A.M.H.J., Johannesma, P.I.M., Hermes, D.J.: Spectrotemporal receptive fields of auditory neurons in the grassfrog. II. Analysis of the stimulus-event relation for tonal stimuli. Biol. Cybern. 38, 235-248 (1980)
- Boer, E. de, Kuyper, P : Triggered correlation. IEEE Trans. Biomed. Eng. BME-15, 169-179 (1968)
- Boer, L. de, Jongh, H R. de: On cochlear encoding: potentialities and limitations of the reverse correlation technique. J. Acoust. Soc. Am. 63, 115–135 (1978)
- Bullock, T II. (ed.): Recognition of complex acoustic signals. (Life Sciences Res. Rep., Vol. 5.) Berlin, Heidelberg, New York: Springer 1977
- Capranica, R.R.. The evoked vocal response of the bullfrog; a study of communication by sound. Cambridge, MA: M.I.T. Press 1965
- Fano, R.M.: Short-time autocorrelation functions and power spectra J. Acoust. Soc. Am. 22, 546-550 (1950)
- Flanagan, J.L : Speech analysis, synthesis, and perception. Berlin, Heidelberg, New York : Springer 1965
- Frishkopf, L.S., Goldstein, M.H., Jr.: Responses to acoustic stimuli from single units in the eighth nerve of the bullfrog. J. Acoust. Soc. Am. 40, 469-472 (1963)
- Gabor, D. Theory of communication. J.I.E.E.(London), Part III 93, 429-457 (1946)
- Gelder, J.J. van, Evers, P.M.G., Maagnus, G.J.M.: Calling and associated behaviour of the common frog, Rana temporaria, during breeding activity. J. Anim. Ecol. 47, 667–676 (1978)
- Gerhardt, H.C.: The significance of some spectral features in mating cell recognition in the green treefrog (Hyla cinerea). J. Exp. Biol. 61, 229 241 (1974)
- Gerhardt, H C. Mating call recognition in the green treefrog (Hyla cinerea) the significance of some fine-temporal properties. J. Exp. Biol. 74, 59–73 (1978)
- Grashuis, J.L.: The pre-event stimulus ensemble: an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons. Ph. D. Thesis, Nijmegen 1974
- Johannesma, P I.M.: The pre-response stimulus ensemble of neurons in the cochlear nucleus. In. Proc. of the IPO Symp. on Hearing Theory, pp. 58–69. Cardozo, B L. (ed.). Eindhoven 1972
- Johannesma, P.I.M., Acrtsen, A.M.H J.: Neural image of sound in the grassfrog. In: Hearing mechanisms and speech, pp. 79–86. Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.). Exp. Brain Res. [Suppl. II] (1979)

- Johannesma, P.I.M.: Functional identification of auditory neurons based on stimulus-event correlation. In: Psychophysical, physiological, and behavioural studies in hearing, pp. 77–84. Brink, G. van den, Bilsen, F.A. (eds.). Delft: Delft University Press 1980
- Kiang, N.Y.-S., Moxon, E.C.: Tails of tuning curves of auditory nerve fibers. J. Acoust. Soc. Am. 55, 620–630 (1974)
- Lavine, R.A. Phase-locking in response of single neurons in cochlear nuclear complex of the cat to low-frequency tonal stimuli. J. Neurophysiol. 34, 467–483 (1971)
- Lee, Y.W., Schetzen, M.. Measurement of the Wiener kernels of a nonlinear system by cross-correlation. Int. J. Control 2, 237–254 (1965)
- Marmarelis, P.Z., Marmarelis, V.Z.: Analysis of physiological systems. The white-noise approach. New York, London: Plenum Press 1978
- Pedersen, J.G.: On the existence of stimulus-event relations for sensory neurons: a statistical method. Biol. Cybern. 38, 201-212 (1980)
- Potter, R.K., Kopp, G.A., Green, H.C.: Visible speech. New York: Van Nostrand 1947
- Richaczek, A.W.: Signal energy distribution in time and frequency. IEEE Trans. Inf. Theory 1T-14, 369-374 (1968)
- Schroeder, M.R., Atal, B.S.: Generalized short-time power spectra and autocorrelation functions. J. Acoust. Soc. Am. 34, 1679–1683 (1962)
- Smolders, J.W.T, Aertsen, A.M.H.J., Johannesma, P.I.M.: Neural representation of the acoustic biotope: a comparison of the response of auditory neurons to tonal and natural stimuli in the cat. Biol. Cybern. 35, 11–20 (1979)
- Swerup, C.: On the choice of noise for the analysis of the peripheral auditory system. Biol. Cybern. 29, 97-104 (1978)
- Wilson, J.P., Evans, E.F.: Systematic error in some methods of reverse correlation. J. Acoust Soc. Am. 57, 215-216 (1975)
- Worden, F.G., Galambos, R. (eds.): Auditory processing of biologically significant sounds (Neurosci. Res. Prog. Bull., Vol. 10.) Brookline, MA: Neurosciences Research Program 1972

Received: September 1, 1980

A. Aertsen

Department of Medical Physics and Biophysics University of Nijmegen Geert Grooteplein Noord 21 NL-6525 EZ Nijmegen The Netherlands

CHAPTER 7

The Spectro-Temporal Receptive Field

A Functional Characteristic of Auditory Neurons

Aertsen, A.M.H.J. and Johannesma, P.I.M.

THE SPECTRO-TEMPORAL RECEPTIVE FIELD

A Functional Characteristic of Auditory Neurons.

A.M.H.J. Aertsen and P.I.M. Johannesma Department of Medical Physics and Biophysics University of Nijmegen Nijmegen The Netherlands.

Abstract. The Spectro-Temporal Receptive Field (STRF) of an auditory neuron has been introduced experimentally on the base of the average spectro-temporal structure of the acoustic stimuli which precede the occurrence of action potentials (Aertsen et al., 1980, 1981).

In the present paper the STRF is considered in the general framework of nonlinear system theory, especially in the form of the Volterra integral representation. The STRF is proposed to be formally identified with a linear functional of the second order Volterra kernel.

The experimental determination of the STRF leads to a formulation in terms of the Wiener expansion where the kernels can be identified by evaluation of the system's input-output correlations.

For a Gaussian stimulus ensemble and a nonlinear system with no even order contributions of order higher than two, it is shown that the second order cross correlation of stimulus and response, normalized with respect to the spectral contents of the stimulus ensemble, leads to the stimulus-invariant spectro-temporal receptive field.

The investigation of stimulus-invariance of the STRF for more general nonlinear systems and for stimulus ensembles which can be generated by nonlinear transformations of Gaussian noise involve the evaluation of higher order stimulus-response correlation functions.

1. INTRODUCTION

One of the possible ways to study the sensory nervous system is the investigation of the input-output relationships of the constitutive elements, the single neurons. A commonly used experimental approach is the extra-cellular recording of single unit activity under a variety of stimulus conditions. Various procedures are used to establish a relation between the sensory stimulus and the trains of action potentials, generated by the neuron. The result

of this approach is a functional description of the neuron as a signal transforming element, participating in the neural representation of sensory stimuli. This functional description can be given in closed form (e.g. a 'transfer function' of some kind) or by means of a set of characteristic parameters (e.g. latency, receptive field properties for visual neurons, best frequency for auditory neurons, etc.). One of the most important issues in this approach is the question whether the functional description of a neuron can be given in a form, such that it comprises the stimulus-response properties for a variety of stimulus conditions. In other words: is it possible to infer the response to a particular type of stimuli from knowledge of the response to another type of stimuli? A long standing issue in auditory electrophysiology is the question whether the responses to 'complex' stimuli, like natural sounds, and the responses to 'simple' stimuli, like tones or clicks, can be reconciled into one satisfactory model (e.g. Worden and Galambos, 1972; Bullock, 1977). For this purpose it evidently is necessary that the functional description of the neuron can be formulated as a complete and stimulus-invariant neural characteristic: it should cover all relevant aspects of the signal transformation in a form from which the stimulus properties have been eliminated, both explicitly and implicitly.

An important contribution to the functional description of sensory neurons has been the concept of the 'receptive field', originally introduced by Adrian et al. (1931) and Hartline (1940) (quoted by Grüsser and Grüsser-Cornehls, 1973). Applied to the visual system, the receptive field of a neuron can be defined as 'the area on the retina from which the discharges of that neuron can be influenced' (Kuffler and Nicholls, 1977). Clearly this original definition of the receptive field is expressed in terms of the receptor surface. Later investigators shifted the attention towards the 2

Unit 161–4 3164 averages



Fig. 1. STRF of a unit determined for a Gaussian wideband stimulus ensemble by averaging the pre-event dynamic spectra and subtraction of the a priori expected stimulus spectrogram. Positive regions are displayed darker than background, negative regions are lighter than background (from Hermes et al., 1981).

neural preference for particular relations in the stimulus domain. Lettvin et al. (1959), in their famous frog-paper, considered the output from the frog's retina as a set of different 'operations on the visual image', where each of the operations was described in terms of 'what common factors in a large variety of stimuli cause response and what common factors have no effect'. In this approach they followed earlier suggestions from Pitts and McCulloch (1947). This suggests that it may be sensible to generalize the receptive field concept to a more abstract formulation as 'the collection of characteristics of stimuli which affect the neural firing behaviour' (e.g. Sejnowski, 1976). In the auditory literature the concept has been used with reference to the spatial properties of adequate acoustic stimuli (Wickelgren, 1971; Knudsen and Konishi, 1978). The present authors have coined the term 'spectro-temporal receptive field' to describe the sensitivity of single auditory neurons with respect to the spectro-temporal intensity density of acoustic stimuli. (Aertsen et al., 1980). A somewhat related, although implicit, use of this concept has been made by Webster and Aitkin (1975).

The spectro-temporal receptive field (STRF) of a neuron has been investigated by averaging the spectro-temporal intensity density of the ensemble of stimuli, preceding the neural events, recorded from that neuron. This intensity density can be measured by means of a dynamic spectrum analyser (DSA; Aertsen and Johannesma, 1980) or by more abstract measures (Johannesma and Aertsen, 1979; Johannesma et al., 1981). This statistical approach to the functional characterization of auditory neurons has been applied to extra-cellular single unit recordings from the auditory midbrain of the grassfrog, using such various stimulus ensembles as tones (Aertsen et al., 1980), a wide ensemble of natural sounds (Aertsen et al., 1981) and Gaussian wideband noise (Hermes et al., 1981). An example of this type of analysis of the neuron's stimulus-response relation is

shown in Fig. 1: the spectro-temporal receptive field of an auditory neuron, obtained by measuring the average pre-event spectrogram for a Gaussian wideband stimulus ensemble (cf. Hermes et al., 1981).

In the present paper we aim to consider the general properties of the 'spectro-temporal receptive field'- concept in the context of the more general theory of nonlinear systems as outlined by Volterra and Wiener (e.g. Marmarelis and Marmarelis, 1978; Schetzen, 1980). The requirement of stimulus-invariance of the STRF-concept will be shown to lead in a natural way to a formulation in terms of the second order Volterra kernel of the system governing the transformation from acoustic stimulus to firing probability of the neuron. The experimental determination of this formally defined STRF, however, requires a formulation in terms of the Wiener kernels, which can be estimated by various order cross correlation functions of stimulus and neural activity. It will be shown that the 'measurement problem' of the STRF imposes certain conditions on both the system and the stimulus ensemble. The conditions become more severe the lower the order of the stimulus-event correlation functions, used to estimate the receptive field.

2. FUNCTIONAL REPRESENTATION OF SYSTEMS

This section deals with the general question of stimulus-invariant system characteristics: the definitions and possible conditions for their existence. The attention will be restricted to single input-single output systems. Furthermore the system is required to be timeinvariant: the structure and parameters of the system do not vary as functions of time.

What is meant by a stimulus-invariant system characteristic is best exemplified by the *linear* system which transforms the input x(t)into the output y(t) according to

$$y(t) = \int d\tau h(\tau) x(t-\tau)$$
(1)

In this familiar linear convolution integral the function $h(\tau)$, known as *impulse response* acts as the kernel of an integral functional which transforms the input x(t) into the output y(t), irrespective of the form x(t) takes: $h(\tau)$ is a stimulus-invariant system characteristic or, shortly, system characteristic. In this particular example of a linear system it furthermore is the only one of interest: it completely characterizes the system. An equivalent formulation can be given in the frequency domain by Fourier transformation of (1):

$$y(\omega) = h(\omega) x(\omega)$$
(2)

where $\omega = 2\pi f$ denotes the angular frequency. In the spectral domain the role of system characteristic is played by the Fourier transform of the impulse response, $\hat{h}(\omega)$, or transfer function. If the linear system is causal this implies that $h(\tau) = 0$ for all $\tau<0$, which in turn implies certain conditions for $\hat{h}(\omega)$ (e.g. Papoulis, 1962).

This result for the linear system can be extended to the case of *nonlinear systems* y(t) = (Sx)(t). If the system S satisfies certain smoothness conditions it can be represented by the functional expansion

$$S(x) = \sum_{n=0}^{\infty} K_n(x)$$
(3)

8

for a bounded input space X, where the K_n are bounded homogeneous polynomial functionals of degree n (Palm and Poggio, 1977a). Furthermore it has been shown that the functionals K_n can be rewritten as Volterra-like integrals

$$(V_n x)(t) = \int d\tau_1 \dots \int d\tau_n v_n(\tau_1, \dots, \tau_n).$$

$$\cdot \prod_{i=1}^n x(t-\tau_i) \quad (4)$$

provided that the integral kernels v_n are allowed to be distributions (i.e. using the Dirac δ -function) to include the representation of e.g. algebraic nonlinearities. This representation of the system S by a functional power series according to Fréchet and Volterra is analogous to the Taylor expansion of ordinary functions (e.g. Barrett, 1963; Palm and Poggio, 1977a; Hung and Stark, 1977). The Volterra-type power series expansion can be shown to be unique, provided that all the kernels $v_n(\tau_1, \ldots, \tau_n)$ are made symmetrical with respect to the arrangement of the variables τ_i , which can always be accomplished by a summation over all permutations of the suffices (Barrett, 1963), leading, to the appearance of an additional factor $\frac{1}{n}$, in the formulation used by some authors (e.g. Bedrosian and Rice, 1971). This type of symmetry in the temporal domain implies the same symmetry in the frequency domain.

The n-th order Volterra functional $(V_n x)(t)$ completely describes the contribution of the *n-th* order system nonlinearity to the output y(t). It is expressed as an n-fold convolution integral of the input x(t) and the integral kernel $v_n(\tau_1, \ldots, \tau_n)$, the *n*-th order Volterra kernel. The kernel v_n completely characterizes the n-th order nonlinearity in the system. Since it is fully determined by the system function S, irrespective of the input x(t), the Volterra kernel $v_n(\tau_1,\ldots,\tau_n)$ can be considered as an n-th order stimulus invariant system characteristic. In the case of a linear system the expansion (3) reduces to the linear convolution (1), the first order Volterra kernel $v_1(\tau)$ then equals the impulse response $h(\tau)$. In analogy to the linear case the higher order Volterra kernels are sometimes referred to as higher order impulse responses, their n-fold Fourier transforms as higher order transfer functions (e.g. Alper and Poortvliet, 1964; Brillinger, 1970; Bedrosian and Rice, 1971).

If the summation in (3) extends over only a finite number of terms, say N, the power series expansion reduces to a functional polynomial. In that case S is called a polynomial system of order N. The practical importance of the polynomial system is evident: if all (=N) Volterra kernels are known, the output to any input-signal is fully specified because of 1) the completeness and 2) the stimulus-invariance of the reasons why polynomial systems are widely used for the representation and approximation of nonlinear systems (e.g. Palm, 1978).

