Quarterly Reviews of Biophysics 16, 3 (1983), pp. 341–414
Printed in Great Britain

Reverse-correlation methods in auditory research

J. J. EGGERMONT, P. I. M. JOHANNESMA
AND A. M. H. J. AERTSEN*
Department of Medical Physics and Biophysics,
University of Nijmegen, Nijmegen, The Netherlands

* Present address: Department of Physiology, School of Medicine, University of Pennsylvania, Philadelphia 19104, U.S.A.

1. INTRODUCTION

Single unit recordings have provided us with a basis for understanding the auditory system, especially about how it behaves under stimulation with simple sounds such as clicks and tones. The experimental as well as the theoretical approach to single unit studies has been dichotomous. One approach, the more familiar, gives a representation of nervous system activity in the form of peri-stimulus-time (PST) histograms,

period histograms, iso-intensity rate curves and frequency tuning curves. This approach observes the neural output of units in the various nuclei in the auditory nervous system, and, faced with the random way in which the neurons respond to sound, proceeds by repeatedly presenting the same stimulus in order to obtain averaged results. These are the various histogram procedures (Gerstein & Kiang, 1960; Kiang *et al.* 1965).

We will call this approach the *experimenter-centred* approach; the experimentally observed variable is the occurrence of action potentials (spikes), commonly expressed as a firing rate. One may also use criteria as just detectable changes in firing rate, e.g. in the construction of frequency-tuning curves, and more recently also changes in synchrony between the occurrence of the spikes and the stimulus. The stimuli used in this approach are the familiar ones such as clicks, tone- and noise-bursts or continuous tones of various frequencies. The idea once again is to obtain estimates of firing probabilities or changes therein by repeatedly presenting the same stimulus. Post-stimulus time-histograms and period histograms are thus obtained by averaging or, using a different expression, by correlation techniques. One calculates the occurrence of spike activity as a function of time, τ, *after* the onset of stimulus presentation. The experimenter-centred approach is therefore a method of *forward correlation.*

The second approach, pioneered by De Boer (1967, 1968), and extended and elaborated by Johannesma (1972), Møller (1973), Grashuis (1974), and van Gisbergen *et al.* (1975) may be called a *subject-centred* approach. This way of thinking considers the occurrence of an action potential (event) as a sign or as an indication that something particular happened with the stimulus preceeding that action potential. Each action potential is considered as signalling a stimulus that was of interest to the neuron and may be even to the animal. For this type of approach the stimulus preferably is diverse in nature, and the pioneering papers reported results based on Gaussian wide-band noise as a stimulus. By correlating the events with the stimulus in a given interval prior to the event one obtains an estimate of the average stimulus that caused the spikes. This correlation procedure where one looks at the average stimulus as a function of time, τ, prior to the spike is therefore a method of backward- or *reverse-correlation* (De Boer, 1968). The procedure consists of averaging an ensemble of signals of a given length (e.g. 20 ms) that are immediately preceding a spike. This particular ensemble is called the pre-event stimulus ensemble (PESE). One can imagine that if, for instance, an auditory nerve fibre is tuned to a given

sound frequency, say 1 kHz, only those segments out of the noise stimulus that contain energy in the 1 kHz region will elicit action potentials. Therefore the PESE will have more energy in the 1 kHz region than the overall stimulus ensemble. This distinguishes nervous activity evoked by a stimulus from that termed as spontaneous activity: in this case the PESE will have the same properties as the stimulus ensemble.

The subject-centred approach is generally based on the use of Gaussian wide-band noise as a stimulus and is therefore also called the white-noise approach (Marmarelis & Marmarelis, 1978). It is basically one of the familiar identification methods used in the study of linear systems but having the powerful option to be used in the study of non-linear systems. The earlier studies used this noise stimulus and obtained an estimate of the neurons impulse response, thereby in fact considering the auditory system up to the auditory nerve fibres or cochlear nucleus units as basically linear. Assuming such a linear system, it was expected that Fourier transformation results in a spectrum that in shape is comparable to the frequency-tuning curves as they will result from the experimenter-centred approach.

The introduction of the concept of the PESE (Johannesma, 1972) allowed a much broader scope: without loss of generality the reverse correlation method could be used for all statistically structured stimulus ensembles. One of these consists of a sequence of tonepips for which the amplitude as well as carrier frequency are randomly selected out of 127 respectively 255 values (Aertsen & Johannesma, 1980) and in which the intervals can also be selected freely. Another potentially important extension is to use an ensemble of natural sounds, the acoustic biotope (Aertsen, Olders & Johannesma, 1981).

Segundo (1970) remarks: 'a relation between two activities can be approached prospectively and retrospectively. The viewpoint applied to the forward or prospective way is that of the current experimental strategy in which the investigator imposes certain conditions and then observes their effects. The backward-, reverse-, or retro-spective viewpoint reflects how a subject must proceed when using his sensory input to identify some stimulus condition.'

The white-noise approach has been applied with potential success in vision research and in the study of the vestibular system (cf. Marmarelis & Marmarelis, 1978). Although its application in auditory research was earlier than in the other fields mentioned, it stayed relatively limited in its use. In fact all developments originated from

Europe, in particular The Netherlands (De Boer, Johannesma) and Sweden (Møller), which may explain part of its restricted use. When Johnson (1980) in a theoretical investigation to the use of the white-noise approach in auditory research remarks that since the auditory system is more non-linear, i.e. contains higher-order non-linearities, than the visual or the vestibular system the white-noise approach is in principle less suited because of computational problems in evaluating the higher-order kernels, it is overlooked that the alternative, more familiar forward methods are *all* based on linear assumptions. There is in our opinion no reason to dismiss a method on theoretical grounds because it is not ideal and instead to stay with methods which may be far less ideal but have a longer history of application.

We have the feeling that unfamiliarity with the method in general, both from a conceptual point of view – one must learn the retrospective approach and to use statistically structured stimuli – and from a computational point of view, has limited its use. It is our aim therefore to describe as lucid as possible the underlying principles. This will lead us into the field of signal analysis and system theory as well as to a comparison of results obtained for the auditory nervous system by the two approaches: the forward approach and the reverse approach. Especially we will investigate whether one gains information by using the subject-centred approach that is not obtainable by the more traditional experimenter-centred one. Are there conflicting results and what does that teach us about the auditory system? Should one approach be favoured above the other or should one be inclined to use the best of both worlds?

Application of the reverse-correlation method so far has been to the auditory nerve in cat, rat, guinea pig and caiman, to the cochlear nucleus complex in cat and rat, to the torus semicircularis (a homologue of the inferior colliculus) in the grassfrog, and in the medial geniculate body in the cat. We will review the results in detail with respect to those from the more traditional approaches, elucidate where results diverge or new features appear.

2. Linearity and non-linearity in the auditory system

The initial observations about non-linear action in the auditory system came from psycho-acoustics. It appeared that usually there are more frequencies present in the audible spectrum than in the

sound stimulus, among these the well known 'residue' and the quadratic and cubic difference tones (for a review: De Boer, 1976*a*). The cubic difference tone is, as its name suggests, a third-order distortion product of the form $2f_1-f_2$, where f_1 is the lower frequency and f_2 the higher-frequency tone in a two-tone complex. Distortion products of this order can arise when a third-order non-linearity is present,

$$F(a) = c_1 a + c_2 a^2 + c_3 a^3, \tag{2.1}$$

that transforms the sound input $x(t)$

$$y(t) = F\{x(t)\}. \tag{2.2}$$

These cubic difference tones can be detected in the activity of auditory nerve fibres (Goldstein & Kiang, 1968), in the response of inner haircells (Sellick & Russell, 1979), in the evoked acoustical emissions from the ear (Kemp & Chum, 1980), and it is likely that they originate from a non-linearity inherent to the combination basilar membrane and outer haircells. There is only marginal doubt that the activity that gives rise to the perception of the cubic difference tone will be found in the vibration pattern of the basilar membrane in animal cochleas which are physiologically in excellent condition (Sellick, Patuzzi & Johnstone, 1982).

Linearity in the auditory system seems to be confined to the action of the middle ear, with the additional restriction that sound intensity should be within the physiological range. When studying the auditory nervous system one therefore studies a system with a non-linearity of at least third order. All measures of auditory nervous system activity therefore are special to the particular stimulus type and stimulus level used. Current interest in the use of complex sounds such as speech to study the auditory system (e.g. Young & Sachs, 1979) has made it abundantly clear that the nerve fibre responses thereto cannot be predicted in an analytic way from the information gained by the responses to simple stimuli as clicks or tonebursts. We are therefore in dire need for a method that really *identifies* the non-linear auditory system. Such a method is offered in principle by the reverse-correlation approach, although at the expense of lengthy computations. However, an essential assumption is that the system is time-invariant, has finite memory, and has constant parameters. Furthermore, the order of the system has to be finite.

A disturbing point common to the work on nervous systems is the discrete output of the neuron: a random series of uni-directional pulses. The solution to the identification problem is in this case

potentially less accurate than in case the input as well as the output signal are continuous (Johnson, 1980).

3. Linear systems analysis with examples for the middle ear

A linear system can be characterized uniquely by its impulse response, $h(\tau)$, or equivalently by its transfer function $\hat{h}(\omega)$. Consequently two techniques are in use which result in either one of these measures. The first technique comprises exciting the system under study with an impulse and observing the output of the system thereto. The other technique requires stimulation with a pure tone of varying frequency while keeping track of the output signal which is also a pure tone of the same frequency; the amplitude ratio and phase difference with respect to the input signal characterise the system. It is well known that $\hat{h}(\omega)$ is the Fourier transform of $h(\tau)$.

The uniqueness of the characterization results from the linearity of the system:

$$f(x_1+x_2) = f(x_1)+f(x_2) \tag{3.1}$$

$$f(cx) = cf(x). \tag{3.2}$$

Therefore it is irrelevant at what stimulus level we study the system, or if we investigate it with a single frequency at a time or with a stimulus containing all frequencies of interest such as the impulse or noise.

Because of the linearity the system, $y(t)$ to any stimulus $x(t)$ can be predicted, provided we know its impulse response, $h(\tau)$, by the convolution integral

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, \mathrm{d}\tau. \tag{3.3}$$

Because of the causality, $h(\tau) = 0$ for $\tau < 0$. Conversely if we know the transfer function $\hat{h}(\omega)$, the spectrum of the response $\hat{y}(\omega)$ can be computed from the spectrum of the stimulus by multiplication

$$\hat{y}(\omega) = \hat{h}(\omega)\, \hat{x}(\omega), \tag{3.4}$$

in which $\hat{y}(\omega)$, $\hat{h}(\omega)$ and $\hat{x}(\omega)$ are the Fourier transforms of respectively $y(t)$, $h(t)$ and $x(t)$. We note that a convolution in the time domain transforms into a multiplication in the frequency domain.

This small piece of theory forms the basis for the understanding why such a variety of systems have been investigated as if they were linear. Furthermore we can now understand why the use of tones,

clicks or noise can result in the same basic information. We will illustrate this with measurements on the middle ear system in amphibians.

(*a*) *The harmonic analysis or sweep frequency approach*

This is the classical way to study input–output relationships of any system. For a constant sinusoid of varying frequency the amplitude ratio and phaseshift between output sinewave and input sinewave are determined. The results are usually presented in plots of gain and phase versus frequency, also known as Bode plots. For the middle ear this type of analysis has been made by, for example, recording vibrations of the stapes foot-plate for sinusoidal sound stimulation of the tympanic membrane. Recording these minute vibrations ($\simeq$ 0·1 to 100 nm) requires laser interferometry, Mössbauer techniques or capacitive probe techniques. An example is shown in Fig. 1, where the vibration pattern of the tympanic membrane of the grassfrog as recorded with laser–doppler interferometry is represented by a Bode plot (De Vlaming *et al.*, to be published).

(*b*) *The structured multi-frequency stimulus approach*

In this context, particularly, the impulse or click has been used. We have already seen that Fourier transformation of the impulse response results in the transfer function $\hat{h}(\omega)$. This $\hat{h}(\omega)$ is a complex valued function consisting of modulus, $|\hat{h}(\omega)|$ and argument $\phi(\omega)$ such that

$$\hat{h}(\omega) = |\hat{h}(\omega)|\, e^{-i\phi(\omega)}. \tag{3.5}$$

Modulus and argument are identical to gain respectively phase in the Bode plot. Because of the smallness of the vibrations the click response for the tympanic membrane is obtained by averaging over a large number of click presentations. An example again recorded by laser–doppler interferometry is shown in Fig. 2. Since the middle ears were identical to those shown in Fig. 1, the results should be fully comparable.