3. EXPERIMENTAL DETERMINATION OF THE SYSTEM CHARACTERISTICS.

The functional (or 'black-box') approach to the analysis of unknown systems, technical or biological, aims to determine the system kernels, introduced in the foregoing section. If there is a priori evidence for linearity and timeinvariance of the system, several methods are available to determine the system characteristic: direct measurement of the impulse response or determination of the transfer function by harmonic analysis. An attractive alternative is the cross correlation of the system's input and output for a wideband stationary input ensemble. In the frequency domain this results in

$$\hat{h}(\omega) = \frac{\hat{R}_{xy}(\omega)}{\hat{R}_{xx}(\omega)}$$
(5)

with $R_{xy}(\omega)$ the cross power spectrum of input and output and $R_{xx}(\omega)$ the input power spectrum (Lee, 1960; Papoulis, 1965). From (5) it is obvious that $R_{xx}(\omega)$ should not equal zero, which implies that information about $h(\omega)$ can only be gained for frequencies which are present in the stimulus ensemble. The cross correlation approach has quite fruitfully been applied the peripheral auditory nervous system for Gaussian wideband noise as stimulus ensemble (De Boer and Kuyper, 1968; Møller, 1973; Grashuis, 1974; Van Gisbergen et al., 1975; De Boer and De Jongh, 1978). First order cross correlation has been applied by the present authors to the neural activity from the grassfrog's auditory midbrain for a natural stimulus ensemble (Aertsen et al., 1981). The results in that case, however, were not very informative.

Many biological systems, including ones encountered in neurophysiology are not linear. This emphasises the need for the experimental determination of higher order system characteristics. When a priori information about the structure of the system is available, methods to determine the system kernels have been developed (for a review see Hung and Stark, 1977). For the case of a general unknown system, however, schemes for the experimental determination of the Volterra kernels have not been developed so far, which is connected to the lack of orthogonality of the Volterra functionals V_n . Wiener (1958) succeeded in rearranging the Volterra functionals in the power series expansion of a time-invariant, finite memory, analytic nonlinear system as given in (3) in such a way as to obtain a new expansion

$$y(t) = \sum_{n=0}^{\infty} (W_n x)(t)$$
(6)

where the new functionals, the Wiener functionals W_n , are orthogonal with respect to a Gaussian white input process, where orthogonality is defined in terms of the expected value:

$$E \{ (W_n x)(t), (W_m x)(t) \} = 0$$

for all t; $n \neq m$ (7)

(e.g. Lee, 1964; Marmarelis and Marmarelis, 1978).

The Wiener functionals W_n , obtained by a Gram-Schmidt type orthogonalization procedure applied to a Gaussian white stimulus ensemble turn out to be inhomogeneous polynomial functionals: W_n is a linear combination of the V_n , V_{n-2} , V_{n-4} etc. Associated with the Wiener functionals are new integral kernels, the Wiener kernels $w_n(\tau_1,\ldots,\tau_n)$. Like the Volterra kernels the Wiener kernels are symmetrical with respect to the arrangement of their arguments. One of the basic advantages of the Wiener expansion is the possibility of experimental determination of the various kernels, thanks to the orthogonality of the functionals. Various identification schemes have been put forward, apart from the Laguerre expansion procedure proposed by Wiener himself. The most commonly applied ones are based on higher order cross correlation functions of the system's input and output signals (Lee and Schetzen, 1965; Schetzen, 1974) or, equivalently, cross spectra in the frequency domain (French and Butz, 1973).

The Wiener approach of orthogonalizing the functional expansion with respect to a Gaussian white noise input (in fact to the integrated process, the Brownian motion) has been adopted and applied to other types of stimulus ensembles as well, e.g. Gaussian pseudo-noise, based on maximum length pseudo-random sequences (e.g. Hewlet Packard 3722), quasi-white signals with a symmetric probability density function (Marmarelis, 1977), Poisson process impulse sequences (Krausz, 1975; Brillinger et al., 1976), Markov chain inputs (Kroeker, 1980) and signal ensembles, generated as the sum of incommensurate sinusoids (Victor and Knight, 1979). In all these cases an orthogonal functional series expansion can be derived, with Wiener-type kernels associated with the various functionals. For a review of deeper mathematical problems, associated with various functional expansions and the identification

of the kernels (e.g. the domain of convergence) the reader is referred to Palm and Poggio (1977a, b) and Palm (1978).

Although related in origin there appear to be remarkable differences between the Volterra representation and the Wiener-type representations. Where the *n-th* order Volterra functional V_n completely describes the n-th order nonlinearity in the system, the inhomogeneity of the Wiener functionals causes this nonlinearity to be distributed over the various functionals W_n , W_{n+2} , W_{n+4} ..., or, conversely, the *n-th* order Wiener-type functional W_n does not represent completely the n-th order nonlinearity in the system. Another distinction can be found when studying systems which can be considered to be built from more elementary systems such as cascade systems, systems with feedback etc. It has been possible to derive explicit and relatively simple expressions for the Volterra kernels of the compound system in terms of the Volterra kernels of the constitutive elements: algebraic expressions in the frequency domain versions of the system kernels (Barrett, 1963). The inhomogeneity of the Wiener functionals forms an obstacle in deriving similar expressions for the Wiener kernels. Marmarelis and Marmarelis (1978) in fact describe an application of the Volterra formalism to the Wiener kernels of a neuron model, which in that case was only allowed because the nonlinearities in that system were supposed to be not higher than second order, causing the Volterra and Wiener kernels to be identical. In a more general model this interchange of Volterra kernels and Wiener kernels, however, is no longer allowed. Another important distinction in the present context is that, while the Volterra kernels are independent of input properties, the Wiener-type kernels are inherently coupled to the input ensemble: orthogonality is only guaranteed for the test input process used, with its specific mean and variance. Obviously the particular definition of orthogonality which was adopted also influences the outcome of the Gram-Schmidt procedure. This input-coupling is reflected in the 'specific input problem': the Wiener-type representation is valid and shows the strongest convergence of all possible representations for the class of input signals the system was tested with, however no a priori guarantees can be given regarding the behaviour of the expansion, especially of truncated versions of it, when any other specific input signal is used (Palm and Poggio, 1977b; Johnson, 1980).

The foregoing considerations suggest that for the investigation of neural responses to such various stimuli as noise, tones and natural sounds a system representation in terms of the stimulus-invariant Volterra formalism offers more possibilities of reconciling the various experimental results than the Wiener-type representations do. The Volterra kernels, however, are not directly accessible for experimental determination. A possible way out of the dilemma might be to use the Volterra

œ

expansion for *representation* of the system, while using the Wiener kernels and especially the cross correlation or cross spectrum techniques, for *identification*. This compound approach then should be combined with an algorithm to extract the Volterra system kernels from the experimentally determined Wiener kernels.

4. SPECTRO-TEMPORAL RECEPTIVE FIELD AND SYSTEM CHARACTERISTICS.

If the necessary convergence criteria for the V- and W-series are fulfilled, it is possible to give an explicit algebraic expression of the Wiener kernels w_n in terms of the Volterra kernels v_n , v_{n+2} , v_{n+4} , ... and, conversely, of the Volterra kernels v_n in terms of the Wiener kernels w_n , w_{n+2} , w_{n+4} , ... (Yasui, 1979). The Volterra kernels are a function of the Wiener kernels as expressed in (8):

$$v_{n}(\tau_{1},\ldots,\tau_{n}) = \sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{(n+2\nu)!}{n!\nu!} \left(\frac{P}{2}\right)^{\nu} .$$

$$\cdot \int d\sigma_{1} \cdots \int d\sigma_{\nu} w_{n+2\nu}(\tau_{1},\ldots,\tau_{n},\sigma_{1},\sigma_{1},\ldots,\sigma_{\nu},\sigma_{\nu})$$
(8)

where P denotes the power spectral density of the Gaussian white noise for which the Wiener functionals have been orthogonalized. The reciprocal relation is obtained by interchanging v and w and omitting the alternating factor $(-1)^{v}$.

Relation (8) in principle offers the possibility to obtain the system kernels by appropriate combination of the Wiener kernels which can be measured by cross correlation techniques. This result, however, is not very useful in practice because the summation in (8) extends to infinite order Wiener kernels. For a continuous system S it will always be possible to obtain a satisfactory approximation of the system function by a polynomial nonlinearity of some finite order N (Weierstrass' theorem); in that case $v_n = w_n = 0$ for all n > N. This result, in combination with (8) leads to an *algorithm* for the calculation of the system kernels v_n :

- the even order system kernels v_{2n} can be obtained from the even order cross correlation functions of order 2n, 2n + 2, ..., 2M, where M = entier (N/2).
- the odd order system kernels v_{2n+1} can be obtained from the odd order cross correlation functions of order 2n + 1, 2n + 3, ..., 2M + 1.

The immediate practical use of this algorithm for the moment is limited to fairly low order nonlinearities because of the rather elaborate computations involved.

An immediate result of (8) is that for a system of order N it holds that $v_N = w_N$ and $v_{N-1} = w_{N-1}$, as was already noted for the special case of a second order system. Another, although related, result from (8) and its reciprocal version is that the contribution to the even resp. odd order system kernels only order from the higher even resp. odd order cross correlation functions. For instance, no matter how large the order N, be it 3 or 27, as long as the highest even order nonlinearity in the system is of order 2, the second order system kernel can simply be measured by evaluation of only the second order cross correlation function. The representation of the system in that case is evidently not complete, so output predictions will surely fail, nevertheless a stimulus invariant system characteristic has been derived relatively easily and can be evaluated independently from the system characteristics that remain to be identified.

Summarizing:

- The first order stimulus-invariant system kernel v_1 of a finite order nonlinear system S can be determined from the various odd order (1, 3, 5,...) cross correlation measurements of a single Gaussian noise experiment. It can be determined by a single first order cross correlation function for systems of order up to 2, with possible additional higher even order nonlinearities (4,6,8,...).
- The second order stimulus-invariant system kernel v_2 of a finite order nonlinear system S can be determined from the various even order (2, 4, 6,...) cross correlation measurements of a single Gaussian noise experiment. It can be determined by a single second order cross correlation function for systems of order up to 3, with possible additional higher odd order nonlinearities (5, 7, 9,...). These results can immediately be extended to the general case of the *n*-th order system kernels.

Having discussed the existence and possible determination of general stimulus-invariant system characteristics we now return to the original question regarding the concept of the spectro-temporal receptive field (STRF). By the choice of the dimensions of frequency and time the STRF clearly is a second order system characteristic. In view of what has been said about stimulus-dependence we propose to let the second order Volterra kernel $v_2(\tau_1, \tau_2)$, or rather its single Fourier transform $\vartheta_2(\omega,\tau)$ be the source of the STRF. This choice implies that a stimulus-invariant STRF of a neuron can be defined, i.e. it exists, provided that the neural system function meets the conditions of a convergent Volterra series expansion. This imposes a priori assumptions about the system: it is required to be time-invariant, have finite memory and should be sufficiently smooth in the stimulus-domain of interest. These requirements in general are not trivially met:





а



- Fig. 2. Schematic diagram of neural system's analysis.
 - 2a. The transformation of acoustic stimulus x(t) into neural response y(t).
 - 2b. The cascade system T formed by the tandem connection of stimulus machine M and neuron system S. Further explanation in the text.

for instance threshold-like system functions are clearly excluded (Palm and Poggio, 1977b). When considering the 'smoothness' of the neuron's response function it should be noted that this applies to the probability of firing, or event density, and not to the actual firings themselves. The experimental *identification* of the STRF requires additional conditions on the system function. For a finite order nonlinearity the STRF in principle can be determined from the Wiener kernels by cross correlation analysis of a Gaussian noise experiment. The present-day available computational technology necessitates even further restrictions: the STRF can be measured from a single second order cross correlation function for neurons with at most a second order even nonlinearity in the system function. It should be realized, however, that the additional conditions mentioned here have to be imposed because of a 'measurement problem'. In the sense of abstract existence of the STRF there is no inherent necessity of them whatsoever.

5. STIMULUS NORMALIZATION OF THE STRF.

The possibility of experimental determination of the Volterra type STRF by means of cross correlation analysis of the system's in- and output has been discussed in Sect. 4, mainly in connection with a Gaussian white process as stimulus ensemble. In the present section we will consider the question whether the Volterra-type STRF can be considered as a stimulus-invariant neural characteristic, or, for which classes of stimuli can the spectro-temporal receptive field, disposed of obvious stimulus properties, be considered as a single model of neural signal representation.

The effect of the characteristics of the stimulus ensemble on the outcome of cross correlation of the system's in- and output can be separated into two components: (1) the direct effect that the probability density function of various spectro-temporal intensity patterns varies for the different stimulus ensembles, which directly influences the result of cross correlation analysis, and (2) the indirect effect that the overall characteristics of the stimulus ensemble may force the neuron into a specific state of responsiveness: the neuron's system function is not stimulus-invariant. In order to investigate the effect (2) the result of cross correlation should be normalized with respect to the direct stimulus influence (1). Only then it is possible to make reliable statements regarding the stimulus-dependence of neural characteristics like the spectrotemporal sensitivity.

The experimental situation is schematically depicted in Fig. 2a.

The system S indicates the total transformation which takes place from the acoustic waveform x(t) into the neural response y(t). The acoustic stimulus x(t) may belong to any type of stimulus ensemble SE, e.g. noise, tonal, natural etc. In order to establish a relation with the formal theory we make the following assumption: for each particular stimulus ensemble SE it is possible to define a machine M, such that M is able to generate the SE from Gaussian white noise. Furthermore it is required that M, like S, satisfies the conditions necessary to apply the Volterra representation of nonlinear systems (Sect. 2). This extension of the actual experiment is depicted in Fig. 2b, where now the composite system T = MS is investigated with a Gaussian white stimulus ensemble. This approach is inspired by the procedure used by Lee and Schetzen (1965).

It has been shown that the Volterra kernels of a cascade system can be expressed in terms of the Volterra kernels of the constitutive elements (Barrett, 1963). For the first and second order kernels we have in the frequency domain:

$$\hat{v}_1^T(\omega) = \hat{v}_1^M(\omega) \cdot \hat{v}_1^S(\omega)$$
(9)

$$\hat{\hat{v}}_{2}^{T}(\omega, v) = \hat{v}_{1}^{M}(\omega) \hat{v}_{1}^{M}(v) \hat{\hat{v}}_{2}^{S}(\omega, v) + \\ + \hat{\hat{v}}_{2}^{M}(\omega, v) \hat{v}_{1}^{S}(\omega+v)$$
(10)

It has been assumed here that $v_o^M = v_o^S = 0$ (zero input gives zero output). The first order cascade kernel thus equals the linear *S*-kernel weighed with the linear *M*-kernel. The second order cascade kernel is composed of two terms: the quadratic *S*-kernel, doubly weighed with the linear *M*-kernel plus the linear *S*-kernel, weighed with the quadratic *M*-kernel. This result can be extended to the higher order Volterra kernels.

In the present context the main interest is in determining v_2^S , which forms the basis of the spectro-temporal receptive field concept. The combination of (9) and (10) leads to

$$\hat{\hat{v}}_{2}^{S}(\omega, v) = \frac{\hat{\hat{v}}_{2}^{T}(\omega, v)}{\hat{v}_{1}^{M}(\omega)\hat{v}_{1}^{M}(v)} - \frac{\hat{\hat{v}}_{1}^{T}(\omega+v)}{\hat{v}_{1}^{M}(\omega)\hat{v}_{1}^{M}(v)\hat{v}_{1}^{M}(\omega+v)}$$
(11)

provided that the denominators do not equal zero. Depending on the experimental context, several ways to proceed are conceivable.

I. Suppose the machine M is a real. one, especially designed to generate, from Gaussian white noise, a stimulus ensemble which has properties that are attractive from the experimenter's point of view. That might, for instance, be a neuro-ethological interest: noise is transformed into quasi-natural sounds. Furthermore, we assume, the machine has been designed as a polynomial system, with the system kernels v^M known. Since in this case both the noise input u(t) and the neural response y(t) are observable it is possible, in principle, to calculate the various order cross correlation functions $R_{u...uy}$. This enables the determination of the various Wiener kernels of the cascade w^T , which, by application of (8) leads to the Volterra kernels v^T . Substitution into the equations (9), (10) and their higher order equivalents thus offers the opportunity to determine all the S-kernels, with v_2 ^S in particular from (11). Obviously in practice the order of both M and S should be as low as possible (with only *M* liable to direct influence) to avoid an enormous amount of computer work on the correlation functions. Although perhaps unfeasible at the moment, this approach in principle offers the possibility to reconcile the use of e.g. neuro-ethologically 'relevant' signals with a formal, system theoretical type of analysis of the stimulus-response relation. The 'bio-acoustical dimensions' of sound (Scheich, 1977) thus might be explored, still retaining the connection with formal theory.

II. Another approach starts from the stimulus ensemble as such. Suppose that closer inspection of the ensemble of interest, e.g. a wide ensemble of natural sounds, indicates that the ensemble can be imagined to be generated by a probability density function which is symmetrical with respect to the stimulus waveform: f(x) = f(-x), with x the vector representation in signal space of the acoustic stimulus x(t) with a duration which is long as compared to the time constants in S. This symmetry property, we assert (without proof), is equivalent to a machine M with only odd order system kernels: $v_2^M = v_4^M = \ldots = 0$. In that case the relation (11) reduces to

$$\hat{\hat{v}}_{2}^{S}(\omega, v) = \frac{\hat{\hat{v}}_{2}^{T}(\omega, v)}{\hat{\hat{v}}_{1}^{M}(\omega)\hat{\hat{v}}_{1}^{M}(v)}$$
(12)

If now 1) either v_1^M is known or it can be determined from observation of u(t) and x(t)and the analysis of the odd order functions $R_{u...ux}$ and furthermore 2) v_2^T can be determined from the observation of u(t) and y(t)and the analysis of the even order cross correlation functions $R_{u...uy}$, then the second order system kernel v_2^S can be calculated from (12). The principal advantage in this case as compared to the general relation (11) is that for the determination of v_2^S now only the even order cascade correlation functions are needed. A similar advantage in this case holds for v_3^S (only odd order cascade correlations), it does not hold for the higher order S-kernels, although also there the amount of contributing terms is strongly reduced.

III. A very rigorous simplification can be obtained, finally, if we replace the Volterra kernels in the right-hand side of (12) by their corresponding Wiener kernels, to give

$$\hat{\hat{v}}_{2}^{S}(\omega, v) = \frac{\hat{\hat{w}}_{2}^{T}(\omega, v)}{\hat{w}_{1}^{M}(\omega)\hat{w}_{1}^{M}(v)}$$
(13)

The aim of this replacement is to restrict the number of cross correlation functions, necessary for the identification of the noise-stimulated systems in Fig. 2b (i.e. M and T) to the extreme limit of one: a single first order correlation R_{uxy} for M and a single second order correlation R_{uuy} for T. This simplification indeed is a very severe one, since some combinatorics, using the algebra for the Volterra kernels of cascade systems (Barrett, 1963) shows that the transition from (12) to (13) implies that:

- 1) the machine M should be linear: $v_n^M = 0$ for all n other than 1, and
- 2) S and (as a consequence) T should not contain even order nonlinearities of order higher than 2: $v_{2n}^{S} = v_{2n}^{T} = 0$ for all $n \ge 2$.

The simplification (13) thus reduces M to a linear filter: the stimulus ensemble should be 'coloured' Gaussian noise (cf. Lee and Schetzen, 1965). Furthermore the restriction on S implies that in (13) v_2^{S} may as well be replaced by the Wiener kernel w_2^{S} . Note that this very condition on S was already derived in Sect. 4, when the experimental determination of the neuron's STRF from a single second order cross correlation function for a Gaussian white stimulus ensemble was discussed.

The restrictions in case III allow to derive from (13) a very simple stimulus-normalization procedure for the neuron's STRF (Schetzen, 1974)

$$\hat{\hat{v}}_{2}^{S}(\omega, v) = \frac{\hat{\hat{R}}_{xxy}, (\omega, v)}{2\hat{R}_{xx}(\omega)\hat{R}_{xx}(v)}$$
(14)

where y'(t) denotes the deviation of the system's response y(t) from its average value during the present stimulus conditions: y'(t) = y(t) - E(y). The latter correction is only effective for $\omega + \nu = 0$, in which case it effectively results in an additional subtraction of the stimulus spectrum:

$$\hat{v}_{2}^{S}(\omega, v) = \frac{\hat{R}_{xxy}(\omega, v)}{2\hat{R}_{xx}(\omega)\hat{R}_{xx}(v)} - \frac{\hat{R}_{xxy}(\omega, v)}{2\hat{R}_{xx}(\omega)\hat{R}_{xx}(v)}$$
(15)

In (14) and (15) the quadratic S-kernel is expressed in the second order cross spectrum of the system's in- and output, 'normalized' with respect to the stimulus ensemble by double division by the stimulus spectrum at the appropriate frequency values. Finally, it should be noted that (14) and (15) describe the quadratic Volterra kernel, which in this case is identical to the second order Wiener kernel for Gaussian white noise. The Wiener kernel, corresponding to the coloured noise, however, has a different form, involving the spectral factorization of R_{xx} (Lee, 1964; Lee and Schetzen, 1965).

6. DISCUSSION

The concept of the Spectro-Temporal Receptive Field (STRF) of an auditory neuron was introduced (Aertsen et al., 1980) in order to give a concise description of stimulus parameters which appear to be relevant for the neuron's firing probability to various acoustic stimuli. Basically it may be conceived as the 'spectacles' in the frequency-time domain through which the neuron 'views' the acoustic environment. The STRF was originally defined in operational terms: the average spectro-temporal structure of stimulus elements which in time precede the occurrence of action potentials. Depending on the type of stimuli (narrowband vs. wideband), various procedures for the experimental determination of the neuron's spectro-temporal sensitivity have been developed (Aertsen et al., 1980, 1981; Hermes et al., 1981). In the present paper an attempt was made to connect this intuitive concept of the STRF with the formal theory of nonlinear systems. In that context various aspects of



- Fig. 3. Schematic diagram of the system theoretical context of the spectro-temporal receptive field STRF.
 - 3a. General scheme of relations between Volterra and Wiener approach to nonlinear systems.
 - 3b. Separation of even and odd order contributions.
 - 3c. Special case of a system with only a quadratic even order component. Further explanation in the text.

the STRF have been considered:

- existence: under what conditions is it possible to define in a formal way something like a STRF of an auditory neuron.
- measurability: under what conditions is it possible to determine the STRF experimentally.
- 3. stimulus-dependence: under what conditions can the descriptions of spectro-temporal sensitivity of a neuron, determined for different types of stimulus ensembles, be connected in a formal way to give the neuron's STRF which is disposed of direct stimulus effects: 'stimulus-normalization'.

Existence and Measurement

It has been shown that the combination of 1) the second order (i.e. spectro-temporal) transfer function-like character and 2) the desired intrinsic stimulus-invariance of the STRF leads to the proposal to identify the STRF with the frequency-time representation of the second order Volterra kernel of the neuron's system function. As a consequence the STRF of a neuron can be defined provided that the requirements for convergence of the Volterra expansion are fulfilled.

The possibility of experimental determination of the STRF leads to the requirement of a system function behaving as a finite order nonlinearity, such that it can be represented by a polynomial system, combined with a formulation in terms of the Wiener kernels. In that case the STRF in principle can be determined from a finite number of correlation functions of the system's in- and output for a Gaussian stimulus ensemble, as schematically indicated in Fig. 3a.

Because of the explicit relations between the Volterra and Wiener kernels (cf. (8)), the STRF, like any other even order system kernel, can be determined from the even order cross correlation functions, whereas the odd order kernels can be measured by evaluating the odd order correlation functions. The terms 'summated' and 'compound' correlation in Fig. 3b serve to indicate the relation with the summated and compound PSTH (Goblick and Pfeiffer, 1969; Johannesma, 1971) as estimators of the even resp. odd part of the neurons firing probability.

In case the even part of the system function contains only a quadratic term, the STRF can be measured from a single second order correlation for a Gaussian input, irrespective of the form the odd part of S takes (cf. Fig. 3c). This situation is directly connected to the operational definition of the STRF given hitherto and the analysis procedures associated with it.

Stimulus-Invariance

It has been shown that for those stimulus ensembles that can be generated from Gaussian white noise through transformation by a finite order machine M, the neuron's STRF in principle can be determined, disposed of the direct effects of the spectro-temporal structure of the stimulus ensembles: stimulus-normalization. The introduction of M, which should preferably be of low order to minimize computational effort, evidently imposes a restriction on the stimulus ensembles which come into consideration. Further investigation into the possible range of stimuli to be generated in this way, particularly in view of the ideas about artificial, quasi-natural sounds originating from neuro-ethology, appears to be worthwile in this context.

For the simple situation, indicated in Fig. 3c, combined with a linear machine *M* (i.e. the stimulus ensemble consists of 'coloured' Gaussian noise) the stimulus-normalization appears to be a straightforward double spectral division of the system's second order cross spectrum, in fact already a result from Lee and Schetzen (1965).

After stimulus-normalization has been applied, the neuron's STRF can be used to investigate the possible stimulus-effects on the neuron's firing behaviour. The STRF then may turn out to be completely stimulus-invariant i.e. identical for various technical and natural stimuli, or it may be invariant only over a restricted range of stimulus variation. The idea of stimulus invariance is strongly connected to the procedure of output-prediction, although certainly not equivalent to it. For a succesful prediction of the response to a particular type of stimuli from knowledge of the response properties to other stimuli, the invariance over the specified stimulus ensembles of system characteristics like the STRF is a necessary prerequisite. It is however not a sufficient condition since the STRF only provides the second order contribution to the response estimation. The more and the stronger other contributions in the neuron's system function are present, the more the prediction in general is bound to deviate from the actual response. Conversely, if the predicted response fails to meet the actual response this points to one or both of two possible reasons: the STRF is not stimulus-invariant over the specific stimulus range and/or it does not represent the complete or dominating term in the system function. The STRF provides a faithful representation of the system's quadratic component within the range of its stimulusinvariance.

In case an adequate stimulus-normalization procedure cannot be applied, because of theoretical or computational restrictions, the descriptive value of the spectro-temporal sensitivity as measured by correlation analysis is, strictly spoken, limited to the domain of stimuli it was determined with. There are no formal guarantees of its significance outside that particular stimulus ensemble. However, this should not taken to imply that nothing could be learned from setting foot on theoretical terra incognita, where educated guesses and great care in the interpretation of results should be used side by side.

Additional remarks

An interesting aspect of the STRF not discussed so far is its *reducibility*: does the STRF provide a minimal representation of the quadratic system component or can it be decomposed into separate constitutive elements. It might for instance prove possible to factorize the STRF into separate spectral and temporal components. There are no a priori grounds for reducibility, the results of experimental analysis should provide the arguments in every particular case (Eggermont et al., 1981).

The STRF-concept should not be taken to necessarily comprise, for all neurons, all stimulus properties, relevant for the neuron's firing behaviour. Fundamental questions regarding auditory perception may be condensed into 'what' and 'where', respectively concerning the 'identification' and the 'localization' of a sound source. The present discussion as well as the experiments that inspired it were primarily focussed on the identification properties of the auditory system. Therefore it may well be conceivable that the STRF for some neurons should rather be viewed as a particular cross-section of a multi-dimensional (e.g. a spatio-spectro-temporal) domain in stimulus space, reflecting the neuron's sensitivity function.

The analysis of the STRF forms a specific example of the more general approach of stimulus functional correlation (Johannesma, 1980). Other examples have been described in the literature (e.g. Møller, 1973; De Boer, 1979). A qualitative formulation of this statistical approach can be found as early as 1959 in the visual literature (Lettvin et al., 1959): 'We should present the frog with as wide a range of visible stimuli as we could, not only spots of light but things he would be disposed to eat, other things from which he would flee, sundry geometrical figures, stationary and moving about, etc. From the variety of stimuli we should then try to discover what common features were abstracted by whatever groups of fibers we could find in the optic nerve..... What, then, does a particular fiber in the optic nerve measure? We have considered it to be how much there is in a stimulus of that quality which excites the fiber maximally, naming that quality'. The actual choice of the stimulus funtional used in the correlation reflects some of the basic model assumptions about the system under investigation. The choice of a spectro-temporal analysis leading to the STRF implicitly supposes the dynamic spectrum analysis to form an important contribution to the neuron's system function. If, for instance, information is available that the neural transducer mechanism contains nonlinearities like exponentials it might be wise to reconsider the approach using a polynomial representation and, instead, to use this information in the formulation of the relevant stimulus (or response-) functional. All relevant information regarding possible models should be used in order to lend additional support to the functional analysis of neural function which, when used in its barest form of 'blackbox' analysis undoubtedly shows several shortcomings (e.g. Johnson, 1980).

The system theoretical investigation of the auditory nervous system, especially the more central parts of it, should be based on two main sources of inspiration: arguments from *ethology* should be used for the construction of a stimulus ensemble which poses the relevant questions to the system under investigation, whereas findings from *electrophysiology* should be incorporated into possible models of neural function. The analysis of the stimulus-normalization procedure of the STRF (Sect. 5) suggests that a straightforward formal approach to the neural representation of acoustic stimuli requires that the use of stimuli deviating more and more from being Gaussian, i.e. involving a stimulus machine *M* of increasing complexity, in principle necessitates the use of ever more sophisticated analysis procedures and a corresponding increase in the necessary computer power.

Acknowledgements. This investigation forms part of the research program of the workgroup Neurophysics at the Department of Medical Physics and Biophysics, University of Nijmegen (The Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (Z.W.O.). The authors wish to thank Jos Eggermont, Dik Hermes and Ton Vendrik for helpful discussions on the material presented in this paper.

REFERENCES

- Adrian, E.D., Cattell, M., Hoagland, H.: Sensory discharges in single cutaneous nerve fibers. J. Physiol. (Lond.) 72, 377-391 (1931)
- Aertsen, A.M.H.J., Johannesma, P.I.M.: Spectrotemporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli. Biol. Cybernetics 38, 223-234 (1980)
- Aertsen, A.M.H.J., Johannesma, P.I.M., Hermes, D.J.: Spectro-temporal receptive fields of auditory neurons in the grassfrog. II. Analysis of the stimulus-event relation for tonal stimuli. Biol. Cybernetics <u>38</u>, 235-248 (1980)
- Aertsen, A.M.H.J., Olders, J.H.J., Johannesma, P.I.M.: Spectro-temporal receptive fields of auditory neurons in the grassfrog. Biol. Cybernetics 39, 195-209 (1981)
- Alper, P., Poortvliet, D.C.J.: On the use of Volterra series representation and higher order impulse responses for nonlinear systems. Revue A Tijdschrift 6, 19-33 (1964) Barrett, J.F.: The use of functionals in the
- Barrett, J.F.: The use of functionals in the analysis of non-linear physical systems. J. Electronics and Control 15, 567-615 (1963)
- Bedrosian, E., Rice, S.O.: The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs. Proc. IEEE 59, 1688-1707 (1971)
- Boer, E. de, Kuyper, P.: Triggered correlation. IEEE Tr. on Biomed. Eng. BME-<u>15</u>, 169-179 (1968)
- Boer, E. de, Jongh, H.R. de: On cochlear encoding: potentialities and limitations of the reverse-correlation technique. J. Acoust. Soc. Am. <u>63</u>, 115-135 (1978)
- Boer, E. de: Polynomial correlation. Proc. IEEE <u>67</u>, 317-318 (1979)
- Brillinger, D.R.: The identification of polynomial systems by means of higher order spectra. J. Sound Vib. <u>12</u>, 301-313 (1970)

Brillinger, D.R., Bryant, Jr., H.L., Segundo, J.P.: Identification of synaptic interactions. Biol. Cybernetics 22, 213-228 (1976)

- Bullock, T.H. (ed.): Recognition of complex acoustic signals. Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Eggermont, J.J., Aertsen, A.M.H.J., Hermes, D.J., Johannesma, P.I.M.: Spectro-temporal characterization of auditory neurons: redundant or necessary? Hearing Research, in press (1981)
- French, A.S., Butz, E.G.: Measuring the Wiener kernels of a non-linear system using the fast Fourier transform algorithm. Int. J. Control 17, 529-539 (1973)
- Gisbergen, J.A.M. van, Grashuis, J.L., Johannesma, P.I.M., Vendrik, A.J.H.: Neurons in the cochlear nucleus investigated with tone and noise stimuli. Exp. Brain Res. 23, 387-406 (1975)
- Goblick, Th.J., Pfeiffer, R.R.: Time domain measurements of cochlear non-linearities using combination click stimuli. J. Acoust. Soc. Am. 46, 924-938 (1969)
- Grashuis, J.L.: The pre-event stimulus ensemble: an analysis of the stimulus-response relation for complex stimuli applied to auditory neurons. Ph.D. Thesis, Nijmegen, The Netherlands 1974
- Grüsser, O.-J., Grüsser-Cornehls, U.: Neuronal mechanisms of visual movement perception and some psychophysical and behavioral correlations. In: Handbook of sensory physiology. Vol VII/3A, Central processing of visual information, pp. 333-429. Jung, R. ed. Berlin, Heidelberg, New York: Springer 1973
- Hartline, H.K.: The receptive fields of the optic nerve fibers. Amer. J. Physiol. <u>130</u>, 690-699 (1940)
- Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M., Eggermont, J.J.: Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetized grassfrog, Rana temporaria L., investigated with noise stimuli. Hearing Research, in press (1981)
- Hung, G., Stark, L.: The kernel identification method (1910-1977) - Review of theory, calculation, application, and interpretation. Math. Biosci. <u>37</u>, 135-190 (1977)
- Johannesma, P.I.M.: Dynamical aspects of the transmission of stochastic neural signals.
 In: Proc. First European Biophysics Congress, pp. 329-333. Broda, E., Locker, A., Springer-Lederer, H. (eds.). Vienna: Verlag der Wiener Medizinischen Adakemie 1971
- Johannesma, P.I.M., Aertsen, A.M.H.J.: Neural image of sound in the grassfrog. In: Hearing Mechanisms and Speech, pp. 79-86. Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.). Exp. Brain Res. Suppl. II (1979)
- Johannesma, P.I.M.: Functional identification of auditory neurons based on stimulus-event correlation. In: Psychophysical, physiological and behavioural studies in hearing, pp. 77-84. Brink, G. van den, Bilsen, F.A.