(*c*) *The use of unstructured multi-frequency stimuli : the white-noise approach*

In this case wide-band noise or pseudo-random noise is used as input signal. While noise originates from a physical fluctuation process such as Brownian motion or other thermal agitation processes, pseudo-random noise is usually generated digitally using a shift-register with proper feedbacks and a subsequent low-pass filtering. This pseudo-

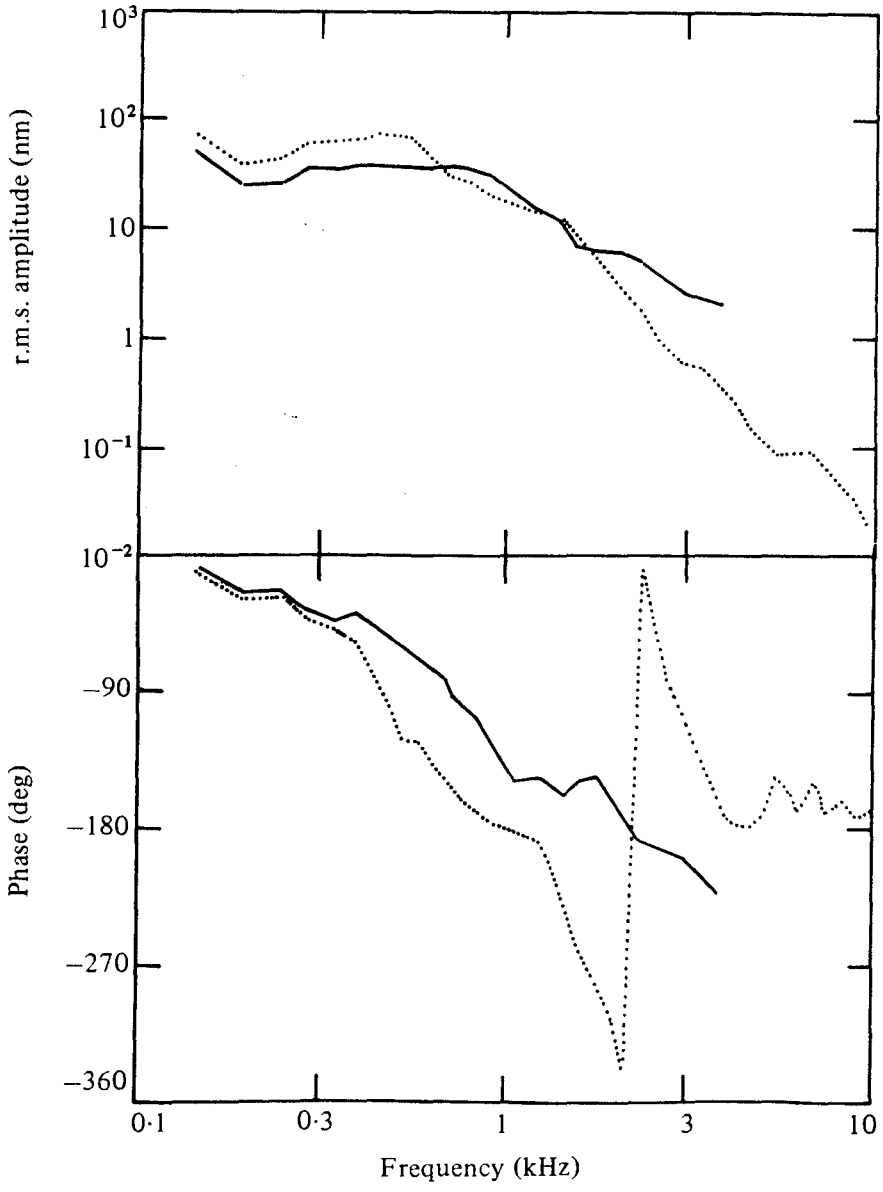


Fig. 1. Frequency-response of the tympanic membrane of the grassfrog. Movements of the tympanic membrane at discrete frequencies were recorded using a laser–doppler velocity meter. The upper part of the figure shows the movement amplitude for two frog ears, the lower part of the figure the corresponding phase-shifts. The response is low-pass of approximately third order.

random noise therefore is reproducible while real noise is not; the reproducibility is an advantage when averaging is needed. The auto-correlation function

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,x(t+\tau)\,dt \qquad (3.6)$$

for white noise is a Dirac δ-function:

$$R_{xx}(\tau) = P\delta(\tau),$$

where P is the power or variance of the noise. For wide-band or

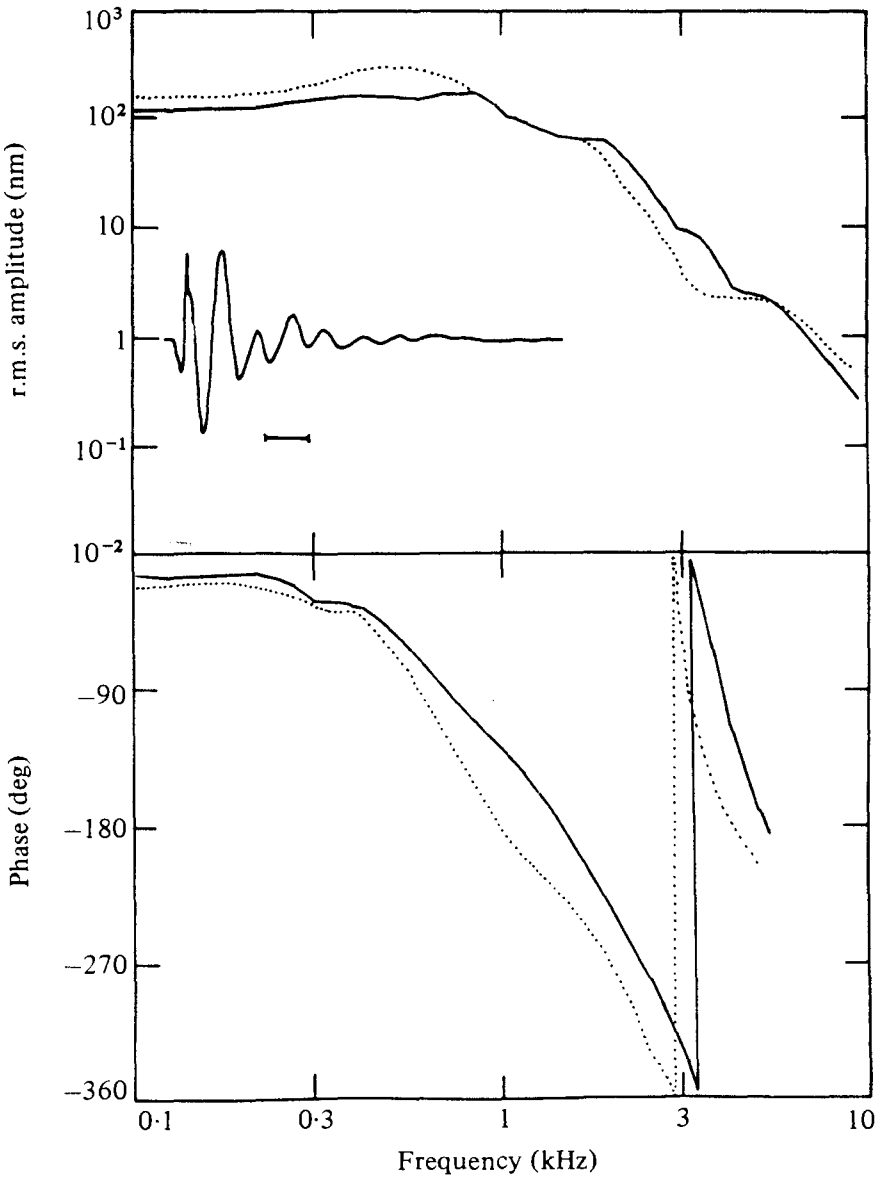


Fig. 2. Impulse response and transfer function of the tympanic membrane of the grassfrog. The same recording technique and same ears as in Fig. 1. Stimulation was with an acoustic click of slightly different level as the tonal stimulus in Fig. 1. The overall behaviour is about the same; the phase-spectra are more equal in this case than in the case for tonal stimulation. Marker duration at the insert is 1 ms.

band-limited white noise the auto-correlation resembles the waveform of an acoustic click. A well-designed pseudo-random noise signal has the same autocorrelation structure and power spectral density as band-limited white noise but deviations in the higher order auto-correlation functions may exist (Swerup, 1978; Eckhorn & Pöpel, 1979). This aspect of pseudo-random noise is particularly relevant in the analysis of non-linear systems.

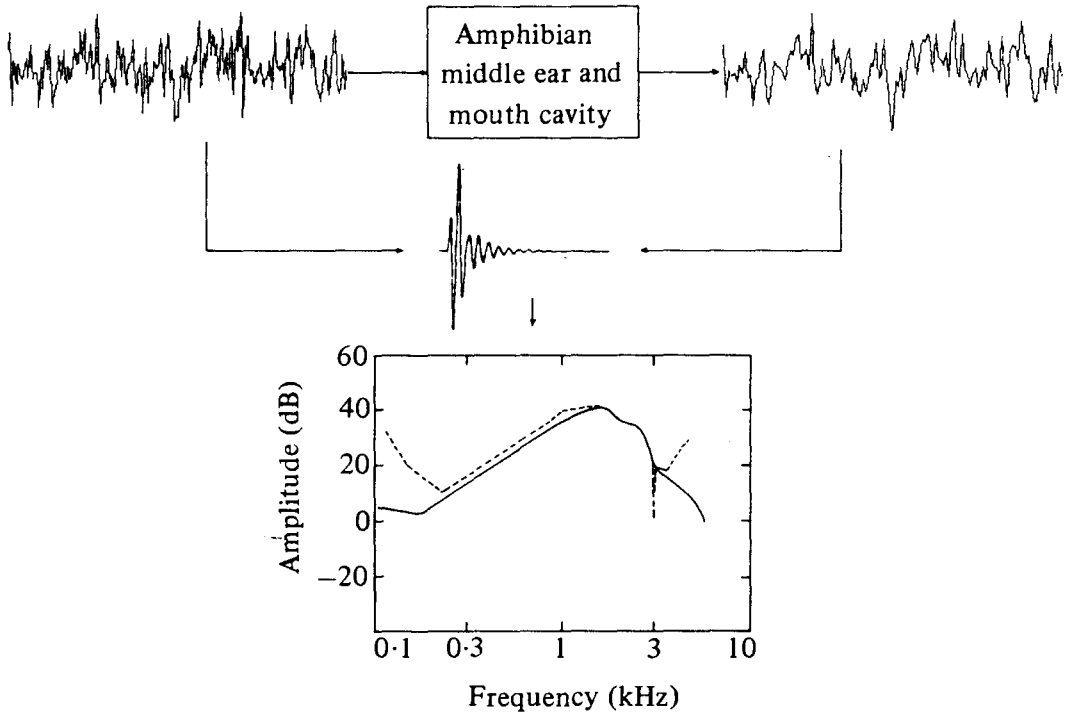


Fig. 3. Cross-correlation method to determine the impulse response and transfer function of the frog's middle ear and mouth cavity. Stimulation was with 5 kHz wide-band noise using a closed sound system; pressure was recorded in the mouth cavity, which has an open connexion to the middle ear using a probe microphone. By cross-correlating the recorded sound-pressure variations with the applied wide-band noise the impulse response as shown was obtained. Fourier transformation resulted in the amplitude spectrum (drawn line). For comparison the transfer function as obtained from click stimulation is shown as a dashed line. Amplitude scale is to an arbitrary reference.

By determining the auto-correlation function of the input signal

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,x(t+\tau)\,\mathrm{d}t \tag{3.7}$$

and the cross-correlation function of output and input signal

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,y(t+\tau)\,\mathrm{d}t \tag{3.8}$$

one is able to compute either the systems impulse response or its transfer function. From (3.8) we derive for $x(t)$ is white noise and using equation (3.3) that

$$R_{xy}(\tau) = h(\tau), \tag{3.9}$$

i.e. the cross-correlation function is equal to the impulse response. The Fourier transforms of the various correlation functions are the power spectral densities $\hat{R}_{xx}(\omega)$, $\hat{R}_{yy}(\omega)$ and the cross-spectral density

$\hat{R}_{xy}(\omega)$. The following relation holds and covers also the case of non-white noise:

$$\hat{h}(w) = \frac{\hat{R}_{xy}(\omega)}{\hat{R}_{xx}(\omega)}. \tag{3.10}$$

The method is illustrated in Fig. 3 together with a result for the frogs middle ear.