(eds.). Delft University Press 1980

- Johannesma, P.I.M., Aertsen, A.M.H.J., Cranen, L., Erning, L.J.Th.O. van: The Phonochrome: a coherent spectro-temporal representation of sound. Submitted to Hearing Research (1981)
- Johnson, D.H.: Applicability of white-noise nonlinear system analysis to the peripheral auditory system. J. Acoust. Soc. Am. <u>68</u>, 876-884 (1980)
- Knudsen, E.I., Konishi, M.: Center-surround organization of auditory receptive fields in the owl. Science 202, 778-780 (1978)
- Krausz, H.J.: Identification of nonlinear systems using random impulse train inputs. Biol. Cybernetics 19, 217-230 (1975)
- Kroeker, J.P.: Wiener analysis of functionals of a Markov chain: application to neural transformations of random signals. Biol. Cybernetics <u>36</u>, 243-248 (1980)
- Kuffler, S.W., Nicholls, J.G.: From neuron to brain: a cellular approach to the function of the nervous system. Sunderland, MA: Sinauer Associates Inc. 1977
- Lee, Y.W.: Statistical theory of communication. New York: Wiley 1960
- Lee, Y.W.: Contributions of Norbert Wiener to linear theory and nonlinear theory in engineering. In: Selected papers of Norbert Wiener, with contributions from Lee, Levinson and Martin, pp. 17-33. Cambridge, MA: M.I.T. Press 1964
- Lee, Y.W., Schetzen, M.: Measurement of the Wiener kernels of a nonlinear system by cross-correlation. Int. J. Control 2, 237-254 (1965)
- Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., Pitts, W.H.: What the frog's eye tells the frog's brain. Proc. IRE 47, 1940-1951 (1959)
- Marmarelis, V.Z.: A family of quasi-white random signals and its optimal use in biological system identification. Part I: Theory. Biol. Cybernetics 27, 49-56 (1977)
- Marmarelis, P.Z., Marmarelis, V.Z.: Analysis of physiological systems. The white-noise approach. New York, London: Plenum Press 1978
- Møller, A.R.: Statistical evaluation of the dynamic properties of cochlear nucleus units using stimuli modulated with pseudo random noise. Brain Res. 57, 443-456 (1973)
- Palm, G., Poggio, T.: The Volterra representation and the Wiener expansion: validity and pitfalls. SIAM J. Appl. Math. <u>33</u>, 195-216 (1977a)
- Palm, G., Poggio, T.: Wiener-like system identification in physiology. J. Math. Biology <u>4</u>, 375-381 (1977b)
- Palm, G.: On representation and approximation of nonlinear systems. Biol. Cybernetics 31, 119-124 (1978)
- Papoulis, A.: The Fourier integral and its applications. New York: McGraw-Hill 1962
- Papoulis, A.: Probability, random variables and stochastic processes. Tokyo: McGraw-Hill Kogakusha 1965
- Pitts, W.H., McCulloch, W.S.: How we know uni-

versals: the perception of auditory and visual forms. Bull. Math. Biophysics <u>9</u>, 127-147 (1947)

- Scheich, H.: Central processing of complex sounds and feature analysis. In: Recognition of complex acoustic signals, pp. 161-182. Bullock, T.H. (ed.) Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Schetzen, M.: A theory of non-linear system identification. Int. J. Control 20, 577-592 (1974)
- Schetzen, M.: The Volterra and Wiener theories of nonlinear systems. New York: Wiley 1980
- Sejnowski, T.J.: On global properties of neuronal interaction. Biol. Cybernetics <u>22</u>, 85-95 (1976)
- Victor, J.D., Knight, B.W.: Nonlinear analysis with an arbitrary stimulus ensemble. Quart. Appl. Math. 37, 113-136 (1979)

- Webster, W.R., Aitkin, L.M.: Central auditory processing. In: Handbook of psychobiology, pp. 325-364. Gazzaniga, M.S., Blakemore, C. (eds.). New York: Academic Press 1975
- Wickelgren, B.C.: Superior colliculus: some receptive field properties of bimodally responsive cells. Science <u>173</u>, 69-71 (1971)
- Wiener, N.: Nonlinear problems in random theory. Cambridge, MA: M.I.T. Press; New York: Wiley 1958
- Worden, F.G., Galambos, R. (eds.): Auditory processing of biologically significant sounds. Neurosci. Res. Prog. Bull., Vol. 10. Brookline, Mass.: Neurosciences Research Program 1972
- Yasui, S.: Stochastic functional Fourier series, Volterra series and nonlinear systems analysis. IEEE Trans. Automat. Contr. AC-24, 230-242 (1979)

CHAPTER 8

A Comparison of the Spectro-Temporal Sensitivity of Auditory Neurons to Tonal and Natural Stimuli

Aertsen, A.M.H.J. and Johannesma, P.I.M.

A comparison of the Spectro-Temporal Sensitivity of Auditory Neurons to Tonal and Natural Stimuli.

A.M.H.J. Aertsen and P.I.M. Johannesma Dept. of Medical Physics and Biophysics University of Nijmegen Nijmegen, The Netherlands.

Abstract. The spectro-temporal sensitivity of auditory neurons has been investigated experimentally by averaging the spectrograms of stimuli preceding the occurrence of action potentials or neural events (the APES; Aertsen et al., 1980, 1981). The properties of the stimulus ensemble are contained in this measure of neural selectivity. The spectro-temporal receptive field (STRF) has been proposed as a theoretical concept which should give a stimulus-invariant representation of the second order characteristics of the neuron's system function (Aertsen and Johannesma, 1981).

The present paper investigates the relation between the experimental and the theoretical description of the neuron's spectro-temporal sensitivity for sound. The aim is to derive a formally based stimulus-normalization procedure for the results of the experimental averaging procedure.

Under particular assumptions, regarding both the neuron and the stimulus ensemble, an integral equation connecting the APES and the STRF is derived. This integral expression enables to calculate the APES from the STRF by taking into account the stimulus spectral composition and the characteristics of the spectrogram analysis. The inverse relation, i.e. starting from the experimental results and by application of a formal normalization procedure arriving at the theoretical STRF, is effectively hindered by the nature of the spectrogram analysis.

An approximative 'normalization' procedure, based on intuitive manipulation of the integral equation, has been applied to a number of single unit recordings from the grassfrog's auditory midbrain area to tonal and natural stimulus ensembles. The results indicate that spectrogram analysis, while being a useful realtime tool in investigating the spectro-temporal transfer properties of auditory neurons, shows fundamental shortcomings for a theoretical treatment of the questions of interest.

1. INTRODUCTION

The combined spectro-temporal characterization of the stimulus-response relation of auditory neurons can be given by the analysis of the spectro-temporal structure of stimuli preceding the occurrence of action potentials or neural events (Johannesma, 1972). This analysis can be performed by determination of the Average Pre-Event Spectrogram (APES; Aertsen et al., 1980, 1981). It has been applied to the responses of single neurons in the auditory midbrain of the anaesthetized grassfrog (Rana temporaria L.) to tonal (Aertsen et al., 1980) and natural (Aertsen et al., 1981) stimuli. In that case the spectro-temporal analysis had been implemented in hardware in the form of a real-time Dynamic Spectrum Analyser (DSA; Aertsen and Johannesma, 1980). A closely related approach has been used to study the responses of neurons in the lightly anaesthetized grassfrog to stationary wideband noise (Hermes et al., 1981), with the spectro-temporal representation of sound given in a more general and formalized manner (Johannesma and Aertsen, 1979; Johannesma et al., 1981). The formerly mentioned neurons which were investigated with tonal and natural stimuli did not respond to stationary noise, probably due to the higher level of anaesthesia used in those experiments.

The form of the APES is obviously highly influenced by the spectro-temporal characteristics of the stimulus ensemble. In order to arrive at a proper representation of the neuron's spectro-temporal sensitivity, irrespective of the way it has been measured, the result of averaging should be normalized for the a priori known stimulus properties. A somewhat intuitive approach has been described as the 'stimulus-filtering' procedure: the APES was divided by the power spectral density of the overall stimulus ensemble (Aertsen et al., 1981). The goal of a stimulus-normalization procedure is to investigate whether the nor-

AVERAGE PRE-EVENT SPECTROGRAM FOR TONAL STIMULI



Fig. 1 Average Pre-Event Spectrogram (APES) of eight torus semicircularis units (a-h), determined with a tonal stimulus ensemble. The a priori expected result for this stimulus ensemble, the Average Spectrogram (AS), is shown under i. The intensity scale in all Figs. refers to arbitrary, identical units. High values are indicated by darker, low values by lighter grey-values. The numbers N_1 and N_2 in the Figs. indicate the number of events presented to the DSA (N_1), resp. the number of events actually considered in averaging (N_2). Further explanation in the text.

malized APES for such various stimulus ensembles like tones and natural sounds can be reconciled into one model of the neuron's spectro-temporal sensitivity.

In the present paper the relation between the APES for tonal and natural stimuli will be investigated, using a more formal approach, based on the concept of the Spectro-Temporal Receptive Field (STRF; Aertsen and Johannesma, 1981). The STRF has been formally identified with the second order Volterra kernel of the neuron's system function. It has been shown that under certain conditions on both the neuron and the stimulus ensemble a stimulusnormalization procedure can be given which, when applied to the second order cross correlation function of the neuron's in- and output leads to the stimulus-invariant STRF. Although in the present case these conditions are far from being fulfilled, the use of such

AVERAGE PRE-EVENT SPECTROGRAM FOR NATURAL STIMULI



Fig. 2 Average Pre-Event Spectrogram (APES) of eight torus semicircularis units (a-h), determined with a natural stimulus ensemble. The a priori expected result for this stimulus ensemble, the Average Spectrogram (AS), is shown under i. These results refer to the same neurons as have been shown in Fig. 1 for a tonal stimulus ensemble. Further details as in Fig. 1.

an approach may nevertheless provide information which can be relevant in tackling the long-standing issue of reconciling the neural responses to 'simple' and 'complex' stimuli (e.g. Worden and Galambos, 1972; Bullock, 1977). In this context it will be necessary to consider the formal relation between the dynamic spectrum analysis (DSA), leading to the APES and the evaluation of second order inputoutput cross correlation.

2. METHODS

The experimental data used in the present investigation were obtained from extra-cellular single unit recordings in the torus semicircularis (the auditory midbrain area) of the grassfrog, Rana temporaria L., described in earlier papers (Aertsen et al., 1980, 1981).

The stimulus-event (s-e) relation was investigated for two different stimulus ensembles (Aertsen and Johannesma, 1980). The first one consisted of sequences of short (16 ms) tonal elements, the γ -tones, the form of which was
inspired by the short sound segments in the male grassfrog vocalizations. The carrier frequency and the amplitude factor of the γ -tones were varied in a pseudo-random way throughout the sequence. The second stimulus ensemble consisted of a wide selection of sounds, occurring in the natural environment of the grassfrog: the Acoustic Biotope.

The properties of the s-e relation have been investigated by analysing the ensemble of stimuli preceding the occurrence of action potentials, the Pre-Event Stimulus Ensemble (PESE; Johannesma, 1972). The spectro-temporal characteristics of the PESE were evaluated by determining the ensemble average of the dynamic power spectra of individual pre-event stimuli, measured by a real-time Dynamic Spectrum Analyser (DSA; Aertsen and Johannesma, 1980). This average measure obviously is heavily influenced by the overall spectro-temporal properties of the stimulus ensemble. The average pre-event dynamic power spectra of eight different torus semicircularis units are shown in Fig. 1 for the tonal stimulus ensemble and in Fig. 2 for the natural stimulus ensemble. A few of these results have been published earlier (Aertsen et al., 1980, 1981). Results are given in a grey-scale representation where higher function values are indicated by darker grey-values. The results for the neural event sequences are shown in the Fig. elements a-h, whereas element i shows the outcome of averaging for a random sequence of events. The latter result provides an indication of the overall spectro-temporal structure of the stimulus ensemble. In each Fig. the results were scaled using the same intensity scale to allow a comparison. The numbers of events, involved in the analysis, are indicated in the Figs.: the first number (N,) indicates the number of events recorded during the presentation of the stimulus ensemble, the second number (N_2) denotes the number of pre-event stimuli actually considered in averaging. Due to a 'dead-time' in the averaging processor of the DSA of 29.5 ms (Aertsen and Johannesma, 1980) it generally appears that $N_2 < N_1$.

The frequency range of the dynamic spectrum analysis ran from 125 Hz to 3200 Hz on a logscale, i.e. $15\frac{1}{3}$ octave filters. The tonal stimulus sequences covered either one of two frequency ranges: 125-2000 Hz or 250-4000 Hz. For 6 out of 8 units (elements a-d, g and h in Fig. 1) the average dynamic spectrum for tonal stimuli was obtained by combining the two separate averages corresponding to the respective tonal sequences in the following way: in the overlapping frequency band a weighed average was determined, while in the non-overlapping frequency bands only the average corresponding to one sequence of tonal stimuli was considered. The same procedure was applied to the results for the random event sequence (Fig. 1i). For the remaining 2 units (elements e and f in Fig. 1) the average dynamic spectrum for only the upper one of the two tonal sequences was available, as can be seen from the corresponding results.

3. FROM THE APES TO A NEURAL CHARACTERISTIC

The Spectro-Temporal Receptive Field (STRF) of an auditory neuron has been formally defined as the frequency-time representation of the second order Volterra kernel of the neuron's nonlinear system function, transforming the acoustic stimulus x(t) into the firing probability y(t)(Aertsen and Johannesma, 1981). As such it constitutes a stimulus-invariant functional characteristic of the neuron, it may be viewed as a kind of spectro-temporal transfer function. It has been shown that for (1) a Gaussian stimulus ensemble and (2) a neuron's system function with no even order contributions of order higher than two, the second order cross correlation of stimulus and response, normalized with respect to the spectral contents of the stimulus ensemble, leads to the STRF. The normalization-procedure in this case is given by

$$\hat{\hat{v}}_{2}(\omega,\nu) = \frac{\hat{R}_{xxy}, (\omega,\nu)}{2\hat{R}_{xx}(\omega) \hat{R}_{xx}(\nu)}$$
(1a)

where \hat{v}_2 denotes the double Fourier transform of the system's second order Volterra kernel $v_2(\sigma, \tau)$, \hat{R}_{xx} the power spectral density of the stimulus ensemble and \hat{R}_{xxy} , the second order cross spectrum of stimulus x and the deviation of the response y from its average value during the stimulation: $y'(t) = y(t) - \overline{y}$. An equivalent formulation of (1) is given by

$$\hat{\hat{v}}_{2}(\omega,\nu) = \frac{\hat{\hat{R}}_{xxy}(\omega,\nu)}{2\hat{R}_{xx}(\omega)\hat{R}_{xx}(\nu)} - \frac{\hat{\hat{R}}_{xxy}(\omega,\nu)}{2\hat{R}_{xx}(\omega)\hat{R}_{xx}(\nu)}$$
(1b)

where the second term on the right-hand side differs only from zero for $\omega+\nu=0$: for a Gaussian stimulus ensemble it holds that

$$\hat{\hat{R}}_{xxy}(\omega, v) = \overline{y}\hat{R}_{xx}(\omega) \ \delta(\omega+v).$$

The APES has been introduced to provide an experimental spectro-temporal characterization of the neuron's stimulus-response relation. For every stimulus segment, preceding a neural event, the spectro-temporal structure is measured by a dynamic spectrum analyser and the resulting spectrograms are averaged for all events occurring during the presentation of the stimulus ensemble (Aertsen et al., 1980). The derivation of a mathematical relation between the STRF and the APES necessitates a more formal description of the latter concept. The goal is to arrive at a stimulus-normalization for the APES, inspired by the relation given in (1).

3.1 Formal description of the APES

The Spectrogram

The operation of determining the dynamic power spectrum or spectrogram of a signal x(t) by a DSA-type analyser can be formalised as follows. The output of the DSA consists of a set of nonnegative functions of time $\{P_{\underline{l}}(t) ; \underline{l} = 1, L\}$, shortly $P(\underline{l}, t)$ representing the time-dependent intensity of the sound in neighbouring 1/3octave bands with center frequencies $\omega_{\underline{l}}$. For each separate time function $P(\underline{l}, t)$, which forms a particular cross section of the dynamic power spectrum $P(\omega, t)$ it holds that (cf. Appendix 2 in Aertsen and Johannesma, 1980)

$$P(l,t) = (f_l * x)^2(t) + (f_l * x)^2(t)$$
(2)

with f_l(t) = impulse response of l-th bandpass filter in the DSA * = convolution operator

~ = Hilbert transform (e.g. Papoulis, 1962)

By introduction of the 'analytic signal' $\phi_1(t) = f_1(t) + i \tilde{f}_1(t)$, with $i = \sqrt{-1}$ (Gabor, 1946; Ville, 1948) and using the relation (e.g. Deutsch, 1969)

$$f_{l} \ast x = f_{l} \ast x = f_{l} \ast \widetilde{x}$$
(3)

the relation (2) can be rewritten as

$$P(l,t) = \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} ds' \phi_l(s) \phi_l^*(s') x(t-s) x(t-s')$$
(4a)

which expresses the spectrogram as a quadratic functional of the signal x(t). Taking the Fourier transform of (4a) we arrive at its spectral equivalent:

$$\hat{P}(l,\mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \, \hat{\phi}_{l}(\lambda) \hat{\phi}_{l}^{*}(\lambda-\mu) \hat{x}(\lambda) \hat{x}^{*}(\lambda-\mu)$$
(4b)

where indicates the Fourier transform and ϕ_{1}^{*} denotes the complex conjugate of ϕ_{1} . The spectrum $\phi_{1}(\lambda)$, associated with the analytic filter impulse response $\phi_{1}(t)$ equals (Gabor, 1946)

The frequency integral in (4b) should be taken from $-\infty$ to $+\infty$, the properties of $\phi_l(\lambda)$ imply that for $\mu \ge 0$ the result is equivalent to an integration from μ to $+\infty$. Furthermore, since P(l,t) is real, it holds that $\hat{P}(l,-\mu) = \hat{P}^*(l,\mu)$.

Averaging of Spectrograms and Cross Correlation

The operation of averaging spectrograms is a particular form of statistical analysis of the stimulus-response relation of a neuron. The stimulus ensemble SE is presented to the animal



Fig. 3 Model of a primary auditory neuron (from Johannesma, 1980).

and the electrical activity of a single neuron is measured extracellularly. The result is a sequence of action potentials, the neural events. Repetition of the experiment leads to another sequence of events, generally not identical to the first one. For our purposes the neural activity z(t) can be modelled as a single realization of a stochastic point process:

$$z(t) = \sum_{n=1}^{N} \delta(t-t_n)$$
(6)

with $\delta(t)$ = Dirac delta function

- ${t_n}$ = moments of occurrence of action potentials
 - N = number of action potentials during presentation of the stimulus ensemble (duration T).

A simple model of the single neuron system is depicted in Fig. 3: a transducer followed by an event generator (De Jongh, 1978, Johannesma, 1971, 1980). The acoustic stimulus x(t) is transformed by the transducer into the generator potential y(t). The transducer mechanism can be described by a deterministic, time-invariant, continuous, dynamic nonlinear system S, i.e. a Volterra system. All stochastic aspects of the neural signal transformation are comprised in the event generator: the neural response function or firing probability g(t) is a continuous, non-negative, algebraic function of the generator potential. Although the model, strictly speaking, only applies to primary auditory neurons, its formalism will be adopted here.

The event-density or PSTH n(t), defined as the ensemble average $\langle z(t) \rangle$ of the neuron's firing pattern over a number of presentations of the same stimulus repertoire, is used as an experimental estimate of the firing probability density g(t). It has been shown that the cross correlation function $R_{\Psi n}$ of any stimulus functional $\Psi(x)$ and the event density n(t) equals the expected value of the stimulusevent correlation $\langle R_{\Psi z} \rangle$ (Johannesma, 1972, 1980):

$$R_{\psi_z} > = R_{\psi_n} = experimental estimate of R_{\psi_q}$$
 (7)

When we adopt the simple model of a *linear* event generator g(y) (which implies that the transducer should be such that y(t)>0 for all t, or, the possible nonlinearity in the firing probability has been absorbed in the transducer

$$\langle R_{\psi z} \rangle = R_{\psi y}$$
 (8)

The 'stochastic' stimulus-event correlation $R_{\Psi z}$ thus has been related to the 'deterministic' Volterra system correlation $R_{\Psi y}$. A related approach has been used for an 'y exponential event generator (Johannesma, 1980).

In case a long and wide stimulus ensemble is used, i.e. long and wide as compared to the range of $\langle R_{\psi_Z} \rangle$, this effectively implies an averaging over the stochastics of the event generator. In that case the stimulus-event correlation over a single presentation of the SE forms a reasonable estimate of its expected value over a number of presentations:

$$R_{\psi_{\mathcal{Z}}} = \langle R_{\psi_{\mathcal{Z}}} \rangle \tag{9}$$

Combining (7), (8) and (9) we have

$$R_{\Psi z} = R_{\Psi y} \qquad (10)$$

or: under the assumption of a linear event generator the experimental result of stimulusevent correlation over a long and wide stimulus ensemble can be used as an estimate of the theoretical stimulus-response correlation of the deterministic nonlinearity S.

When applied to the case of the dynamic power spectrum $P(\omega, t)$ as stimulus functional Ψ this leads to the stimulus-event correlation R_{PZ} as the experimental estimate of the cross correlation of stimulus spectrogram and generator potential. In connection to (9) it may be noted that for a long ensemble of various natural stimuli the result of stimulusevent correlation R_{PZ} appeared to be highly reproducible upon repeated presentation of the stimulus ensemble (Aertsen et al., 1981). The stimulus ensemble SE, considered as a single time-function x(t) with large duration T, gives for the cross correlation R_{PZ} ,

$$R_{Py}(\omega,\tau) = \frac{1}{T} \int_{0}^{T} dt P(\omega,t-\tau)y(t)$$
(11)

In case the SE is thought to be a long, single realization of a Gaussian process and furthermore the system S is time-invariant, the timeintegral in (11) can be considered as an estimate of the expected value $E(P(\omega, t-\tau)y(t))$ for the stochastic input process x and the corresponding output process y.

The experimentally obtained stimulus-event correlation $R_{PZ}(\omega, \tau)$, in analogy to (11), is given by

$$R_{Pz}(\omega,\tau) = \frac{1}{T} \int_{0}^{T} dt P(\omega,t-\tau)z(t)$$
(12)

with z(t) as described in (6). The event character of z(t) causes this integral to result in an averaging operation:

$$R_{Pz}(\omega,\tau) = \frac{1}{T} \sum_{n=1}^{N} P(\omega,t_n-\tau)$$
(13)

with t_n the moment of occurrence of the n-th action potential. This result leads to a more general interpretation of stimulus-event correlation.

The Average Pre-Event Spectrogram

The Pre-Event Stimulus Ensemble (PESE) is defined as the collection of stimuli preceding the neural events (Johannesma, 1972; Aertsen et al., 1980, 1981). As a consequence the PESE forms a subset of the SE, its structure reflecting the neuron's selective properties. We have investigated the spectro-temporal structure of the PESE by measuring the ensemble average of dynamic power spectra, associated with the individual pre-event stimuli: the Average Pre-Event Spectrogram (APES) $p_e(u,\tau)$ given by

$$P_{e}(\omega,\tau) = \langle P(\omega,\tau) \rangle_{PESE} = \frac{1}{N} \sum_{n=1}^{N} P_{n}(\omega,\tau)$$
(14)

where $\tau > 0$ indicates time *before* the neural event and $P_n(\omega, \tau)$ denotes the spectrogram of the n-th pre-event stimulus $x_n(\tau) = x(t_n - \tau)$. The comparison of (10), (13) and (14) immediately shows the following proportionality

$$\frac{N}{T} p_e(\omega, \tau) = R_{Py}(\omega, \tau)$$
(15)

The relations (6)-(15) thus lead to the conclusion that the product of the average firing rate N/T and the average pre-event spectrogram is equal to the cross correlation across the deterministic system S of the stimulus spectrogram and the generator potential.

For the practical implementation of the APES we return to the DSA-equation given in (4a) and (4b). The combination of (4) and (14) leads to the APES $p_{\rho}(l,\tau)$, given by

$$p_e(l,\tau) = \int ds \int ds' \phi_l(s) \phi_l^*(s').$$

$$\cdot \langle x(t_n - \tau - s) x(t_n - \tau - s') \rangle_{PESE} \qquad (16a)$$

or, in the frequency domain,

$$\hat{p}_{e}(l,\mu) = \frac{1}{2\pi} \int d\lambda \, \hat{\phi}_{l}^{*}(\lambda) \, \hat{\phi}_{l}(\lambda-\mu).$$

$$\cdot \langle x_{n}(\lambda)x_{n}(\lambda-\mu) \rangle_{PESE} \qquad (16b)$$

Following (15) the APES forms an estimate of R_{Py} . As a consequence we have the theoretical relation

$$\frac{N}{T} p_e(l,\tau) = \frac{1}{T} \int_0^l dt \frac{1}{2\pi} \int ds \int ds' \phi_l(s) \phi_l^*(s')$$

$$\cdot x(t-\tau-s)x(t-\tau-s')y(t)$$
(17)

Changing order of integration this leads to

$$\frac{N}{T} p_e(l,\tau) = \frac{1}{2\pi} \int ds \int ds' \phi_l(s) \phi_l^*(s')$$
$$\cdot \frac{1}{T} \int_0^T dt \ x(t-\tau-s)x(t-\tau-s')y(t)$$
(18)

or, equivalently,

$$\frac{N}{T} P_{e}(l,\tau) = \frac{1}{2\pi} \int ds \int ds' \phi_{l}(s)\phi_{l}^{*}(s').$$

$$\cdot R_{xxy}(\tau + s, \tau + s')$$
(19a)

In the frequency domain this reads

$$\frac{N}{T}\hat{P}_{e}(l,\mu) = \frac{1}{2\pi} \int d\lambda \hat{\phi}_{l}^{*}(\lambda)\hat{\phi}_{l}(\lambda-\mu) \hat{R}_{xxy}(\lambda,\mu-\lambda)$$
(19b)

This implies that the APES $p_{c}(l,\tau)$ is a linear functional of the second order input-output cross correlation function: for every *l*-th DSAchannel (19a) describes the 2-dimensional convolution of the time-inverted cross correlation function R_{xxy} and the matrix $\phi_{l}\phi_{l}^{*}$ associated with the filter inpulse response ϕ_{l} . The corresponding relation (19b) in the frequency domain thus may be viewed as a 2-dimensional DSA-filtering of the second order cross spectrum.

It should be noted that the transformation from R_{xxy} to the APES cannot be inverted, due to the projection-like integration in (19). This is quite a general characteristic of the sonograph-type of spectrum analysers, described by (4), of which the DSA forms an example: phase relations of different spectro-temporal components in the signal are not preserved.

3.2 Relation between APES and STRF

For a Gaussian white stimulus ensemble the second order input-output correlation R_{xxy} is tightly connected to the system's second order Wiener kernel w_2 (Lee and Schetzen, 1965). In case the system does not contain even order nonlinearities of order higher than two the Wiener kernel w_2 is identical to the system's quadratic Volterra kernel v_2 ; in the general case w_2 can be considered as the second order 'Wiener-contribution' to the system kernel v_2 which forms the base of the formally defined STRF (Aertsen and Johannesma, 1981). The relation (19) implies that for a Gaussian white stimulus ensemble the APES $p_e(l,\tau)$ minus Average Spectrogram (AS) $p(l,\tau)$ of the overall stimulus ensemble, by its connection to the difference $R_{xxy} - R_{xxy}$, represents the 'DSA-mediated' second order Wiener contribution to the neuron's STRF. For the special class of systems with $v_2 = w_2$ the measurement of the Difference Spectrogram (DS) $q_e(l,\tau)$, defined by

$$q_{\rho}(l,\tau) = p_{\rho}(l,\tau) - p(l,\tau)$$
(20)

directly represents the DSA-filtered STRF. The magnitude of the higher even order nonlinearities in the neuron's system function determines the degree in which the DS, for a Gaussian white stimulus ensemble, represents a faithful image of the STRF. The distortion introduced by the use of stimulus ensembles which deviate from Gaussian white noise will now be considered. The stimulus-normalization of the APES will be considered on the base on the normalization procedure of the STRF as given in (1), i.e. for linearly filtered Gaussian noise. A more general approach for nonlinearly transformed Gaussian noise, involving the use of higher order correlation functions, as indicated in Aertsen and Johannesma (1981) is beyond the scope of the present discussion.

The correlation function R_{xxy} can be expressed, by using (1), in terms of the system kernel v_2 and the stimulus autocorrelation R_{xx} . This results in a formal relation between the APES and the STRF. When stated in the frequency domain this leads to

$$\frac{N}{T} \left(\hat{p}_{e}^{(l,\mu)} - \hat{p}^{(l,\mu)} \right) = \frac{1}{\pi} \int d\lambda \, \hat{\phi}_{l}^{*}(\lambda) \phi_{l}^{(\lambda-\mu)} \, \hat{R}_{xx}^{(\lambda)} (\lambda) \hat{R}_{xx}^{(\mu-\lambda)} \, \hat{\hat{v}}_{2}^{(\lambda,\mu-\lambda)}$$

$$(21)$$

The kernel $v_2(\lambda,\mu-\lambda)$ appearing in this relation equals the double Fourier transform of $v_2(\tau+\sigma,\tau)$ with respect to τ and σ . It may be noted here that for Gaussian noise the Fourier transformed AS $\hat{p}(l,\mu)$ in fact represents a filter-dependent constant $p_l \,\delta(\mu)$, measuring the contribution to the power spectrum of the overall stimulus ensemble which falls within the passband region of the *l*-th DSA-filter. It can experimentally be estimated by measuring the average spectrogram of stimuli preceding a random sequence of events (cf. Figs. 1i and 2i for the tonal resp. the natural SE).