4. Linear systems analysis applied to the auditory nerve fibre responses

So far we have dealt with continuous input signals and continuous output signals. Because our interest moves toward neural phenomena, i.e. situations where the output signal can be considered as a point process, we must evaluate whether the linear systems analysis techniques can still be used.

Before starting the analysis in a formal way, first of all a survey will be given of the various experimental procedures, mostly based on assumptions of linearity, that are in use for the auditory system.

(*a*) *Sweep-frequency methods*

In these procedures the neural firing rate is plotted as a function of stimulus frequency for pure tones or long tone bursts as stimuli. These curves are determined for a large range of stimulus intensities and are called iso-intensity contours (e.g. Rose *et al.* 1971). In addition, the phase differences between stimulus and response are evaluated from period histograms (e.g. Anderson *et al.* 1971). Both measures taken together are formally related to Bode plots. An example is shown in Fig. 4. One observes that the shape of the iso-intensity contours changes as a function of stimulus level, so does the phase. This is indicative for a non-linearity in the system.

While iso-intensity contours are constant input level curves, the so-called frequency-tuning curves (e.g. Kiang *et al.* 1965) depend on a constant output criterion such as an increase of, for example, 20% over the spontaneous firing rate. In a linear system one could convert iso-intensity contours into tuning curves. Typical tuning curves for auditory nerve fibres are shown in Fig. 5, together with their rate-intensity functions.

(*b*) *Methods based upon click stimulation*

By calculating a PST histogram for the firings produced by repeated click stimulation one observes, especially for low characteristic

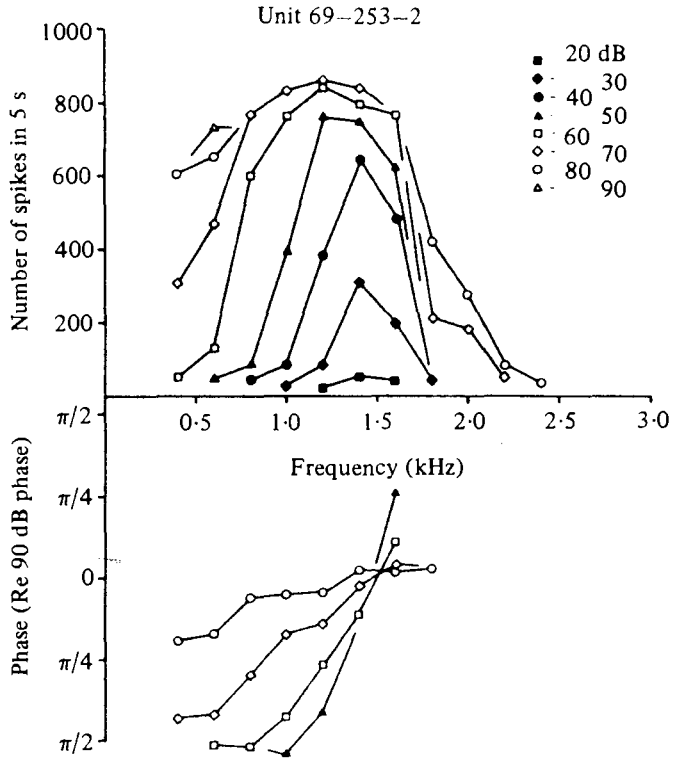


Fig. 4. Frequency-response for firing rate and cumulative phase for an auditory nerve fibre at tonal stimulus levels of 20–90 dB SPL. The change in shape of the response areas (upper half) and slope of the cumulative phase indicates that the fibre's behaviour is non-linear. (From Anderson *et al.* 1971.)

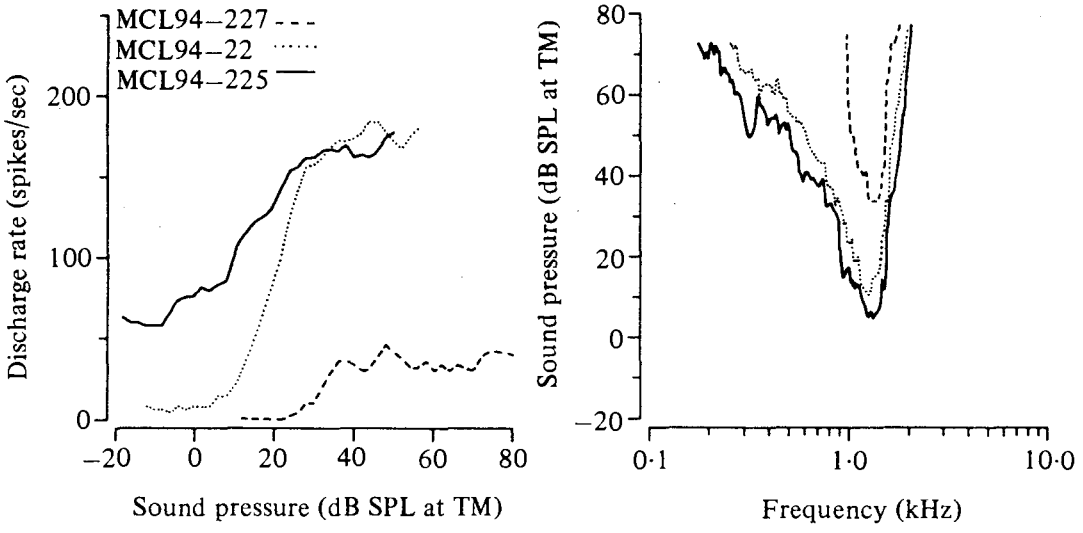


Fig. 5. Rate-intensity functions and frequency tuning curves for three types of cat auditory nerve fibres: low, medium and high spontaneous rate. Low spontaneous rate fibres appear to have elevated threshold, but otherwise show the same tuning as the other two categories. (From Liberman, 1978.)

frequencies, that the firings on average appear to be spaced at intervals equal to the period of the characteristic frequency (CF). Stimulating with clicks of opposite polarity (rarefraction *v.* condensation clicks) one notices that the firings are now at a different position

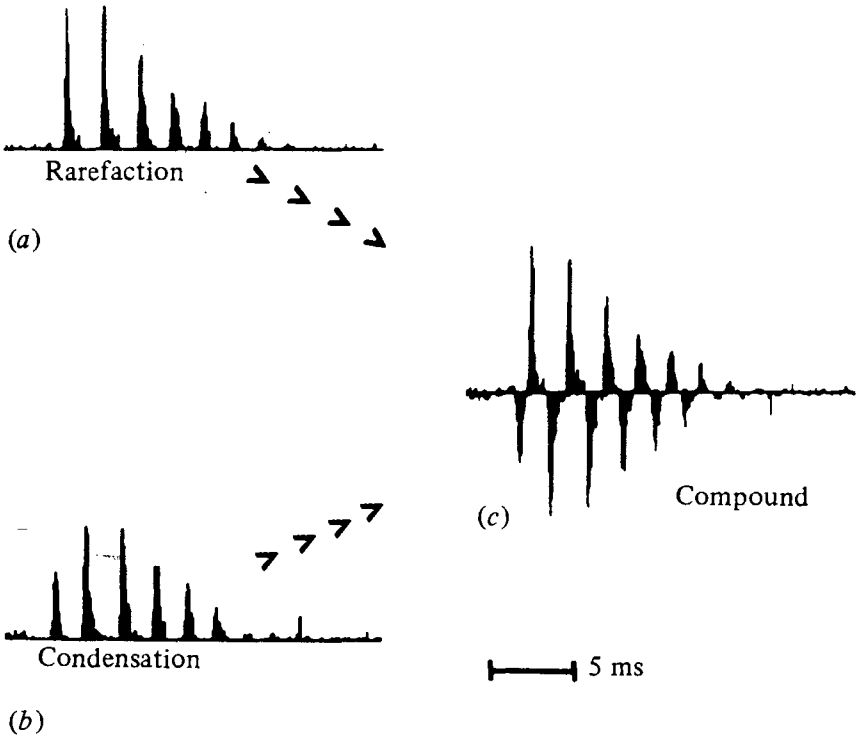


Fig. 6. The construction of a compound PST histogram to click stimulation. The PST histogram obtained for rarefaction clicks and the inverted PST histogram for condensation clicks are put in time registration with each other to give the compound PST histogram. As defined in the text, the appropriate way would be to subtract the PST histogram to condensation clicks from that obtained with rarefaction clicks. (From Pfeiffer & Kim, 1972.)

but that the peaks are still spaced by CF^{-1} ms. This is illustrated in Fig. 6.

Denoting the PST histogram to rarefaction clicks with $n_+(t)$ and that to condensation clicks by $n_-(t)$ one may define the compound PST histogram (Goblick & Pfeiffer, 1969) by

$$n_0(t) = n_+(t) - n_-(t). \tag{4.1}$$

A series of compound PST histograms is shown for fibres of different CF in Fig. 7.

(c) *A model for the peripheral auditory nervous system*

At this point we introduce a model representing current thinking about the auditory periphery (excluding the linear middle ear). It consists of a non-linear filter situated in the basilar membrane that receives energy from an active mechanism (probably the outer haircells); this is followed by a synaptic mechanism that is responsible for auditory adaptation (e.g. Eggermont, 1973, 1975; Smith, 1979; Harris & Dallos, 1979) and finally a pulse-generating mechanism, the

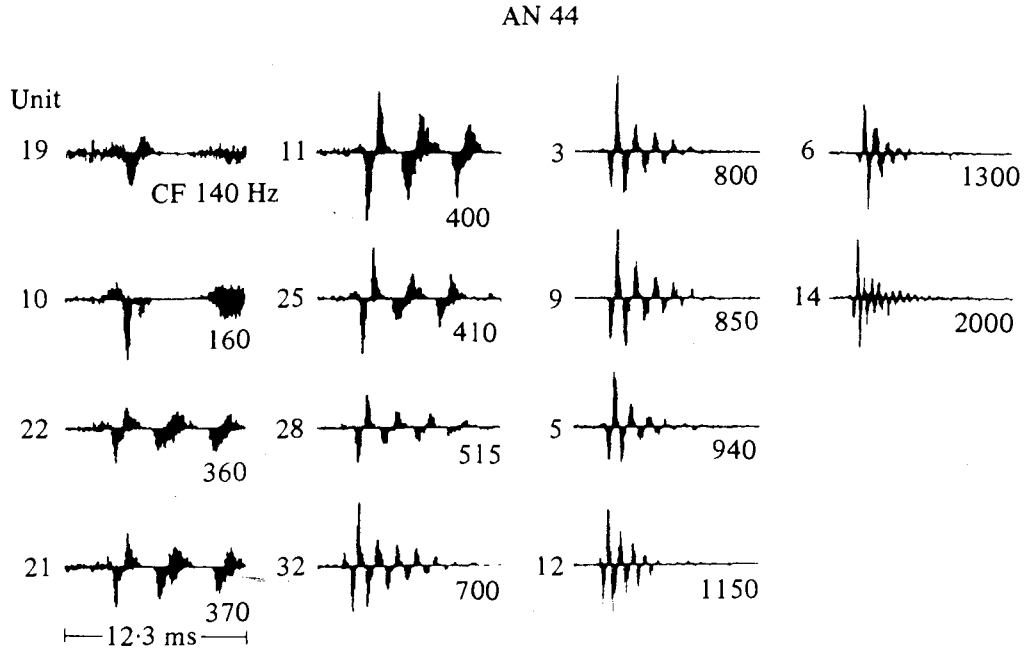


Fig. 7. A collection of compound PST histograms for cat auditory nerve fibres with characteristic frequencies ranging from 150 to 2000 Hz. (From Pfeiffer & Kim, 1972.)

neuron, with a resetting mechanism. This model is visualized in Fig. 8. For our purposes we simplify this model to one that is known as a sandwich model or band-pass non-linearity model. In this type of model the algebraic non-linearity is sandwiched between a bandpass filter representing the tuning properties of the basilar membrane and a low-pass filter representing the synaptic action (e.g. Pfeiffer, 1970; Johannesma, 1971; De Boer, 1976*b*).