The left-hand side of (21) represents an experimental measure: the product of 1) the average firing rate N/T and 2) the deviation of the average pre-event spectrogram p_e from the a priori expected stimulus spectrogram p. The right-hand side shows a frequency integral in which the following factors can be discerned: 1) the quadratic system kernel $\hat{v}_2(\lambda,\mu-\lambda)$, which is filtered 2-dimensionally by 2) the bi-spectral properties of the stimulus ensemble $\hat{R}_{xx}(\lambda) \hat{R}_{xx}(\mu-\lambda)$, the result of which in turn is filtered again by 3) the DSA-filter operator $\phi_{\pi}^{\pm}(\lambda)\phi_{\pi}(\lambda-\mu)$.

If in some way v_2 is known, for instance from correlation analysis of Gaussian white noise experiments, and if all the other v_{2n} are zero, the relation (21) offers the possiblity to calculate the form of the APES for a nonwhite Gaussian stimulus ensemble. This result then may be compared to the experimental result of stimulus-event correlation in order to test the stimulus-invariance of the neuron's STRF. Another point of interest is the stimulus amplitude dependence of the system characteristics. The expression (21) predicts that scaling of the SE-amplitude by a factor A results in a scaling of the product $N/T q_{\rho}(l,\tau)$ by a factor of A4. When the experimentally determined product of average firing rate and difference spectrogram does not show this scaling property the system may either be not second order in its even components and/or the STRF is inherently amplitude dependent, for instance by an amplitude-dependent gain factor ('longterm adaptation').

As already noted the integral in (21) cannot simply be inverted to yield an explicit expression of the system kernel v_2 in terms of the spectrogram, the stimulus properties and the DSA-filter characteristics. Further simplification has to be obtained in order to acquire a practically useful stimulus-normalization procedure of the APES. This will involve some quite rigorous approximations which will be discussed in the following section.

4. APPROXIMATIVE 'STIMULUS-NORMALIZATION' OF THE APES

Under the given assumptions, the formal relations given in Sect. 3, which result in the integral equation (21), are exact. The 'approximations' which have been made were 'probabilistic' approximations: experimental estimators have been identified with theoretically expected values. In this section a 'normalization' procedure will be developed, based on (21). The goal is to arrive at a procedure which can be implemented for practical use, e.g. in the form of a computer program. Several strong approximations will be made, which implies that it will not be possible to attain mathematical rigour at the various steps. We nevertheless consider it to be useful to reach an applicable 'normalization' procedure, in order to explore the boundaries of useful application of stimulus-event characterization procedures as have been described here.

The quantity of basic interest in (21) is the system kernel v_2 . The first step thus is to reach an approximation where v_2 appears before the integral sign instead of being part of the integrand. If now 1) v_2 , in the bi-frequency domain, varies smoothly compared to the remainder of the integrand and if 2) the absolute value of this remainder, for each l, obtains a single sharp maximum, then an approximation is obtained by evaluating v_2 at the bi-spectral location of this maximum. Using the 1/3-octave shape of the DSA-filters and furthermore assuming that also the stimulus power spectrum \hat{R}_{xx} is behaving smoothly as compared to the filter characteristic $|\hat{\phi}_l|$ it follows that for the *l*-th DSA-channel the maximum is obtained for $\lambda_{l} = \omega_{l} + \mu$. As a consequence, under these subsequent smoothness assumptions relation (21) may be approximated Ъy

$$\frac{N}{T} \hat{q}_{e}(l, \mu) = \hat{v}_{2}(\omega_{l} + \mu, -\omega_{l}).$$

$$\cdot \frac{1}{\pi} \int d\lambda \ \hat{\phi}_{l}^{*}(\lambda) \hat{\phi}_{l}(\lambda - \mu) \ \hat{R}_{xx}(\lambda) \hat{R}_{xx}(\mu - \lambda)$$
(22)

or, in the form of a normalization procedure,

$$\hat{v}_{2}(\omega_{l}+\mu,-\omega_{l}) =$$

$$\frac{\frac{N}{T} (p_e(l,\mu) - p(l,\mu))}{\frac{1}{\pi} \int d\lambda \hat{\phi}_l^*(\lambda) \hat{\phi}_l(\lambda - \mu) R_{xx}(\lambda) R_{xx}(\mu - \lambda)}$$
(23)

A mathematical problem which has been ignored here is the fact that both \hat{v}_2 and $\hat{\phi}_1$ are complex and oscillating functions of frequency. The main argument for still making the approximation (22) is the assumption that $|\hat{\phi}_l|$ is very sharp compared to the other functions, so the most prominent contribution to the integral comes from a narrow frequency band around the maximum of $|\hat{\phi}_l^*(\lambda)| |\hat{\phi}_l(\lambda-\mu)|$, i.e. $\lambda_l = \omega_l + \mu$. This point however, should be studied more carefully.

From (23) it follows that the Fourier transformed difference spectrogram DS should be divided by an integral expression which in principle can be evaluated, since all quantities in the integrand are amenable for experimental determination. For every l and every μ we have a different integral with gradually shifting filters and power spectra. Since, however, the numerical evaluation of all these integrals requires a large amount of computing, it is worthwile to find out whether the integral in (23) can be expressed in terms of quantities which can simply be measured. First of all, the comparison of (4b) and (23) shows that the denominator in (23) can be interpreted as the bi-frequency representation of the DSA-spectrogram of the stimulus autocorrelation function $R_{mr}(\tau)$, considered as a signal. In other words, following (23) the Difference Spectrogram has to be inversely filtered with a 2-dimensional filter composed of the DSA-filters and the 'stimulus filter'. Guided by intuition we have made a further approximation, leading from (23) to

$$\hat{v}_2(\omega_1 + \mu_3 - \omega_1) =$$

$$\frac{\frac{N}{T}\left(\hat{p}_{e}(l,\mu)-\hat{p}(l,\mu)\right)}{\frac{1}{\pi}\hat{R}_{xx}(\omega_{l}+\mu)\int d\lambda \hat{\phi}_{l}^{*}(\lambda)\hat{\phi}_{l}(\lambda)\hat{R}_{xx}(\lambda)}$$
(24)

which implies that the spectrogram in the denominator of (23) has been factorized into a purely spectral component and a component representing temporal features of $R_{xx}(\tau)$. The spectral component

$$I(\omega_{l}) = \int d\lambda \,\hat{\phi}_{l}^{*}(\lambda) \hat{\phi}_{l}(\lambda) R_{xx}(\lambda)$$
(25)

is simply the overall stimulus power spectrum as measured by the DSA at the filter center frequencies ω_{l} and thus is readily obtained experimentally (compare the Average Spectrogram $p(l,\tau)$).



Fig. 4 Equalized Difference Spectrogram (EDS) for a tonal stimulus ensemble of eight torus semicircularis units (a-h), vs. the a priori expected results for this ensemble (i). These Figs. represent the result of a partial 'stimulus-normalization' procedure of the tonal APES's in Fig. 1. Further explanation in the text.

The second component $R_{xx}(\omega_{\ell}+\mu)$, contrarily to $I(\omega_{\ell})$, is a function of μ and should be interpreted as the Fourier transform

$$\hat{R}_{xx}(\omega_{l}+\mu) = \mathcal{F}_{\tau+\mu} \quad R_{xx}(\tau) \ e^{-i\omega_{l}\tau}$$
(26)

or: it is the frequency representation of the ω_{L} -component of $R_{xxx}(\tau)$. The combination of the two factors (25) and (26) can loosely be written as

$$I(\omega_{l}) \hat{R}_{xx}(\omega_{l}+\mu) = \mathcal{F}_{\tau+\mu} \quad I(\omega_{l}) e^{-i\omega_{l}\tau} R_{xx}(\tau)$$
(27)

which, when we consider $I(\omega_l)$ as the DSA-version of $\hat{R}_{xx}(\omega)$ in ω_l , shows some reminiscences of the complex spectro-temporal intensity density function of the autocorrelation $R_{xx}(\tau)$ (Rihaczek, 1968; Johannesma and Aertsen, 1979; Johannesma et al., 1981).

By a heuristic type of reasoning the integral relation between APES and system kernel, expressed in (21), has been transformed into the approximative relation (24), which can be considered to define an approximative 'normalization'-procedure, based on experimentally obtainable quantities like the AS and its 1dimensional equivalent, $I(\omega_L)$. The 'stimulusnormalization' of the APES, described by (24),



Fig. 5 Equalized Difference Spectrogram (EDS) for a natural stimulus ensemble of eight torus semicircularis units (a-h), vs. the a priori expected result for this ensemble (i). These Figs. represent the result of a partial 'stimulus-normalization' procedure of the natural APES's in Fig. 2. Further explanation in the text.

leading to the DSA-mediated STRF, can be considered to be composed of three different stages:

- The subtraction from the APES of the a priori expected average spectrogram AS of the overall stimulus ensemble, leading to the Difference Spectrogram DS = APES-AS.
- 2. The division of each separate channel in the DS by the filter-dependent constant $I(\omega_l)$. This correction, in fact an 'equalization' for the spectral composition of the overall stimulus ensemble, was introduced earlier on purely intuitive grounds: the 'stimulus-filtering' procedure (Aertsen et al., 1981). The result at this stage may be called the Equalized Difference Spectrogram (EDS).
- 3. The third stage, the division (in the frequency domain) of the EDS by the shifted spectrum $\hat{R}_{xxx}(\omega_l + \mu)$ sort of demodulates the various channels in the EDS with respect to the spectral composition of the SE at the various frequencies within the passband region of the corresponding DSA-filter. The result after the third stage of the procedure will be called 'Normalized' Pre-Event Spectrogram.

The acutes (') surrounding the adjective 'normalized' should emphasize that the proposed procedure does not follow from straightforward theoretical considerations. Only the first





Fig. 6 'Normalized' Pre-Event Spectrogram for a tonal stimulus ensemble of eight torus semicircularis units (a-h), vs. the a priori expected result for this ensemble (i). These Figs. represent the result of an approximative 'stimulusnormalization' procedure of the tonal APES's in Fig. 1. Compare also the corresponding EDS's in Fig. 4. Further explanation in the text.

stage, subtraction of the AS, follows from formal theory. The other two stages cannot lay a claim to mathematical rigour, rather they represent an intuitively based attempt to escape from the inherent limitations, imposed by an analysis based on DSA-type signal representations.

5. RESULTS

In order to illustrate the effect of the 'stimulus-normalization' procedure (24), developed in Sect. 4, we present the results of its application to the APES's of 8 neurons from the torus semicircularis of the grassfrog, obtained for tonal and natural stimuli as shown in Figs. 1 resp. 2. The results after the second stage of 'correction', the Equalized Difference Spectrograms, are shown in Fig. 4 for tonal stimuli and Fig. 5 for natural stimuli. The final results, the 'Normalized' Pre-Event Spectrograms are shown in Fig. 6 (tonal) and 7 (natural).

Analogously to the APES-results in Figs. 1 and 2 the elements a-h refer to various single neuron event sequences, whereas the element i shows the result for a random sequence of events. The latter results may give an indication regarding the overall effect of the proce-



Fig. 7 'Normalized' Pre-Event Spectrogram for a natural stimulus ensemble of eight torus semicircularis units (a-h), vs. the a priori expected result for this ensemble (i). These Figs. represent the result of an approximative 'stimulusnormalization' procedure of the natural APES's in Fig. 2. Compare also the corresponding EDS's in Fig. 5. Further explanation in the text.

dure and the reliability of its results. Within each Fig. the results in the various elements were scaled using the same intensity scale; between Figs. the intensity scale generally differed to obtain a maximal resolution in the grey-scale representation. For a general discussion on the various phenomena encountered in these (ω, τ) -representations and their neurophysiological interpretation the reader is referred to Aertsen et al. (1980, 1981).

The comparison of APES (Fig. 1) and EDS (Fig. 4) for tonal stimuli shows that the effect of subtraction of the AS and subsequent equalizing in that case is not very drastic, as should be expected for a stimulus ensemble with a rather smooth spectral composition (cf. Fig. li). The main effect is an enhancement of low- ω_{l} regions as becomes evident from the flattening of the result for random events (compare Figs. 1i and 4i), which reflects the properties of the stimulus power spectrum as determined with the DSA. As a consequence the 'activation areas', 1.e. areas in the (ω, τ) domain with conspiquously higher values in the spectrogram than expected a priori, indicated by darker regions in the Figs., for low-frequency units are enhanced somewhat, compared to the high-frequency units (e.g. units 35-6 (d) and 26-4(f)).

The effect of equalizing for natural stimuli (compare Figs. 2 and 5) is more pronounced, as should be expected from a stimulus ensemble with an overall power spectrum which quite strongly varies with frequency (cf. Fig. 2i and also Fig. 6 in Aertsen et al. (1981)). The strong contribution of certain frequency components in the natural SE, notably the 500 Hz component due to the grassfrog vocalizations, which dominates a number of APES's (e.g. units 35-6 (a) and 28-3 (h)), is weakened. The same effect can be observed in the result for random events (Figs. 2i and 5i). This results in natural EDS's which present, in their spectral aspects, a more balanced and smooth impression of the neuron's selectivity.

On the whole these results are very similar to those already obtained by the 'stimulusfiltering' procedure (Aertsen et al., 1981): the spectral preferences as reflected in the EDS for different stimulus ensembles (Figs. 4 and 5) show a better match than the original APES (Figs. 1 and 2). As to be expected from a frequency equalizing procedure, it has no effect whatsoever on the temporal patterns in the APES, which are due to the specific temporal structure of both the tonal ensemble and the natural ensemble, combined with the spectrotemporal sensitivity of the various neurons.

The effect of the third stage of the 'normalization'-procedure can be seen from a comparison of Fig. 4 and 6 for tonal stimuli, resp. Figs. 5 and 7 for natural stimuli. The general tendency here is that the high- $\omega_{\mathcal{I}}$ regions in the EDS are enhanced compared to the low- ω_{1} regions. This is caused by the gradually decreasing values of $\tilde{R}_{xx}(\omega_l + \mu)$ on a linear frequency scale with increasing ω_7 . This decrease holds for both the tonal and the natural ensemble. This decrease of the denominator $R_{rr}(\omega_1 + \mu)$, which induces increasing numerical problems for increasing ω_{l} -channels in the EDS would not have occurred for stimulus ensembles with an overall 'whiter' power spectral density. The general tendency in the results (Figs. 6 and 7), however, is consistent with a log frequency representation (induced by the DSA) of a spectral density which should rather be viewed on a linear frequency scale (compare a log ω representation of white noise). The 'overrepresentation' of high- ω_7 phenomena in the 'normalized' APES more or less suppresses the low-frequency activation areas, as becomes most apparent in the tonal results (e.g. units 28-2 (c), 35-6 (d) and 26-4 (f)), less prominent in the natural results although also there the effect is present. The effect of the third stage on the temporal structure in the EDS is hardly discernible and certainly not capable of removing the temporal stimulus properties. On the whole the results in Figs. 6 and 7 certainly cannot be called an improvement as compared to the equalization results. They much more tend to restore the gross characteristics of the original APES (e.g. the 'natural' results of units 35-6 (d) and 26-4 (f)) than to join the spectro-temporal sensitivity as determined

with tonal and natural stimuli into one single model.

6. DISCUSSION

In the present paper we have investigated the formal relation between the Average Pre-Event Spectrogram (APES) and the Spectro-Temporal Receptive Field (STRF) of an auditory neuron. The APES has been introduced as an experimental measure of the neuron's spectro-temporal sensitivity for acoustic stimuli and in its implementation was based on the dynamic spectrum representation of sound (Aertsen and Johannesma, 1980; Aertsen et al., 1980, 1981). The STRF on the other hand forms a theoretical concept which aims to connect the general spectrotemporal representation of sound and the corresponding description of neural sensitivity, inspired by the work of Gabor (1946) and Ville (1948), with the general theory of nonlinear systems, as developed by Volterra, Wiener and Lee (Aertsen and Johannesma, 1981).

The basic result is given by relation (21) which expresses the APES in the STRF in the form of a quadratic integral expression, which also contains the stimulus spectrum and the transfer function of the filters in the dynamic spectrum analyser (DSA). This relation is based on two assumptions: 1) the even order nonlinearity in the neuron's system function is of second order and 2) the stimulus ensemble is Gaussian. Under these assumptions the relation (21) enables to predict, for any Gaussian stimulus ensemble, the form of the APES, determined with a DSA-type analyser, from the knowledge of the STRF, the stimulus spectrum and the analyser's filter characteristics.

The inverse relation, a straightforward expression of the theoretical STRF in terms of the experimentally obtained APES is effectively rendered impossible, due to the 'destructive' properties of the spectrogram analysis. The average pre-event spectrogram only presents partial information regarding the second order input-output cross correlation function, a necessary tool for the assessment of the quadratic system kernel, the STRF. This implies that the use of DSA-type analysers to describe the spectro-temporal structure of sound and the neural sensitivity to it, does not allow the derivation of a rigorous stimulus-normalization procedure of the APES as was our original goal. The descriptive value of the APES for the characterization of neural sensitivity therefore is restricted to the type of stimuli it was determined with.

An approximative 'stimulus-normalization' procedure has been proposed, based on the assumption of a DSA-analysis with bandpass filters which can be considered 'sharp' as compared to the frequency dependence of the stimulus spectrum and the STRF. Quite apart from a number of mathematical difficulties associated with the 'approximation' which we certainly do not consider to be solved, the results of application of this approximative normalization procedure were far from being impressive. In fact, the procedure could not be considered to give better results than an earlier, more intuitively based procedure of 'equalizing' the APES for the overall power spectrum of the stimulus ensemble did give (Aertsen et al., 1981).

These results lead to the conclusion that in the context of a formal approach to the investigation of neuron characteristics like the spectro-temporal receptive field and its possible inherent stimulus-dependence, an instrument-based analysis like the APES shows severe shortcomings, which can only be overcome by using more refined and theoretically more firmly and elegantly based methods of spectrotemporal representation of sound. A possible example of such a method has been described as the complex spectro-temporal intensity density function CoSTID (Rihaczek, 1968; Johannesma and Aertsen, 1979; Johannesma et al., 1981). One of its essential characteristics is that, unlike the sonogram-analysis, it preserves phaserelations which are present in the signal. It has succesfully been applied to the characterization of auditory neuron responses to Gaussian wideband noise (Hermes et al., 1981). Furthermore a stimulus-normalization procedure for the CoSTID has been developed, quite similar to relation (1), which appears to provide quite satisfactory results in neuron simulation studies, using as stimulus ensembles various mixtures of Gaussian wideband noise and natural sounds (in preparation).

The unsatisfactory results of the approximative 'normalization'-procedure most probably have also causes in the two basic assumptions underlying a second order normalization procedure not being fulfilled. Both the tonal and the natural stimulus ensemble are certainly far from Gaussian, whereas a reliable statement regarding the order of the even part of the neuron's nonlinearity cannot be given. The natural stimulus ensemble contains very specific spectro-temporal structures, is too specific and too complex to be adequately described by its overall power spectrum. The same holds, albeit to a lower extent, for the tonal ensemble. This leads to the conclusion that neuron characteristics determined for these types of stimulus ensembles have, strictly speaking, only descriptive value when remaining within the same type of stimulus ensemble. This is characteristic of the problems one faces when trying to reconcile the various approaches to the study of neural function, especially in the more central regions of the auditory system. From neuro-ethology we have the demands which lead to complex and (species-) specific natural stimuli. The 'classical' approach in auditory neurophysiology leads to the use of 'simple' stimuli like tones. A general theoretical framework is only provided for Gaussian noise stimuli. This creates a field of tension where adequate strategies have to be developed.

Several ways to proceed are conceivable. One possible way is to remain within the relative safety of general theory as long as possible, i.e. to start from Gaussian white noise and gradually transform it in various possible ways e.g. by linear and/or nonlinear filtering or the progressive addition of non-Gaussian signals. This approach, in fact, sets out to explore the experimental boundaries of applicability of the general theory. An attractive alternative is to start from the inspiration of neuro-ethology, where relevant questions are formulated in terms of concepts like 'information bearing elements' (Suga, 1978) or 'bioacoustic dimensions' (Scheich, 1977), leading to a type of 'parameter variation' of the interesting aspects of natural sounds. The question whether this approach can be formalized into a general framework, making use of the various possibilities of signal and system theory available, certainly forms an interesting challenge for any neurophysiologist with a soft spot for the elegance of mathematics and biophysics. Whatever the approach one chooses, the chances for success will certainly be enlarged by using all the available neurophysiological evidence and turning it into possible models of neural function. The inherent weakness of a complete 'black box' approach, both factually and conceptually, would thereby be reduced to a large extent. Our final conclusion would be that the dynamic spectrum analysis can play a useful role in the auditory neurophysiological experiments where rapid, preferably real-time, feedback to the experimenter may indicate what to look for and where, in the spectro-temporal domain, the questions should be focussed. Once the interesting regions have been located, the characteristics have to be investigated by more adequate and precise, at the same time more (computer) time consuming spectro-temporal procedures.

Acknowledgements. This investigation forms part of the research program of the Workgroup Neurophysics at the Department of Medical Physics and Biophysics of the University of Nijmegen (the Netherlands) and was supported by the Netherlands Organization for Advancement of Pure Research (ZWO). The authors wish to thank Ton Vendrik for helpful comments on the manuscript. Technical support has been given by Jan Bruijns and Wim van Deelen.

REFERENCES

- Aertsen, A.M.H.J., Johannesma, P.I.M.: Spectrotemporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli. Biol. Cybernetics 38, 223-234 (1980)
- Aertsen, A.M.H.J., Johannesma, P.I.M., Hermes, D.J.: Spectro-temporal receptive fields of auditory neurons in the grassfrog. II. Analysis of the stimulus-event relation for

12

tonal stimuli. Biol. Cybernetics <u>38</u>, 235-248 (1980)

- Aertsen, A.M.H.J., Olders, J.H.J., Johannesma, P.I.M.: Spectro-temporal receptive fields of auditory neurons in the grassfrog. Biol. Cybernetics <u>39</u>, 195-209 (1981)
- Aertsen, A.M.H.J., Johannesma, P.I.M.: The spectro-temporal receptive field: a functional characteristic of auditory neurons. In: Spectro-temporal characterization of auditory neurons (Ch. 7). Aertsen, A.M.H.J., Ph.D. Thesis, Nijmegen, the Netherlands 1981
- Bullock, T.H. (ed.): Recognition of complex acoustic signals. Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Deutsch, R.: System Analysis Techniques. Englewood Cliffs, N.J.: Prentice Hall 1969
- Gabor, D.: Theory of communication. J.I.E.E. (London), pt. III, <u>93</u>, 429-457 (1946)
- Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M., Eggermont, J.J.: Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetized grassfrog, Rana temporaria L., investigated with noise stimuli. Hearing Research, in press (1981)
- Johannesma, P.I.M.: Dynamical aspects of the transmission of stochastic neural signals. In: Proc. First European Biophysics Congress, pp. 329-333. Broda, E., Locker, A., Springer-Lederer, H. (eds.). Vienna: Verlag der Wiener Medizinischen Akademie 1971
- Johannesma, P.I.M.: The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: Proc. of the IPO Symp. on Hearing Theory, pp. 58-69. Cardozo, B.L. (ed.). Eindhoven, the Netherlands 1972
- Johannesma, P.I.M., Aertsen, A.M.H.J.: Neural image of sound in the grassfrog. In: Hearing Mechanisms and Speech, pp. 79-86. Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.). Exp. Brain Res. Suppl. II (1979)

- Johannesma, P.I.M.: Functional identification of auditory neurons based on stimulus-event correlation. In: Psychophysical, physiological and behavioural studies in hearing, pp. 77-84. Brink, G. van den, Bilsen, F.A. (eds.). Delft University Press 1980
- Johannesma, P.I.M., Aertsen, A.M.H.J., Cranen, L., Erning, L.J.Th.O. van: The Phonochrome: a coherent spectro-temporal representation of sound. Submitted to Hearing Research (1981)
- Jongh, H.R. de: Modelling the peripheral auditory system. Fitting model pulse patterns to neuronal impulse trains. Ph.D. Thesis, Amsterdam, the Netherlands 1978
- Lee, Y.W., Schetzen, M.: Measurement of the Wiener kernels of a nonlinear system by cross-correlation. Int. J. Control 2, 237-254 (1965)
- Papoulis, A.: The Fourier integral and its applications. New York: McGraw-Hill 1962
- Rihaczek, A.W.: Signal energy distribution in time and frequency. I.E.E.E. Tr. on Inf. Theory <u>IT-14</u>, 369-374 (1968)
- Scheich, H.: Central processing of complex sounds and feature analysis. In: Recognition of complex acoustic signals, pp. 161-182. Bullock, T.H. (ed.) Life Sciences Res. Rep., Vol. 5. Berlin, Heidelberg, New York: Springer 1977
- Suga, N.: Specialization of the auditory system for reception and processing of speciesspecific sounds. Fed. Proc. <u>37</u>, 2342-2354 (1978)
- Ville, J.: Théorie et applications de la notion de signal analytique. Cables et Transmission <u>2</u>, 61-74 (1948)
- Worden, F.G., Galambos, R. (eds.): Auditory processing of biologically significant sounds. Neurosci. Res. Prog. Bull., Vol. 10. Brookline, Mass.: Neurosciences Research Program 1972

CHAPTER 9

- A. Summary
- B. Samenvatting

The investigation, described in this thesis, concentrates on the development of a theoretical description and related experimental procedures which provide for neurons in more central parts of the auditory nervous system a functional representation of the neuron as an element of the auditory processor. This approach is based on statistical analysis of the relation between the presented ensemble of sounds and the extra-cellularly recorded single unit activity. A point of special interest is the question to what extent such a description covers the neuron's behaviour under a variety of different acoustic stimulus conditions such as tones and natural sounds.

The analysis is focused on a comparison of the spectro-temporal properties of stimuli preceding the occurrence of action potentials, the neural events, with the characteristics of the complete ensemble of stimuli presented to the animal. This approach, formally related to the evaluation of the second order stimulus-event cross correlation function, has been implemented in different ways connected to the characteristics of the stimuli used. The analysis results in a representation of the neuron's stimulus-event relation by means of the 'spectro-temporal receptive field' (STRF).

In order to get an impression of the extent of stimulus-invariance of this neuron characteristic, the outcome of the correlation analysis has to be normalized with respect to the a priori known spectro-temporal structure of the stimulus ensemble it was determined with. Therefore the analysis as described has been discussed in the context of the general theory of nonlinear systems. Under certain conditions, regarding both the neuron's system function and the stimulus ensemble, indeed a stimulusnormalization procedure can formally be derived.

The concepts and analysis procedures have been illustrated by the results of their application to extra-cellular single unit recordings from peripheral and more central levels in the auditory system of the anaesthetized cat and frog.

B. SAMENVATTING

Het centrale thema in het in dit proefschrift beschreven onderzoek is de mogelijkheid om voor neuronen in meer centrale delen van het auditieve zenuwstelsel te komen tot een funktionele beschrijving van het neuron als element in het auditieve informatieverwerkende systeem. De aanpak baseert zich hierbij op een statistische analyse van de relatie tussen de aangeboden geluidsverzameling en de extracellulair gemeten 'single unit'-aktiviteit. Een punt van speciale aandacht is de mate waarin de aldus verkregen beschrijving overdraagbaar is voor verschillende typen van geluid zoals tonen en natuurlijke geluiden.

De eerste vraag die rijst is de vraag of er überhaupt een samenhang bestaat tussen de aangeboden stimulusverzameling en de geregistreerde reeks van aktiepotentialen, de neurale events (gebeurtenissen). Dit is onderzocht door de stimulusverzameling twee maal aan te bieden en het effekt hiervan op de neurale aktiviteit te analyseren op reproduceerbaarheid (Hoofdstuk 2). Deze analyse geschiedde met behulp van het kruiskoïncidentiehistogram, gebaseerd op de kruiskorrelatiefunktie van de twee reeksen events. De resultaten van deze methode, toegepast op de aktiviteit van neuronen in de nucleus cochlearis (perifeer) en de auditieve cortex (centraal) van de geanaesthetiseerde kat bij stimulatie met een brede verzameling van natuurlijke geluiden, suggereren dat de vraag naar de existentie van een stimulus-event relatie genuanceerd dient te worden, waarbij aspekten als sterkte en temporele dracht van de relatie een belangrijke rol spelen. Deze blijken samen te hangen met het nivo in het auditieve systeem.

Hoofdstuk 3 beschrijft een kwalitatieve benadering van de vraag in hoeverre de neurale responsies op zulke verschillende geluidsstimuli als tonen en natuurlijke geluiden uit elkaar begrepen kunnen worden. De aanpak hierbij is gebaseerd op een kwalitatieve vergelijking van enerzijds de spektro-temporele gevoeligheid van het neuron zoals bepaald met tonen met anderzijds het gemiddelde vuurpatroon tijdens herhaalde aanbieding van de natuurlijke stimulus, aangevuld met een dynamisch vermogensspektrum (spektrogram) van die stimulus. Hierbij bleek dat voor neuronen op het nivo van de nucleus cochlearis van de kat de responsies, op het gehanteerde verklaringsnivo, globaal gesproken redelijk in overeenstemming waren, terwijl voor neuronen in het meer centraal gelegen corpus geniculatem mediale de response karakteristieken niet in overeenstemming te brengen waren. Dit hoofdstuk besluit met een diskussie over de mogelijke samenstelling van een brede verzameling natuurlijke stimuli op basis van kriteria ontleend aan de oekologie en ethologie van het desbetreffende proefdier.

In de hoofdstukken 4, 5 en 6 wordt nader ingegaan op de in hoofdstuk 3 geformuleerde vraag. In hoofdstuk 4 worden methoden beschreven om te komen tot een gekombineerde spektro-temporele beschrijving van geluid. Deze methoden zijn gebaseerd op een komplexe (in mathematische zin) representatie van signalen: het analytisch signaal. Voor spektraal smalbandige signalen, zoals tonen, kan een spektro-temporele beschrijving worden gegeven op een parametrische wijze door middel van de instantane amplitude en de instantane frekwentie. Voor breedbandige signalen, waaronder vele natuurlijke geluiden, is een meer algemene beschrijving nodig, zoals die kan worden gerealiseerd in een spektrogram. Het funktioneel principe van een hiertoe ontwikkeld instrument, de 'dynamisch spektrum analysator' (DSA) wordt beschreven.

De aard van de stimulus-event relatie van het neuron is onderzocht door middel van statistische analyse van het Pre-Event Stimulus Ensemble (PESE): de verzameling van geluidssegmenten welke onmiddellijk voorafgaan aan de neurale events. Dit PESE vormt een deelverzameling van het totale stimulusensemble (SE) waarvan de samenstelling wordt bepaald door het vuurgedrag van het neuron. Een vergelijking van de karakteristieken van PESE en SE levert informatie omtrent de stimulusgevoeligheid van het neuron. Formeel is deze analyse gerelateerd aan het opmaken van de kruiskorrelatiefunktie van een bepaalde stimulusfunktionaal en de neurale aktiviteit: stimulus-event korrelatie. De analyse heeft zich toegespitst op de spektrotemporele struktuur van het PESE, hierbij gebruik makend van de eerder ingevoerde geluidsbeschrijving. Voor het tonale stimulusensemble leidt dit tot een parametrische weergave van het PESE: de amplitude-frekwentie-tijd kubus (AFT). Kwadratische middeling van amplitudewaarden geeft de gemiddelde geluidsintensiteit als funktie van frekwentie en tijd vóór de neurale event (IFT). Voor het natuurlijke stimulusensemble kan met behulp van de DSA een dimensioneel vergelijkbare karakteristiek worden bepaald: het gemiddelde pre-event spektrogram (APES). Deze analyseprocedures zijn geillustreerd aan de hand van resultaten van hun toepassing op een aantal 'single unit'-afleidingen in de torus semicircularis (de auditieve middenhersenen) van de geanaesthetiseerde bruine kikker, op aanbieding van een uitgebreid ensemble van tonale (hoofdstuk 5) en natuurlijke (hoofdstuk 6) stimuli.