It can be shown (Johannesma, 1971) that when $h(\tau)$ is the impulse response of the bandpass filter and $k(\sigma)$ is the impulse response of the lowpass filter, that for an input signal $x(t)$

$$n_0(t) = \int_0^\infty k(\sigma) \int_0^\infty h(\tau)\, x(t-\tau-\sigma)\, \mathrm{d}\tau\, \mathrm{d}\sigma\,. \tag{4.2}$$

This indicates that $n_0(t)$ is linearly related to $x(t)$; if $x(t)$ is a click then $n_0(t)$ is the impulse response of the cascaded linear filter and the low-pass filter. The compound PST histogram therefore does not betray the algebraic non-linearity.

It appears that a logical extension of the methods for the analysis of linear systems to the auditory nervous system is the reverse correlation method using white noise as input signal. For a linear system this method should also yield the impulse response of the neuron. This method has been first applied by De Boer (1967). An

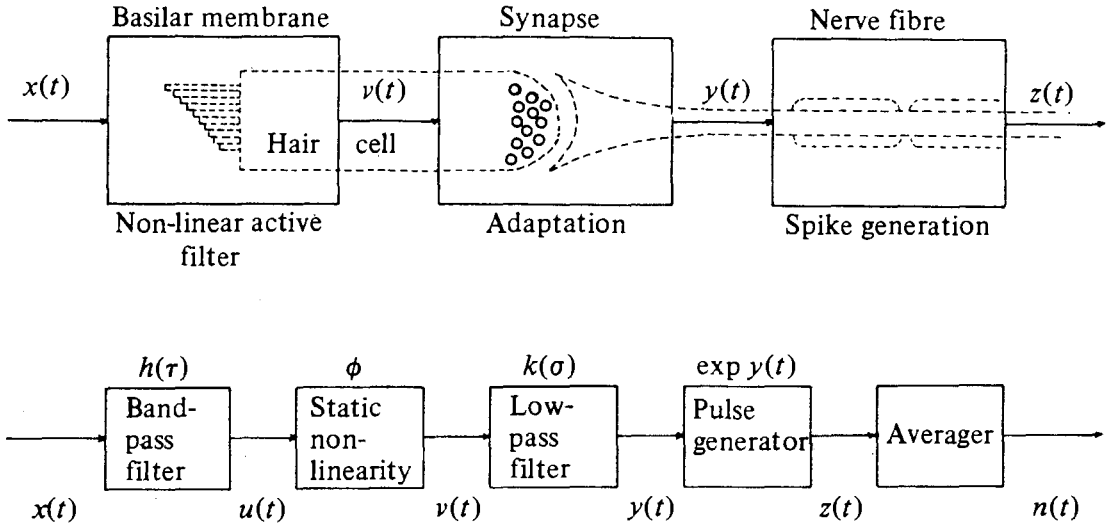


Fig. 8. Model for the peripheral auditory system. The upper part of the figure indicates the current viewpoint: the inner ear consists of a non-linear active filter composed from basilar membrane and hair cells, a synapse incorporating the adaptation properties, and a spike generating element: the nerve fibre. For the purpose of the systems analysis this is modelled into a cascade of simple subsystems as shown in the lower part of the Figure. The non-linear active filter is split up into a sandwich system: a non-linear element sandwiched between two linear filters. The band-pass filter represents the low-intensity tuning of the basilar membrane, the low-pass filter the synaptic action. The spike generating element is modelled as an exponential pulse generator. For comparison with experimental procedures an averager is added.

understanding of this procedure is crucial to the remainder of the paper, thus we will deal therewith in an alternately heuristic and analytic way in the next section.

5. Reverse correlation and its application to auditory nerve fibres

Consider a system consisting of a linear filter and an algebraic non-linear element (e.g. an exponential pulse generator) following it. The input signal is again $x(t)$ and the output of the non-linear element is a series of action potential pulses $z(t) = \Sigma_{n=1}^{N} \delta(t-t_n)$. The input–output correlation function is then equal to

$$R_{xz}(\tau) = \frac{1}{T}\int_0^T x(t)\, z(t+\tau)\, \mathrm{d}t$$

$$= \frac{N}{T}\cdot\frac{1}{N}\sum_{n=1}^{N} x(t_n-\tau), \qquad (5.1)$$

indicating that $R_{xz}(\tau)$ equals the average value of the signal that

precedes the spikes times the neurons average firing rate. Recalling that the compound PST histogram $n_0(t)$ equals

$$n_0(t) = \int_0^\infty l(\tau)\, x(t-\tau)\, \mathrm{d}\tau \tag{5.2}$$

in which

$$l(\tau) = \int_0^\infty k(\sigma)\, h(\tau-\sigma)\, \mathrm{d}\sigma \tag{5.3}$$

can be considered as the impulse compound PST histogram, then, because of (3.9), for white noise as a stimulus

$$l(\tau) = \frac{1}{T}\int_0^T x(t)\, n_0(t+\tau)\, \mathrm{d}t, \tag{5.4}$$

which means that the impulse response of the system can also be found by cross-correlating the compound PST histogram to a white-noise stimulus with the white noise (Johannesma, 1971; Møller, 1977). It can be shown that

$$l(\tau) = 2R_{xz}(\tau). \tag{5.5}$$

Therefore, the following quantities are identical:

(1) the compound PST histogram for click stimulation multiplied by 0·5,

(2) the input–output correlation function for a white-noise input,

(3) the average signal preceding a spike for white noise input times the average neural firing rate during stimulation.

In the general model shown in Fig. 8 the occurrence of action potentials $z(t)$ depends on a signal $y(t)$ which can be considered the generator potential of the auditory neuron. In case the pulse generator gives rise to action potentials at a rate proportional to the generator potential it can be shown (Johannesma, 1980) that

$$R_{xz}(\tau) = R_{xu}(\tau). \tag{5.6}$$

When a pulse generator is modelled as a linear integrator followed by a threshold set at level b for positive-going crossings then it can be shown (De Boer, 1973) that for $x(t)$ is Gaussian

$$R_{xz}(\tau) = bR_{xy}(\tau) - \alpha\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \dot{R}_{xy}(\tau). \tag{5.7}$$

This suggests that there is a correction proportional to the time

derivative of $R_{xy}(\tau)$, the latter being the actual cross-correlation for the BPNL network. For normalization purposes one takes

$$\alpha^2 = \frac{\int R_{xy}^2(\tau)\,d\tau}{\int \dot{R}_{xy}^2(\tau)\,d\tau}. \tag{5.8}$$

This expression (5.7) has been used to test models of the auditory system for their applicability (De Jongh, 1978); it appeared that fixed threshold models are unlikely for the auditory nerve.

A schematic prerequisite for the reverse correlation procedure is shown in Fig. 9. Just as for the compound impulse response, it can be shown that the reverse correlation procedure is insensitive to the non-linearity as long as it is algebraic.

When reverse correlation is performed on-line, high neural firing rates may cause spikes following each other within the time window of the averager to be discarded. This *could* lead to inaccurate estimates of $R_{xz}(\tau)$, Wilson & Evans (1975) argued that the method could lead to impulse responses that show either less damping ('reversed tape-recorder method') or more damping ('delayed noise method') than the theoretically calculated impulse responses. Provided rather trivial precautions are made, no such problems can occur.

We may therefore state that the first-order correlation between input signal and output signal, regardless of whether the output is discrete or continuous (De Boer & Kuijper, 1968), represents the *linear relation* between input and output. There exist two ways to obtain the input–output correlation. If the stimulus is repetitive (e.g. short-duration pseudo-noise sequences, Møller, 1977) one may first form the compound PST histogram and then calculate the cross-correlation, cf. equation (5.4). This is illustrated in Fig. 10. If the stimulus is non-repetitive, as in the case of real noise or very long pseudo-noise sequences, one calculates the average signal before the spike (equation (5.1) and Fig. 9). Note that τ, by definition, runs from right to left (cf. Fig. 9). The first method has the advantage of on-line calculation of the compound PST histogram followed by a short off-line correlation procedure; a drawback can be the inherently poorer statistics of the short pseudo-noise sequences. This, however, seems not to affect the results for the first-order cross-correlation but affects the higher-order cross-correlations (cf. section 17).

The Fourier transform of the reverse correlation function and of the click-and-noise compound PST histogram (Fig. 11) will be equal to the tuning curve provided that the system behaves linearly. This aspect has been explored, especially by De Boer for the cats auditory

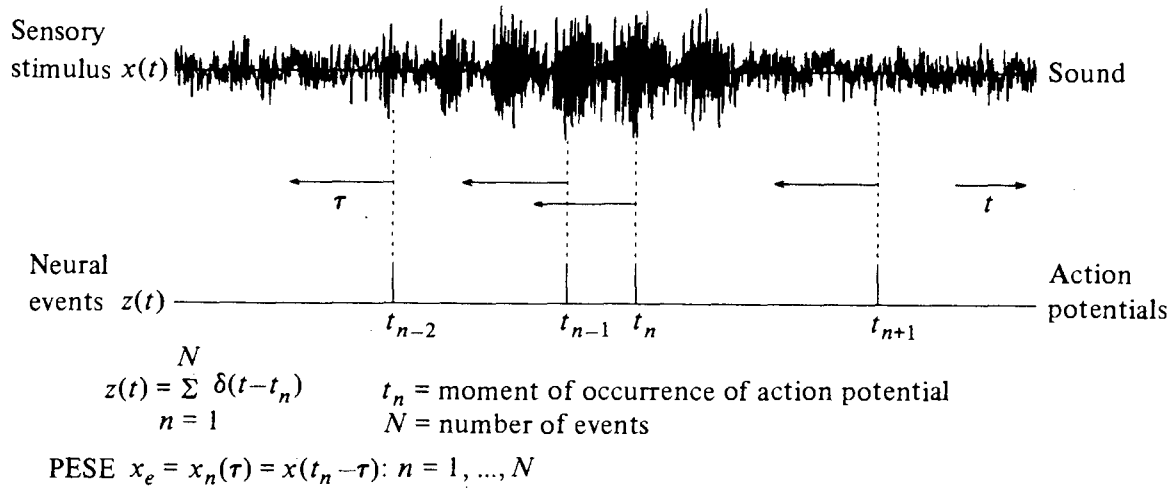


Fig. 9. The operational definition of the pre-event stimulus ensemble, PESE (From Aertsen *et al.* 1980.)

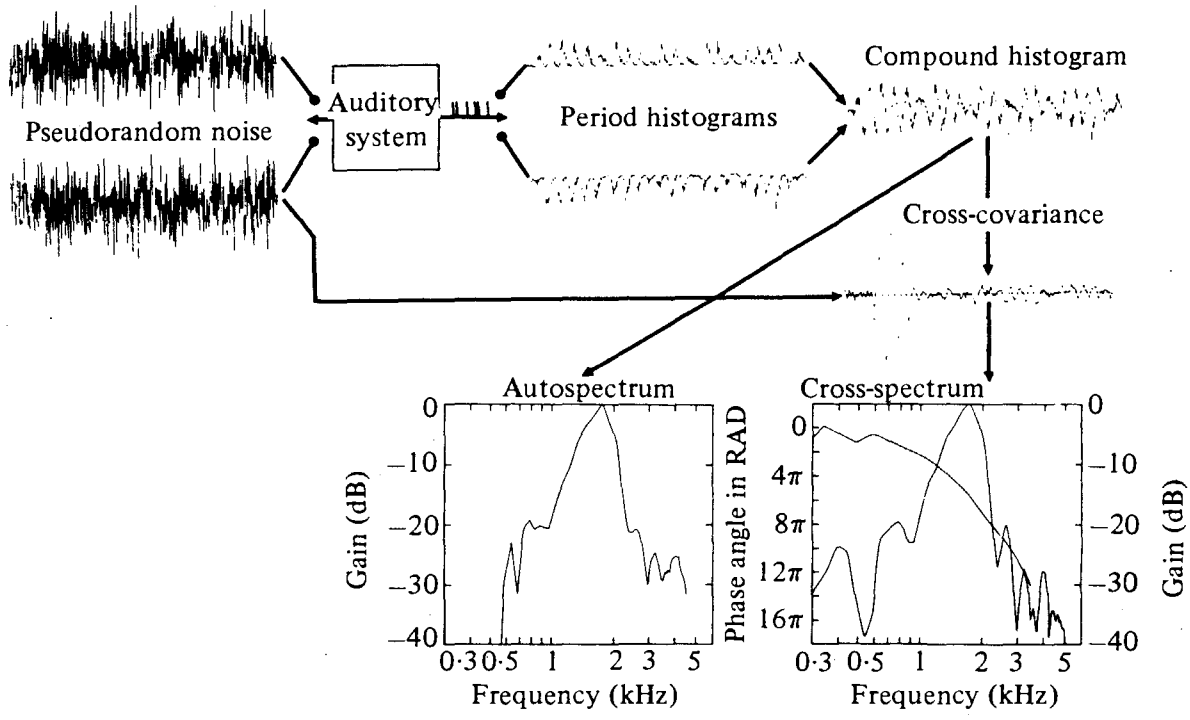


Fig. 10. Schematic illustration of the data processing to obtain cross-correlation functions in case short pseudo-random noise sequences are used. A compound histogram is formed from which the autospectrum, in case the stimulus is sufficiently white, directly gives the modulus of transfer function. From the cross-correlation the cross-spectrum can be obtained by Fourier transformation. This results in amplitude as well as phase information. (From Møller, 1977).

nerve (De Boer, 1967, 1968, 1969, 1973; De Boer and De Jongh, 1978) and also by Evans (1977), and by Harrison & Evans (1982) for the guinea pig.

De Boer (1969) observed some discrepancy in this comparison: the actual tuning curve was slightly narrower and somewhat steeper than

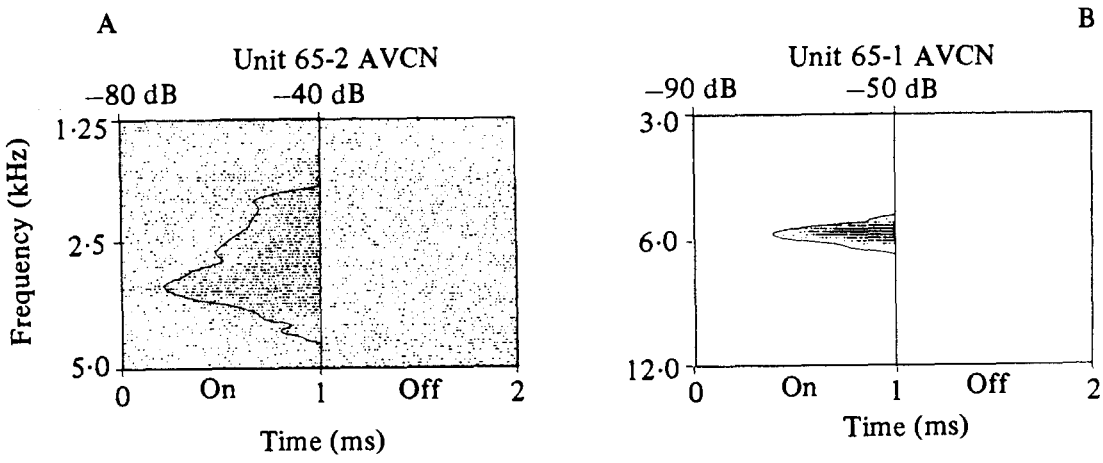


Fig. 11. Filtering characteristics of cochlear nucleus units in the cat as determined by reverse correlation (——) and compared to frequency–response areas (.....). (From Grashuis, 1974.)

the results for the stimulation with noise. This was attributed to the difference in stimulation: highly frequency-specific stimuli to obtain the tuning curve and a broadband stimulus to obtain the reverse correlation function.

For the model shown in Fig. 8, where ϕ represents the non-linearity, and in case the neural pulse generator is a linear one, it is found that

$$n_0(t) = \int_0^\infty k(\sigma)\,\phi\left\{\int_0^\sigma h(\tau)\,x(t-\sigma-\tau)\,\mathrm{d}\tau\right\}\mathrm{d}\sigma. \tag{5.9}$$

If we express the non-linearity as

$$\phi(u) = u + \psi(u), \tag{5.10}$$

where $\psi(u) = \psi(-u)$, then the Fourier transform of the reverse correlation function $n_0(t)$ for $x(t)$, being Gaussian white noise with power P, is

$$\hat{n}_0(\omega) = P\hat{k}(\omega)\,\hat{h}(\omega). \tag{5.11}$$

Obtaining the tuning curve requires that $x(t) = A\cos\omega t$ and it can be shown that if

$$\phi(u) = u + |u|$$

one obtains for the tuning curve

$$T(\omega) = \tfrac{1}{2}A^2\hat{k}(0)\,|\hat{h}(\omega)|. \tag{5.12}$$

By comparing the Fourier transform of the reverse correlation function $\hat{n}_0(\omega)$ with the tuning curve $T(\omega)$ we observe that only if ω is small compared to the cut-off frequency of the low-pass filter, $\hat{k}(\omega)$, both estimates turn out to be about equal. If ω is much larger than that cut-off frequency then no reverse correlation is obtained; this

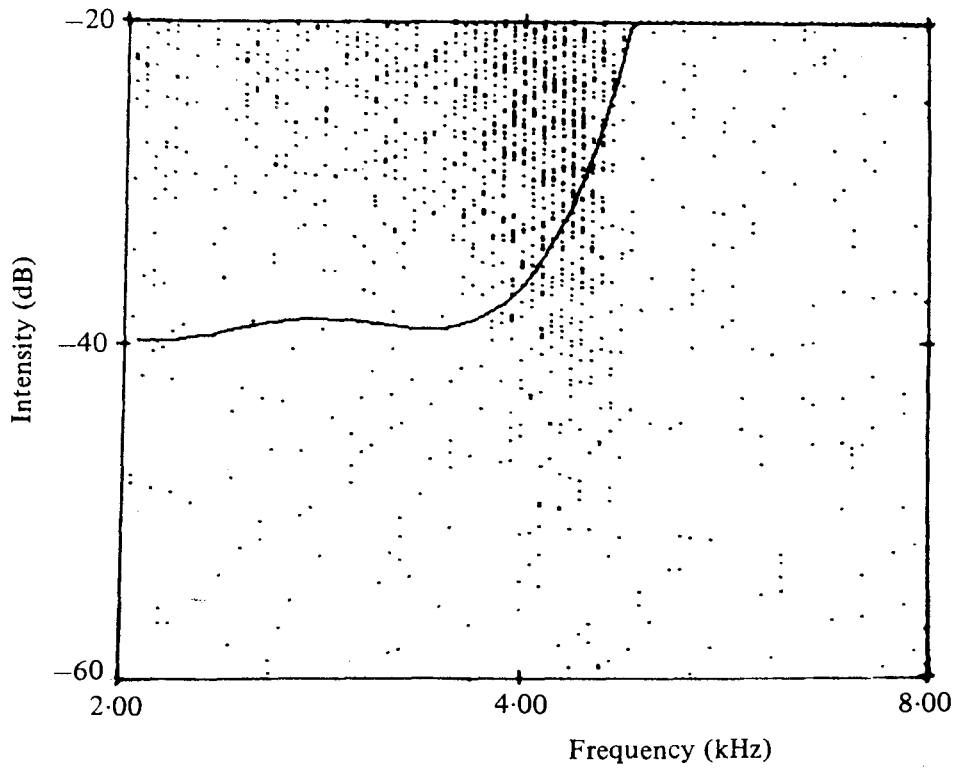


Fig. 12. Response area as determined with tonal stimuli and compared to the filter characteristic derived from reverse correlation (drawn line) for a unit from the VCN in the cat. Although the characteristic frequency of the unit is only slightly above 4 kHz, the reverse correlation function seems only to represent the low-pass filter (cf. Fig. 8). (From Grashuis, 1974.)

is equivalent to saying that the phase lock of the nerve fibre's discharges has disappeared. An example of such a case is shown in Fig. 12. It must be remarked that basically the Fourier transform of the reverse correlation function should not be compared to tuning curves but to iso-intensity contours (Møller, 1977, 1978). De Boer observed less or no effect upon the reverse correlation function when the noise intensity was changed, this was in fact confirmed by Evans (1977) also for the cats auditory nerve. In contrast, Møller (1977, 1978) found for the rat's auditory nerve a quite dramatic change in the shape of the reverse correlation function as a result of a 40 dB change in the noise level as shown in Fig. 13. Comparing the cross-spectra (equation 3.10) Møller (1978) found that they widened considerably for higher noise levels, the low-intensity noise results were quite well comparable to a frequency-tuning curve (Fig. 14). A finding concurrent with the widening of the cross-spectra for higher noise levels is a change in best frequency (BF) toward lower frequencies.†

† We use characteristic frequency (CF) for the frequency at which the neuron has the lowest threshold, and best frequency (BF) for the frequency at which the firing rate is highest in case of super-threshold stimulation.

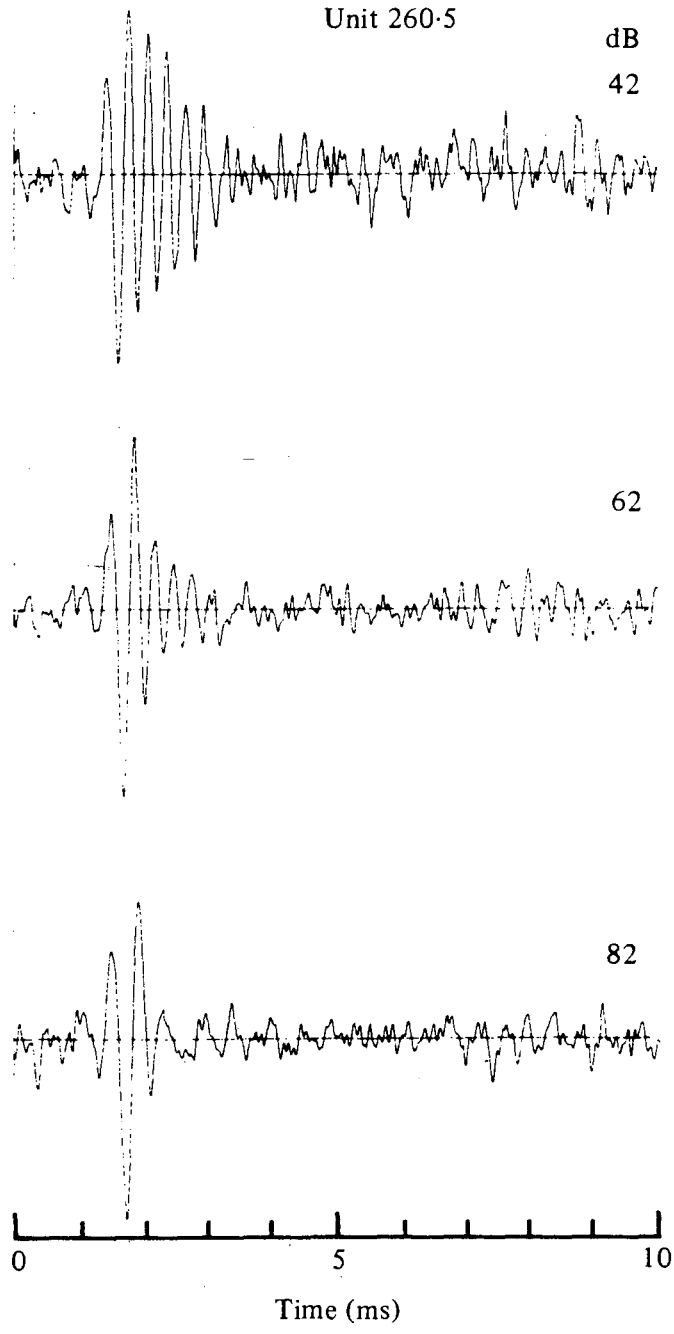


Fig. 13. Strong intensity dependence of reverse correlation functions for an auditory nerve fibre in the rat. (From Møller, 1977.)

Harrison & Evans (1982) compared tuning curves in normal and pathological guinea-pig cochleas with the cross-spectrum or transfer function derived from the reverse correlation technique. When using reverse correlation functions obtained under noise stimulation near threshold they obtained a good correspondence between $Q_{10\,\mathrm{dB}}$ values in the normal as well as in the pathological cochlea. However, they observed quite a strong change in the reverse correlation functions for increasing noise levels. The impulse response became much more damped as in the rat data shown by Møller.