De hier geintroduceerde karakteristieken van het PESE geven een indruk van de spektro-temporele gevoeligheid van het neuron voor het desbetreffende SE. Voor het gevoeligheidsgebied voor bepaalde aspekten van geluid is, geinspireerd op o.a. de visuele neurofysiologie, de term 'receptief veld' geintroduceerd. Toegepast op de hier gebruikte benadering leidt dit tot het begrip 'Spektro-Temporeel Receptief Veld' (STRF) als neuronkarakteristiek. De resultaten van de analyse van het PESE, de IFT resp. het APES, als zodanig geven een onjuist beeld van het STRF aangezien ze, behalve de neurale gevoeligheid, is belangrijke mate de spektro-temporele struktuur van het gebruikte SE reflekteren. Een korrektie voor deze stimuluseigenschappen in de vorm van een stimulus-normalisatie procedure is noodzakelijk. De centrale vraag hierbij is in hoeverre het voor a priori bekende stimuluseigenschappen genormaliseerde STRF kan worden opgevat als een stimulus-invariante neuronkarakteristiek, die de neurale vuurpatronen in uiteenlopende geluidsomgevingen kan verbinden. Stimulus-invariantie van het STRF is geen logisch afleidbare eigenschap, hij dient voor elk neuron afzonderlijk experimenteel onderzocht te worden. Op de resultaten verkregen met het natuurlijke SE is hiertoe in hoofdstuk 6 een intuitief beredeneerde procedure toegepast: het APES werd spektraal gewogen met het gemiddelde vermogen van het SE in de onderscheiden frekwentiebanden: 'stimulus filtering'. Voor een gedeelte van de onderzochte neuronen bleek dit, globaal gesproken, te leiden tot een vergelijkbare gevoeligheid onder tonale en natuurlijke stimuluskondities. Over het geheel genomen evenwel waren de resultaten niet zodanig dat een uitspraak over de stimulus-invariantie van de spektro-temporele gevoeligheid goed mogelijk was.

Om te komen tot een meer formele beschrijving van het STRF en, hieraan gekoppeld, een adekwate stimulus-normalisatie is de beschreven analyseprocedure in hoofdstuk 7 geplaatst in het kader van de theorie voor niet-lineaire systemen zoals ontwikkeld door Volterra en Wiener, waarbij de input en output van het systeem worden gerelateerd via een integraalontwikkeling. Hierbij is het voorstel gedaan om het STRF formeel te relateren aan de 2e orde Volterra kern uit deze integraalreeks. De experimentele bepaling van het STRF noopt tot een formulering in termen van Wiener kernen van even orde, welke via kruiskorrelatie analyse gemeten kunnen worden. Voor een Gaussisch stimulusensemble en een nietlineair systeem met geen even orde bijdragen van orde hoger dan twee blijkt de tweede orde kruiskorrelatiefunktie van stimulus en response te leiden tot het gewenste STRF. Voor meer algemene niet-lineaire systemen en voor stimulusensembles die kunnen worden beschouwd als nietlineair getransformeerde Gaussische ruis zijn evenwel hogere orde kruiskorrelatiefunkties noodzakelijk voor de bepaling van het STRF.

In hoofdstuk 8, tenslotte, wordt een formele relatie gelegd tussen de experimentele analyse via het APES en het theoretisch geformuleerde STRF. Onder de bovenvermelde kondities wat betreft het neuron als systeem en het stimulusensemble wordt een integraalvergelijking afgeleid waarin het APES wordt uitgedrukt in het STRF, de spektrale samenstelling van het SE en de filterkarakteristieken van de DSA. Voor een Gaussisch SE is het mogelijk om op deze wijze vanuit kennis van het STRF de vorm van het APES te berekenen, vooropgesteld dat het STRF inderdaad een stimulus-invariante neuronkarakteristiek is. De omgekeerde weg, vanuit APES naar het STRF, wordt verhinderd door de vorm van

de integraalvergelijking, hetgeen terug te voeren is op 'informatie-vernietigende' eigenschappen van de DSA. Een intuitief beredeneerde, vrij grove benadering van deze integraaluitdrukking leidt tot een 'benaderende stimulusnormalisatie', welke de eerder toegepaste 'stimulus-filtering' als onderdeel omvat. De resultaten van toepassing van deze procedure op een aantal van de resultaten zoals beschreven in hoofdstuk 5 en 6 geven aan dat een spektrotemporele analyse met behulp van DSA-achtige instrumenten, hoewel zinvol voor het verkrijgen van een globaal beeld, voor een meer kwantitatieve benadering van de vraag naar een stimulusinvariant STRF niet toereikend is. Deze dient vervangen te worden door een algemenere en theoretisch gefundeerde spektro-temporele signaalbeschrijving, waartoe inmiddels een alternatief is ontwikkeld in de vorm van de komplexe spektro-temporele intensiteitsdichtheidsfunktie (CoSTID).

Samenvattend kan worden gesteld dat er m.b.t. de funktionele beschrijving van neuronen in meer centrale delen van het auditief zenuwstelsel een duidelijk spanningsveld blijft bestaan tussen enerzijds de wensen zoals die komen vanuit de nu beschikbare algemene theorie, met name de eis van het Gaussische karakter van het stimulusensemble, en anderzijds de uiterst relevante vraag naar het gebruik van komplexe, natuurlijke geluiden zoals die wordt gesteld vanuit de neuro-ethologie. Vooralsnog lijkt een benadering waarbij de inspiratie vanuit theorie en experiment een komplementaire rol vervullen meer kans van slagen te bieden dan enerzijds een louter formele benadering, waarbij de noodzakelijke restrikties wat betreft stimulusensemble en neuroneigenschappen het karakter van een keurslijf dreigen aan te nemen en anderzijds een louter pragmatische benadering waarbij de onderzochte stimulus-response relaties niet in een wijder theoretisch kader kunnen worden geplaatst.

Het in dit proefschrift beschreven werk maakt deel uit van het onderzoek, verricht door de Werkgroep Neurofysika van het Laboratorium voor Medische Fysika en Biofysika aan de Universiteit van Nijmegen. Dit onderzoek wordt gesteund door de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.). De Werkgroep Neurofysika funktioneert zowel in de Werkgemeenschap Auditief Systeem van de Stichting voor Biofysika als in de Nijmeegse Interfakultaire Werkgemeenschap "Hersenen en Gedrag". Afgezien van hetgeen in dit proefschrift is gepresenteerd wordt van het onderzoek van de Werkgroep Neurofysika tevens verslag gedaan in de hier volgende recente publikaties:

- Johannesma, P.I.M., Aertsen, A.M.H.J.: Neural image of sound in the grassfrog. In: Hearing mechanisms and speech, pp. 79-86. Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.). Exp. Brain Res. (Suppl. II) (1979)
- Aertsen, A.M.H.J., Johannesma, P.I.M.: Spectrotemporal analysis of auditory neurons in the grassfrog. In: Hearing mechanisms and speech, pp. 87-93. Creutzfeld, O., Scheich, H., Schreiner, Chr. (eds.). Exp. Brain Res. (Suppl. II) (1979)
- Johannesma, P.I.M.: Functional identification of auditory neurons based on stimulus-event correlation. In: Psychophysical, physiological and behavioral studies in hearing, pp. 77-84. Brink, G. van den, Bilsen, F.A. (eds.). Delft: Delft University Press 1980
- Aertsen, A.M.H.J., Eggermont, J.J., Hermes, D.J., Johannesma, P.I.M., Vendrik, A.J.H.: Spectro-temporal receptive fields of auditory neurons in the midbrain of the grassfrog. In: Stichting voor Biofysika, Verslag over het jaar 1979-80, pp. 108-118. Den Haag 1980
- Eggermont, J.J., Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M.: Response properties and spike waveforms of single units in the torus semicircularis of the grassfrog (Rana temporaria) as related to recording site. In: Neuronal mechanisms in hearing. Aitkin, L., Syka, J. (eds.). New York: Plenum Press (in press) 1981
- Johannesma, P.I.M.: Neural representation of sensory stimuli and sensory interpretation of neural activity. In: Proc. Symp. 'Neural Communication and Control; Facts and Theories'. Debrecen (Hungary), in press (1981)
- Eggermont, J.J., Aertsen, A.M.H.J., Hermes, D.J., Johannesma, P.I.M.: Spectro-temporal characterization of auditory neurons: redundant or necessary? Hearing Research, in press (1981)
- Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M., Eggermont, J.J.: Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetized grassfrog, Rana temporaria L., investigated with noise stimuli. Hearing Research, in press (1981)
- Johannesma, P.I.M., Aertsen, A.M.H.J., Cranen, L., Erning, L.J.Th.O. van: The phonochrome: a coherent spectro-temporal representation of sound. Submitted to Hearing Research (1981)

Curriculum Vitae.

Ik werd geboren op 20 oktober 1948 in Oosterhout (N.B.). De HBS-b opleiding, gevolgd aan het Mgr. Frencken College in Oosterhout, heb ik afgerond in 1966. In september 1966 begon de studie natuurkunde aan de Rijks Universiteit Utrecht. Het afstudeeronderzoek vond plaats in de Vakgroep Medische en Fysiologische Fysika o.l.v. prof.dr. J.J. Denier van der Gon en dr. G.E.P.M. van Venrooy en had betrekking op een onderwerp uit de cytofysika/elektronenmikroskopie: 'Freeze-etching - freezing velocity and crystal size'. Op 5 november 1973 werd deze studie afgesloten met het doctoraalexamen Experimentele Natuurkunde.

Vanaf 1 september 1973 ben ik werkzaam als wetenschappelijk medewerker op het Laboratorium voor Medische Fysika en Biofysika aan de Katholieke Universiteit van Nijmegen, en wel in de werkgroep Neurofysika o.l.v. prof.dr. A.J.H. Vendrik, prof.dr. J.J. Eggermont en dr. P.I.M. Johannesma. Het onderzoek in deze werkgroep was gericht op de spektro-temporele karakterisering van auditieve neuronen in kat en kikker. Een gedeelte van de hierbij ontwikkelde methodes en verkregen resultaten is weergegeven in dit proefschrift. Het onderzoek vond plaats in dienstverband met, achtereenvolgens, ZWO (Stichting voor Biofysika) en de Medische Fakulteit van de KUN. Vanaf 1 november 1980 ben ik betrokken bij onderzoek in deze zelfde werkgroep naar 'de neurale wisselwerking in het centrale zenuwstelsel van de bruine kikker in relatie tot auditieve perceptie', in het kader van de universitaire onderzoekspool, via de Fakulteit Wiskunde en Natuurwetenschappen van de KUN.

Vanaf februari 1974 deel ik mijn leven met Mieke Nijboer. Op 3 januari 1977 kwam onze dochter Woosje ons daarbij gezelschap houden.

STELLINGEN

- 1. De vraag of een neuron funktioneel deel uitmaakt van een bepaald sensorisch systeem kan in kwantitatieve termen worden onderzocht door de reeksen aktiepotentialen die worden afgeleid tijdens herhaalde aanbieding van een gevarieerd stimulus ensemble statistisch te onderzoeken op onderlinge gelijkenis.
 - Pedersen, J.G.: On the existence of stimulus-event relations for sensory neurons: a statistical method. Biol. Cybernetics <u>38</u>, 201-212 (1980).
 Dit proefschrift, hoofdstuk 2.
- Het begrip 'receptief veld' van een sensorisch neuron dient te worden geformaliseerd. Hierbij kunnen begrippen en methoden ontleend aan de systeemen signaaltheorie een zinvolle funktie vervullen.
 - Yasui, S., Davis, W., Naka, K.-I.: Spatio-temporal receptive field measurement of retinal neurons by random pattern stimulation and cross correlation. IEEE Tr. on Biomed. Eng. BME-<u>26</u>, 263-272 (1979).

Dit proefschrift, hoofdstuk 7.

- 3. Voor de vaak impliciet gemaakte vooronderstelling dat de spektro-temporele stimulusgevoeligheid van auditieve neuronen te scheiden zou zijn in twee onafhankelijke komponenten, t.w. een spektrale en een temporele komponent, welke afzonderlijk onderzocht kunnen worden, bestaat geen logische grond terwijl ze in een aantal aanwijsbare gevallen m.n. in meer centrale delen van het zenuwstelsel, feitelijk onjuist is.
 - Eggermont, J.J., Aertsen, A.M.H.J., Hermes, D.J., Johannesma, P.I.M.: Spectro-temporal characterization of auditory neurons: redundant or necessary? Hearing Research, ter perse (1981). Dit proefschrift.
- 4. Bij de keuze uit vier isomorfe 2^e orde representaties van signalen wordt de voorkeur voor een spektro-temporele beschrijving in belangrijke mate bepaald door intuitieve vertrouwdheid ontleend aan de auditieve perceptie.
 - De Weerd, J.P.C.M.: Estimation of evoked potentials. A study of a posteriori "Wiener" filtering and its time-varying generalization. Proefschrift, Nijmegen 1981.
 Dit proefschrift.

5. Geluid kan worden weergegeven in de vorm van een kleurenbeeld: de 'phonochroom'. De inverse operatie, de 'chromophoon', is in principe uitvoerbaar en leidt voor een willekeurig kleurenbeeld tot een ensemble van geluiden.

> Johannesma, P.I.M., Aertsen, A.M.H.J., Cranen, L., Erning, L.J.Th.O. van: The phonochrome: a coherent spectro-temporal representation of sound. Hearing Research, ter perse (1981).

- 6. Indien we al in staat zijn om ons een inzichtelijk beeld te vormen van de buitenwereld gebeurt dit via de elektrische aktiviteit van de neuronen; indien we al in staat zijn om ons een inzichtelijk beeld te vormen van de elektrische aktiviteit van de neuronen gebeurt dit via de buitenwereld.
 - Johannesma, P.I.M.: Neural representation of sensory stimuli and sensory interpretation of neural activity. In: Proc. Symp. 'Neural communication and control; facts and theories'. Debrecen, Hongarije, ter perse (1981).
- 7. Het gebruik van termen als 'komplexe stimuli' en 'feature detectors' in de studie van het auditief zenuwstelsel stimuleert tot weinig verhelderende diskussies over de vermeende specificiteit van meer centraal gelegen neuronen.
- 8. Bij onderzoek aan processen die zich afspelen in neurale netten waarbij de neurale interaktie een belangrijke rol speelt verliezen begrippen als de orde en het receptief veld van een neuron zowel hun intuitieve als hun formele eenvoud.
- 9. Wetenschappelijk onderzoek is gebaseerd op komplementaire bijdragen van kreativiteit en effektiviteit. Een strakke projekmatige organisatie van onderzoek bevordert effektiviteit ten koste van kreativiteit.
- 10. Het organiseren van wetenschappelijke symposia met steun vanuit NAVO-fondsen betrekt deze symposia in het kultureel-wetenschappelijk alibi van de NAVO en dient, gelet op het dominant-militaire karakter van deze organisatie, vermeden te worden.
- 11. Ervaring met het werken met komplexe opstellingen rondom biofysisch-neurofysiologische experimenten wijst uit dat het merendeel van alle mogelijke storingen, denkbare zowel als ondenkbare, in de praktijk ook daadwerkelijk optreedt. Dit ervaringsfeit, gevoegd bij de alleszins redelijke veronderstelling dat kernfysici i.h.a. niet kapabeler zijn dan biofysici doet het ergste vermoeden omtrent de veiligheid van het gebruik van kerncentrales.

Nijmegen, 17 juni 1981

Ad Aertsen

#