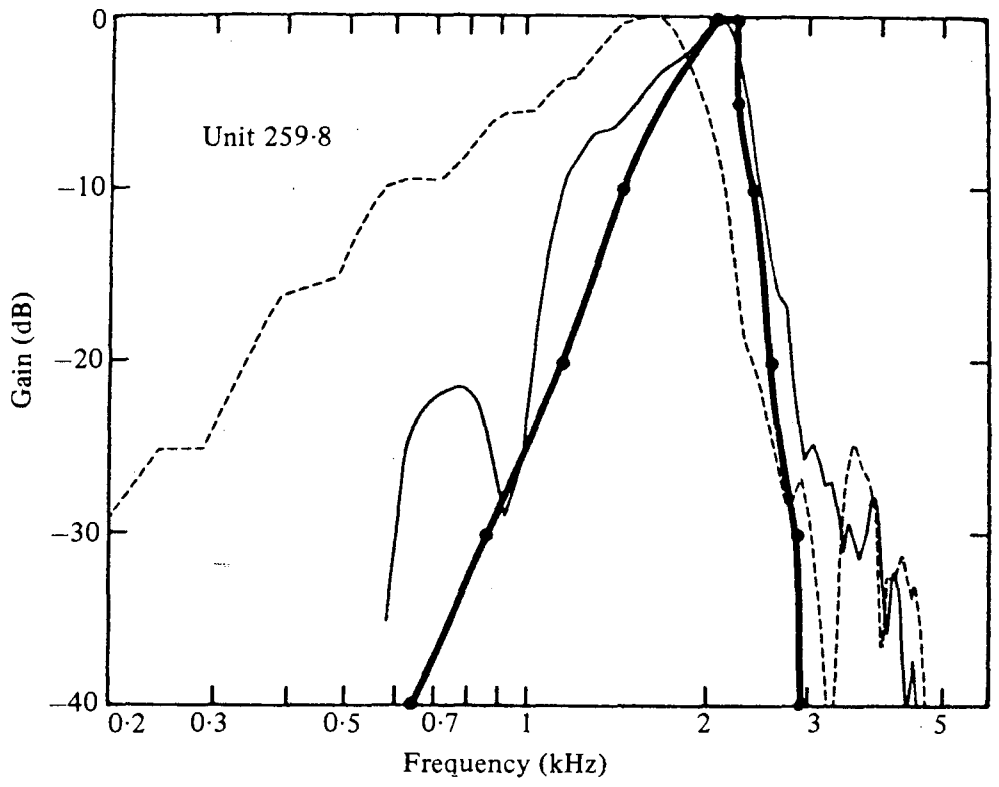


Fig. 14. Transfer functions for single auditory nerve fibres in the rat as derived from reverse correlation functions compared to a frequency-tuning curve obtained with tones (thick line). The thin line shows results for a 15 dB above threshold noise stimulus. (From Møller, 1978.)

Auditory nerve fibres with CF's above 4 kHz do not show multiple peaks in click PST histograms (Kiang *et al.* 1965), and procedures resulting in period histograms (Rose *et al.* 1967) indicate the same. One may therefore expect (see also equation 5.11) that the loss of phaselock will abolish the formation of useful reverse correlation functions in these cases. This has been found indeed; the highest CF fibre for which a reverse correlation procedure gave good results was reported by De Boer & De Jongh (1978) at 4·55 kHz for the cat.

Fengler (1980), who studied reverse correlograms for primary auditory nerve fibres in the caiman, found clear results in the frequency range of 195–1680 Hz at an inner ear temperature (and body temperature) of 27 °C. He observed a clear effect of raising the noise level for which the reverse correlograms were obtained. This consisted in a larger damping, a downward shift of BF and some decrease in latency. In three neurons he was able to study the effect of a rise in body temperature: it was observed that the BF increased and that the damping increased. The increase in BF amounted to 3–8 % per °C.

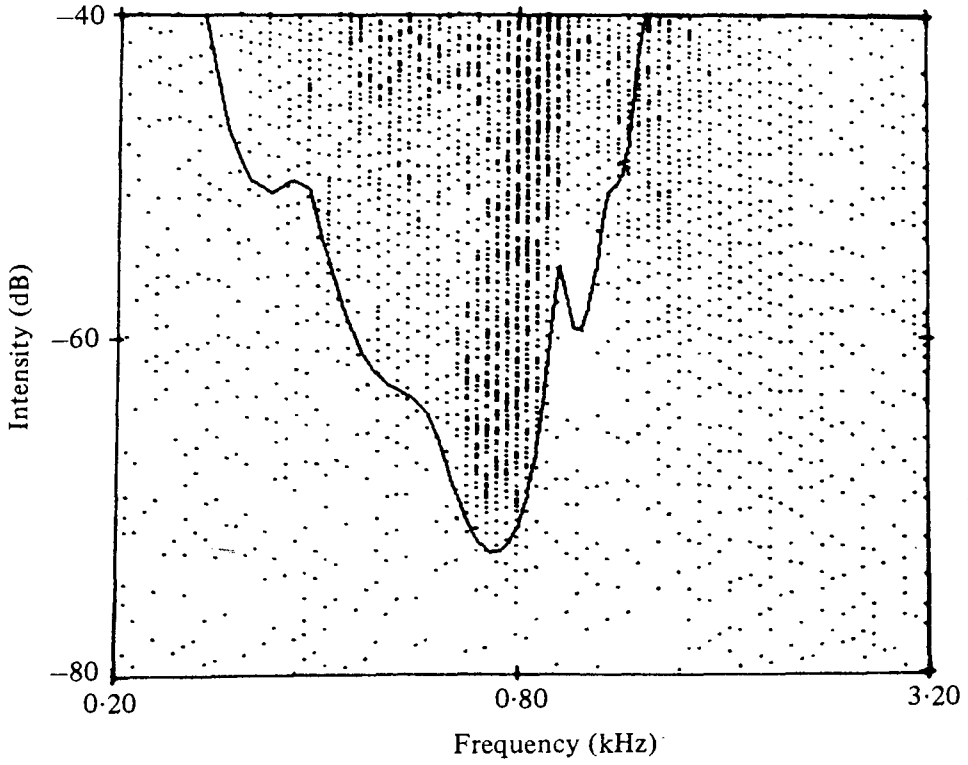


Fig. 15. Comparison of the response area as obtained with tonal stimuli compared to filter characteristics as derived from reverse correlation. Interestingly, the filter characteristic seems to be more simple than the complex, non-monotonic response area in which suppression effects can also be seen. (From Grashuis, 1974.)

6. Reverse correlation for higher auditory centres

Neurons in the cochlear nucleus complex and in higher auditory centres are inherently more complex in their response characteristics than auditory nerve fibres. One may wonder what relation there is in these cases between an estimate of the linear behaviour regarding the tuning on basis of broadband stimulation (clicks and noise) and on basis of simple tonal stimuli. Johannesma *et al.* (1973) and Van Gisbergen *et al.* (1975 *a*) compared cochlear nucleus responses to tone and noise stimuli. Reverse correlation functions could be obtained from neurons with low CF, which in addition had a primary-like temporal pattern and only showed an activation response to tones. Van Gisbergen (1974) and Grashuis (1974) showed that the highest CF for which a reverse correlation could be obtained was close to 5 kHz for a unit from the VCN. In one noteworthy example (Grashuis, 1974), however, it was found that although the neurons CF was only slightly above 4 kHz, the reverse correlation function indicated a BF of 2·7 kHz. This was probably caused by the action of the low-pass filter (cf. Figure 12). Dorsal cochlear nucleus units did yield a reverse correlation but the highest CF in that case was

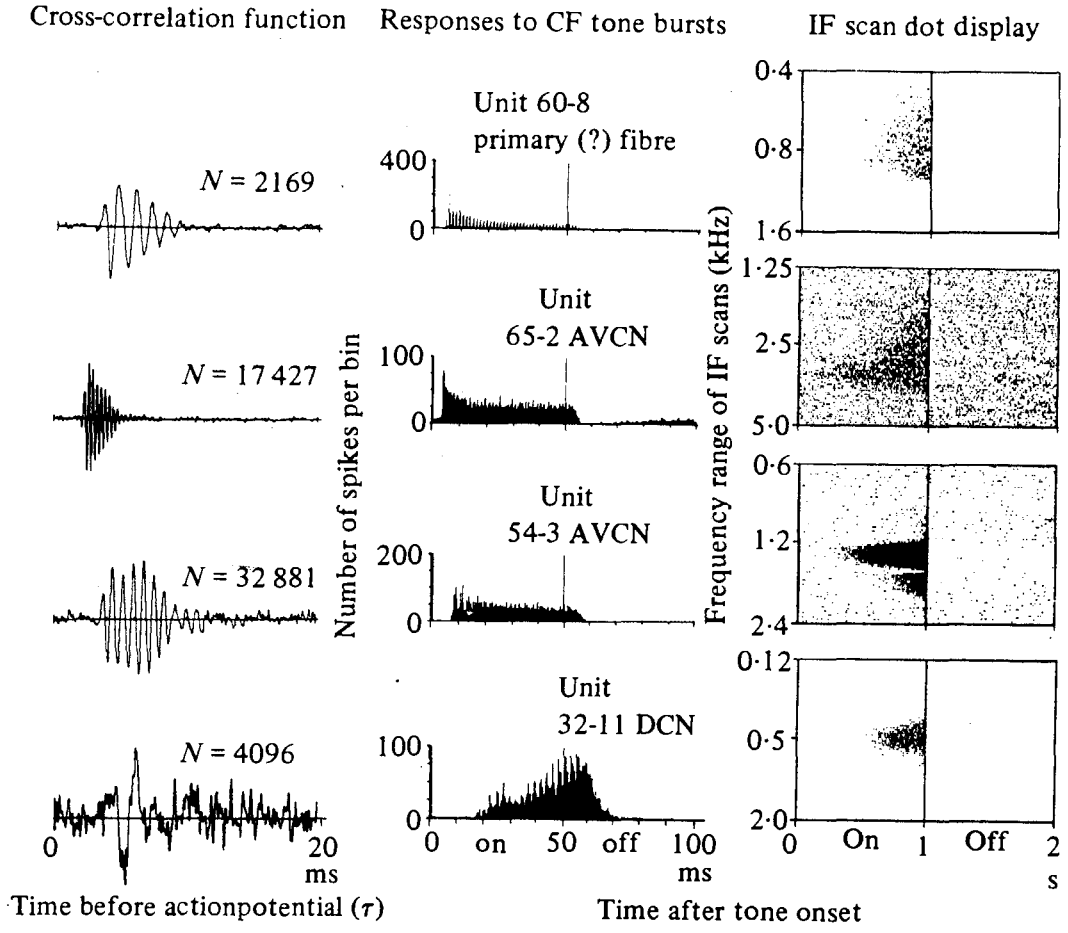


Fig. 16. Cross-correlation functions obtained from units in the cochlear nucleus with different temporal patterns of response to CF tone bursts. No obvious relation seems to exist between the cross-correlation function and the temporal pattern of response to tone bursts. The response areas (right-hand column) are shown for comparison. *N* denotes the number of pre-spike stimuli which have been averaged. (From Van Gisbergen, 1974.)

below 2 kHz. It was generally found that the reverse correlation function strongly reflected the frequency selectivity of the neuron as revealed by its response area (or frequency tuning curve) to pure tones. When, however, inhibitory effects could be seen in the response area (Fig. 15) for spontaneously active neurons the reverse correlation did not show these effects and comparison of tuning curve and Fourier transformed reverse correlation function seems no longer possible.

A great variety of temporal response patterns is found in the cochlear nucleus complex. One may distinguish the primary-like pattern, the chopper, build-up and complex patterns. Van Gisbergen *et al.* (1975*a*, *b*) could, however, not demonstrate any relation between the reverse correlation function and the PST histogram to tonebursts. It was concluded that the reverse correlation function solely reflects the frequency selectivity of the neuron (Fig. 16).

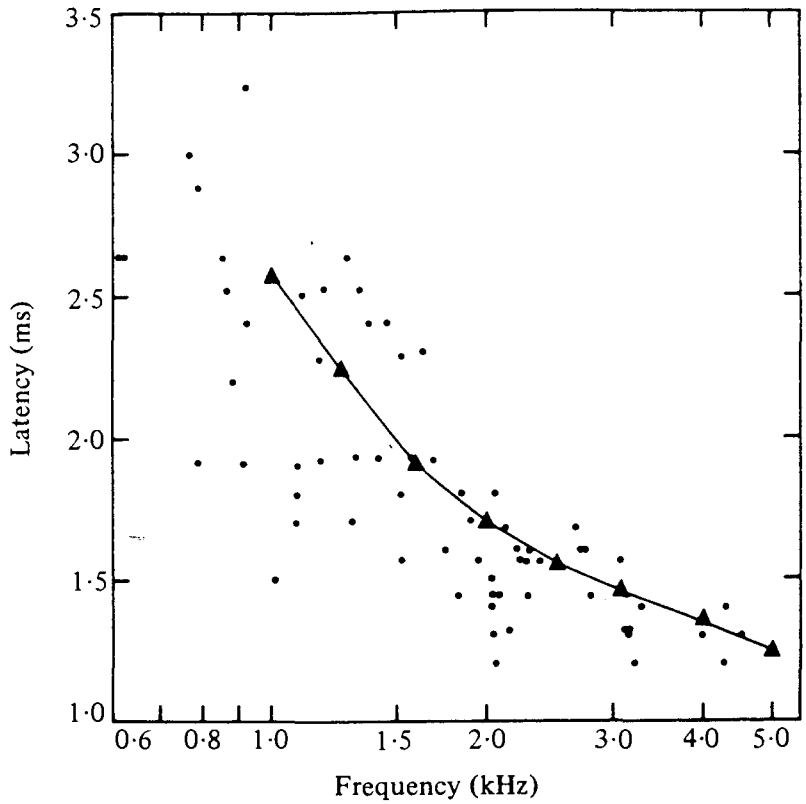


Fig. 17. Comparison of latency from the cross-correlation function and the N_1-latency to one-third-octave filtered clicks (△–△). (From Møller, 1977.)

A further parameter of interest is the response latency. When we consider the waveform of the reverse correlation function it is noted that there is apparently a delay before the oscillation starts to built up. This can be interpreted as a pure time delay composed mainly of acoustic delay, travelling wave delay, synaptic delay(s) and conduction time along the auditory nerve fibre (Van Gisbergen, 1975*a*; De Boer, 1976*b*; Møller, 1977). It appears that this pure time delay plotted against CF decreases toward high CF, probably mainly due to a decrease in travelling wave delay (Fig. 17). Møller (1975) and Van Gisbergen (1975*b*) have directly compared latency for cochlear nucleus neurons to tonebursts and noise stimuli. This resulted in large discrepancies, especially for units in the DCN, which showed activation as well as suppression in the response area or had complex temporal patterns. This discrepancy was attributed to temporal integration. On the other hand, units which showed activation-only to tone bursts had comparable latencies to both types of stimulation. Recently Van Heusden & Smoorenburg (1983) compared frequency selectivity as measured with tones, clicks and noise stimuli for AVCN units before and after inducing acoustic trauma. A 30 min exposure to 105 dB SPL pink noise resulted in average threshold shifts of

30 dB. Response areas and turning curves were measured for sinusoids, compound PST histograms were computed for click stimuli and reverse correlation functions were obtained for noise. A comparison was made between the weighted average group delay to sinusoidal stimuli and the centre of gravity of the reverse correlation function or the compound PST-histogram. A high correlation was found and it was concluded that 'the cochlear latency data suggest that the cochlear filter behaves like a linear filter'. About the relationship between the three measures of frequency selectivity the authors remark that 'from a stimulus point of view, measurements of response areas are closer to measurements with click or noise stimuli since all frequency components of click and noise stimuli are at equal level'. A problem, however, is that response areas are independent of the amount of phase locking. Therefore Van Heusden and Smoorenburg suggest to compare the response spectra of clicks and noise with response areas weighted according to the amount of phaselock at the contributing frequencies. In eight units a direct comparison could be made for the pre- and post-exposure frequency selectivity. It appeared that noise exposure had no effect on the $Q_{10\ \mathrm{dB}}$ derived from the reverse correlation function while the $Q_{10\ \mathrm{dB}}$ for the frequency tuning curve obtained with sinusoids invariably showed a decrease (up to 40 %). This was explained by assuming that the noise had a linearizing effect on the pattern of discharges; a similar phenomenon was found for click stimuli.

Frequency selectivity is mainly a reflexion of a stationary state of the neuron, while latency apart from this also involves dynamic properties. Stimulation with continuous wide band noise will bring the neuron quite rapidly in such a stationary state. De Boer & De Jongh (1978) estimate that this takes only a few milliseconds for the cat's auditory nerve. To study the dynamical aspects of neurons one should therefore prevent the formation of a stationary situation. Noise modulation of carrier tones is one of the methods employed which can also be investigated using correlation techniques.

7. Envelope correlation: the dynamical properties of neurons

Instead of correlating the compound PST-histogram for a short sequence of pseudo-noise with that noise to obtain a measure for the frequency selectivity of the neuron (cf. Fig. 10) one now correlates a compound PST-histogram with the pseudo-noise modulator of a

tone at the CF. The resulting transfer function (after Fourier transformation) now describes the dynamic properties of the system at mean stimulus level at a particular frequency. This modulation transfer function may be compared to sine-modulation transfer functions (Møller, 1973) and it appears that they result in the same gain function.

These gain functions or their corresponding impulse responses or step-responses (Møller, 1976*a*) as obtained for the rat cochlear nucleus could be classified into two types. In the first the step-response decreased only very slowly with time and the impulse response therefore was highly damped with a corresponding low-pass characteristic for the transfer function. This was interpreted as to represent a low degree of adaptation. The second type appeared to show a large degree of adaptation and had a damped oscillation in the impulse response with a corresponding bandpass transfer function peaking between 100 and 200 Hz. It was found that units could show a low degree of adaptation at, for example, 20 dB SPL and a very high degree of adaptation at 60 dB SPL, i.e. the modulation impulse response changes from highly damped to strong oscillatory with increasing carrier intensity. When instead of the CF-tone carrier a noise modulated tone frequency in the inhibitory region was selected, while the CF-tone was still present, the modulation impulse response was, except from a 180° phase change, nearly the same. This indicates that adaptation for inhibition and excitation follows largely the same time course (Møller, 1976*a*): the form of the modulation impulse response does not depend on the carrier frequency, only the sign and size are carrier-frequency-dependent.

This last finding points to a certain independence between the filter properties $\hat{l}(\omega)$ and dynamic amplitude coding characteristics of auditory neurons and cochlear nucleus units: selectivity *v.* sensitivity.

In the case of a tonal stimulus, carrier frequency ω, which is harmonically amplitude-modulated, modulator frequency ν, we may write for the neuron's response y:

$$y_I(\omega,\nu) = g_I(\omega,\nu)\, A[\hat{h}(\omega)\,\hat{x}(\omega)]\,(\nu), \qquad (7.1)$$

where A is an operator, determining the instantaneous amplitude of the $\hat{h}$-filtered input signal (comparable to an 'envelope detector'), g is a modulation-transfer function for carrier frequency ω and modulation frequency ν. h is the spectral sensitivity of the neuron.

The index I has been used to indicate the possibility that g, and

thus y, depend on the mean intensity level I of the stimulus. In case $x(t)$ is a wideband signal, equation (7.1) can be generalized

$$y_I(\omega, \nu) = g_I(\omega, \nu)\, A[\hat{h}\hat{x}]\,(\omega, \nu), \tag{7.2}$$

where $A(\omega, \nu)$ indicates the spectrum of the amplitude modulator (ν-axis) of the ω-component in the $\hat{h}$-filtered signal input. This wide-band generalization bears a relation to the concept of the dynamic spectrum, to be discussed more thoroughly in Section 9. The term $y(\omega, \nu)$ denotes the contribution of the particular (ω, ν) combination to the final response function $y(t)$ of the neuron. The latter can be obtained by integrating over the various frequencies ω to yield the spectrum $\hat{y}(\nu)$ of the neural response $y(t)$.

Following (7.2) one may obtain information regarding the dynamics of $g(\omega, \nu)$ by studying the neuron's firing rate as a function of the dynamic amplitude behaviour of the stimulus x at various carrier frequencies ω. As Møller (1976*a*) showed, the dependence of the modulation transfer function $g_I(\omega, \nu)$ on the carrier frequency (related to the tuning) is only very small. Van Gisbergen *et al.* (1975*b*) came to the same conclusions for cochlear nucleus neurons in the cat. Therefore in good approximation we may write

$$g_I(\omega, \nu) = a(\omega) g_I(\nu). \tag{7.3}$$

Experimental results show that, especially for auditory nerve fibres and type I cochlear neurons (Møller, 1976*a*) $g_I(\nu)$ is a low-pass function. For a certain class of cochlear nucleus neurons, $\hat{g}(\nu, I)$ is peaked around a frequency ν between 100 and 200 Hz. Bibikov & Gorodetskaya (1981) found peaks around 40 Hz for neurons in the torus semicircularis of the grassfrog, i.e. close to the natural repetition frequency in male grassfrog vocalizations. The present, somewhat heuristic, discussion on the separability of frequency selectivity and temporal sensitivity will be discussed in a more quantitative and formalized context in Section 14.

8. Reverse correlation and phase lock

The auditory system is highly non-linear, i.e. characterization of the system will depend on stimulus level as well as on the type of stimulus. Therefore the three different ways we described in Section 3 for characterizing a linear system will now differ when applied to the higher parts of the auditory system. We have seen that under the

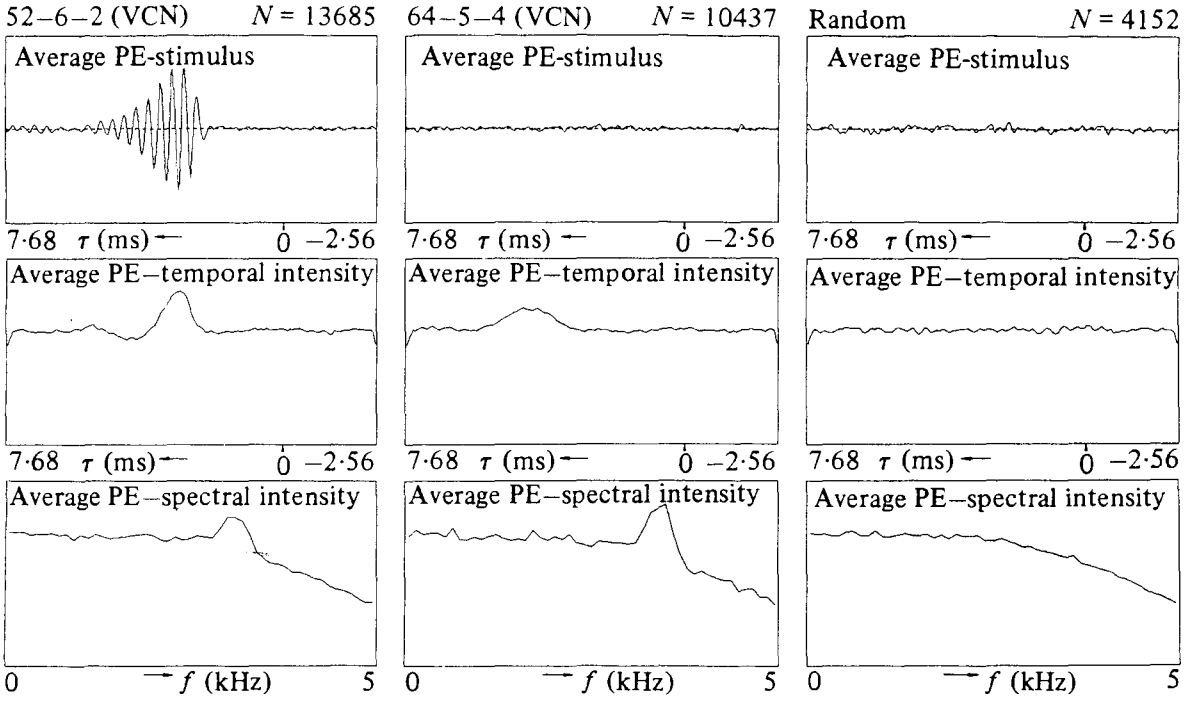


Fig. 18. The effect of phase-lock upon the reverse correlation function. Shown are results from two neurons from the VCN in the cat with nearly the same characteristic frequency: about 3 kHz. One of the units, left column, shows a clear reverse correlation, the other represented in the middle column shows no sign of a reverse correlation. For comparison the third column shows the results for spontaneous activity. In the second row the average temporal intensity of the noise preceding the spikes is shown, both units show a peak superimposed upon the average noise level (cf. the third column) at about 3 resp. 5 ms before the spike. In the lower row the average spectral intensity of the noise preceding the spike is shown: again both neurons show a clear peak around 3 kHz superimposed on the noise spectrum. The lack of phaselock may abolish the reverse correlation function but response properties can still be obtained using intensity averaging (N denotes the number of spikes). (From Johannesma and Aertsen, in preparation.)

assumption of an algebraic, i.e. static non-linearity, the impulse response of the linear part of the system is obtained when using the reverse correlation technique. Linear systems are the first approximation to non-linear systems, an approximation which in many cases is satisfactory. One of the obvious situations in which the reverse correlation as well as the compound PST-histogram method fails is for characteristic frequencies above 4 kHz, where the phase lock disappears.

For mammals this phase lock strongly diminishes (at the level of the auditory nerve) above 3 kHz and is virtually absent above 5 kHz (Rose *et al.* 1967). In reptiles studied at a temperature of 28 °C Klinke & Pause (1980) and Fengler (1980) found reverse correlation functions up to 1700 Hz. For amphibians Narins (1983) found phase lock to

disappear around 900–1000 Hz at 22–24 °C in the auditory nerve. Hermes *et al.* (1981) observed reverse correlation in the torus semicircularis of the grassfrog studied at 15 °C only below 350 Hz. Temperature is therefore an important parameter in setting the upper limit of phase lock. Synaptic mechanisms with their high temperature dependence may play an important role here. Higher up in the central nervous system additional limiting factors may show up by the cascading of synapses (to be considered as low-pass filters) by leaky integrator properties of neurons (e.g. built-up neurons in the dorsal cochlear nucleus (Van Gisbergen *et al.* 1975*a*, *b*) or through neural interaction.

An alternative method to reveal the power-spectrum of the filter characteristics of the neuron for wide-band stimulation is by computing for each part of the signal preceding a spike (the pre-event stimulus ensemble, PESE) the power spectrum and subsequently averaging these power spectra. A difference between the average PESE power spectrum and the power spectrum of the wide-band stimulus ensemble (SE) indicates a stimulus response relation. Fig. 18 shows two examples from neurons in the ventral cochlear nucleus of the cat (Johannesma & Aertsen, to be published). One unit shows a reverse correlation function for which also the average temporal intensity and the average spectral intensity of the PESE are shown. The right-hand column shows the overall stimulus ensemble using random triggers. The middle column shows results for a unit where no reverse correlation could be obtained, but the average temporal intensity as well as the average spectral intensity differ from those for random triggers. This method therefore can give both the latency (on basis of the temporal intensity) and the best frequency (on basis of the spectral intensity).

It may happen, however, that there is an appreciable amount of post-activation suppression. This will cause the average PESE power spectrum to contain less energy in a certain frequency range than the SE in the time window of the suppression. In another window more close to the time of occurrence of the event, we would find on average more energy. Averaging over a large time-window, then, may result in no apparent difference between the power spectra of the PESE and the SE (Hermes *et al.* 1981). A sign of this phenomenon is also present in the left-hand column of Fig. 18.

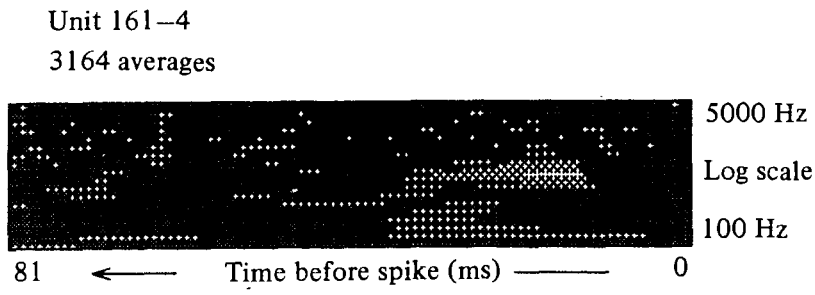


Fig. 19. Spectro-temporal receptive field of a unit from the torus semicircularis in the grassfrog determined for a gaussian wide-band stimulus ensemble by averaging the pre-event dynamic spectra and subtraction of the *a priori* expected stimulus spectrogram. Positive regions are displayed darker than background, negative regions are lighter than background. (From Hermes *et al.* 1981.)

9. The dynamic spectrum and dynamic spectrum analyser

Traditionally sound was studied either as a function of time or as a function of frequency. The first method requires sharply defined instants of time, the second method infinitely long tones of rigorously defined frequency. In everyday life neither sharply defined sounds nor infinitely long tones are abundant, and sound such as speech insist on a description in terms of both frequency and time. This has led to the invention of the 'Sound Spectrograph' (Potter, Kopp & Green, 1947) which produced visible speech. This is a short-time spectro-temporal picture of sound. It has been widely used in (neuro-)ethology for representation of bird songs and frog vocalizations.

When adopting the subject-centred method of investigation to the auditory system, spikes loose their explicit meaning, they only indicate that some sound caused the neuron to respond. The properties of those sounds therefore tell something about the neuron and deserve further study.

This approach, combined with the spectrogram analysis of sound among others, deals with the earlier-described pitfalls of post-activation suppression: the construction of the dynamic spectrum or short-time power spectrum of the stimuli preceding the action potentials (Aertsen & Johannesma, 1980). In this case the spectrum is computed as a function of time before the spikes and as such represents suppression and activation separately, thereby resulting in a more complete picture of the relevant stimulus properties. This dynamic spectrum of the PESE can be seen as the sonogram that the neuron selects from the wide-band stimulus. A drawback of a dynamic spectrum approach is the *a priori* setting of the various filter bandwidths. Commonly used for auditory research is a bandwidth of

one third of an octave. Still the dynamic power spectrum of the PESE reveals the basic advantage of this type of second-order reverse correlation: it provides us with an estimate of both the neurons spectral properties (frequency tuning) and temporal properties (latency, periodicity, short-term adaptation) in one analysis.

Extending the concept of first-order reverse correlation functions or equivalently the average values of the pre-event stimulus ensemble as given by

$$R_e(\tau) = \frac{1}{N} \sum_{n=1}^{N} x_n(\tau) \tag{9.1}$$

to the second moment function or time dependent autocorrelation function of the PESE

$$R_e(\sigma, \tau) = \frac{1}{N} \sum_{n=1}^{N} x_n(\sigma)\, x_n(\tau) \tag{9.2}$$

and defining the dynamic power spectrum $P(\omega, \tau)$ of the stimulus we may calculate the average pre-event dynamic power spectrum

$$P_e(\omega, \tau) = \frac{1}{N} \sum_{n=1}^{N} P_n(\omega, \tau), \tag{9.3}$$

where the subscript n refers to the nth pre-event stimulus $x_n(\tau)$; $R_e(\sigma, \tau)$ and $P_e(\omega, \tau)$ are related by Fourier transformation (Aertsen *et al.* 1981). An example of the average pre-event dynamic power spectrum for gaussian noise as a stimulus is given for a neuron from the torus semicircularis in the grassfrog in Fig. 19. A grey coding is used, dark denotes more than average power and light areas indicate less than average power. This is interpreted as dark areas corresponding to activation, i.e. frequencies within these regions activate the neuron, and light areas as suppression effects. In this case it may be interpreted as post-activation suppression (Hermes *et al.* 1981). The entire constellation of activation and suppression areas has tentatively been called 'Spectro-temporal receptive field' (Aertsen, Johannesma & Hermes, 1980). When we want to study the average properties of sound preceding an action potential the crucial point will be how accurate the sound can be measured. Sonograms are usually made by passing the sound through a bank of one-third-octave filters or fixed bandwidth filters and representing the intensity of the filter output (e.g. Aertsen & Johannesma, 1980). Taking the filter bandwidth as $\Delta\omega$ then the time resolution of the method will be bounded by Gabor's (1946) inequality

$$\Delta\omega \,.\, \Delta t \geqslant \tfrac{1}{2}. \tag{9.4}$$

For the dynamic spectrum analyser with one-third-octave bandwidth Aertsen and Johannesma calculated the uncertainty product to be 0·78 for each of the filters used. Thus if auditory neurons were better analysers in terms of the uncertainty product, i.e. had a $\Delta w . \Delta t$ more close to half, this spectral analyser would result in tuning curves that appear too broad and/or had temporal uncertainties (e.g. in latency) that are slightly exaggerated. In general, however, neurons in the central auditory system of the frog showed receptive field sizes ($\Delta\omega . \Delta t$) of more than 5, in that case justifying the use of such a type of spectrum analyser.

A more serious drawback of a physical, proportional band-width system is that the filter delay depends on centre frequency. The group delay μ_t is then inversely proportional to its centre frequency μ_f. This leads to a distorted picture especially when short time-windows (of the order of 100 ms) and low centre-frequency filters are used.

A method is available, however, to construct a mathematical expression which contains all information about the sound and from which the output of any spectrum analyser – regardless of the bandwidth of the filters – can be calculated. This has been termed the complex energy density (Rihaczek, 1968) or coherent spectro-temporal intensity density (CoSTID) by Johannesma (1972) and Johannesma *et al.* (1981). It is a complex valued function on the combined frequency– time domain.

10. The coherent spectro-temporal intensity density function

Any periodic signal can be expressed as a Fourier series, a sum of cosine and sine terms

$$x(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t) \qquad (10.1)$$

or alternatively by a complex time series

$$\xi(t) = \sum_{n=0}^{\infty} (a_n - ib_n)\, e^{in\omega t}, \qquad (10.2)$$

which can be formed by adding to the real signal $x(t)$ an imaginary signal $i\tilde{x}(t)$, where $\tilde{x}(t)$ is the signal in quadrature to $x(t)$:

$$\tilde{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} ds \frac{x(s)}{t-s}, \qquad (10.3)$$

where $\int_{-\infty}^{\infty}$ is the Cauchy principal value integral. $x(t)$ and $\tilde{x}(t)$ are a pair of Hilbert transforms and $\xi(t)$ is called the analytic signal:

$$\xi(t) = x(t) + i\,\tilde{x}(t). \tag{10.4}$$

For example if $x(t) = a \cos \omega t + b \sin \omega t$ then $\xi(t) = (a - ib)\,e^{i\omega t}$. The spectrum $\hat{\xi}(\omega)$ of $\xi(t)$ equals zero for negative frequencies and equals twice $\hat{x}(\omega)$ for positive frequencies.

If we now form the complex function

$$\Xi\,(\omega, t) = \hat{\xi}^*(\omega)\, e^{-i\omega t} \xi\,(t) \tag{10.5}$$

in which $\hat{\xi}^*(\omega)$ is the complex conjugate of the Fourier transformed $\xi(t)$, we have obtained what we introduced above as the CoSTID. From this complex valued function of ω and t one may calculate dynamic spectra with arbitrary spectral or temporal resolution (Johannesma & Aertsen, 1982).

Being a function of both frequency and time the CoSTID seems intuitively suited to our view of sound properties that are relevant to auditory neurons. Direct integration over frequency or time of equation (10.5) gives the temporal respectively the spectral intensity density

$$I(t) = \frac{1}{2\pi} \int \Xi\,(\omega, t)\, \mathrm{d}\omega \tag{10.7}$$

and

$$J(\omega) = \int \Xi\,(\omega, t)\, \mathrm{d}t, \tag{10.8}$$

which we have encountered alread in Fig. 18. Representation of the CoSTID may be done by separately showing the real and imaginary part or integrating this in one picture by using colour: the phonochrome (Johannesma *et al.* 1981).

11. The spectro-temporal receptive field of auditory neurons

Consider an auditory neuron, e.g. in the cochlear nucleus complex of the cat or the torus semicircularis of the grassfrog for which spikes are recorded during continuous gaussian wide-band noise stimulation. During the experiment the spikes are used (after an appropriate delay) to trigger a dynamic spectrum analyser which then processes the noise signals preceding the spikes over an interval of some length, say 81 ms. After about 3000 averages we may obtain a picture that consists of three regions (cf. Fig. 19): a more or less uniform

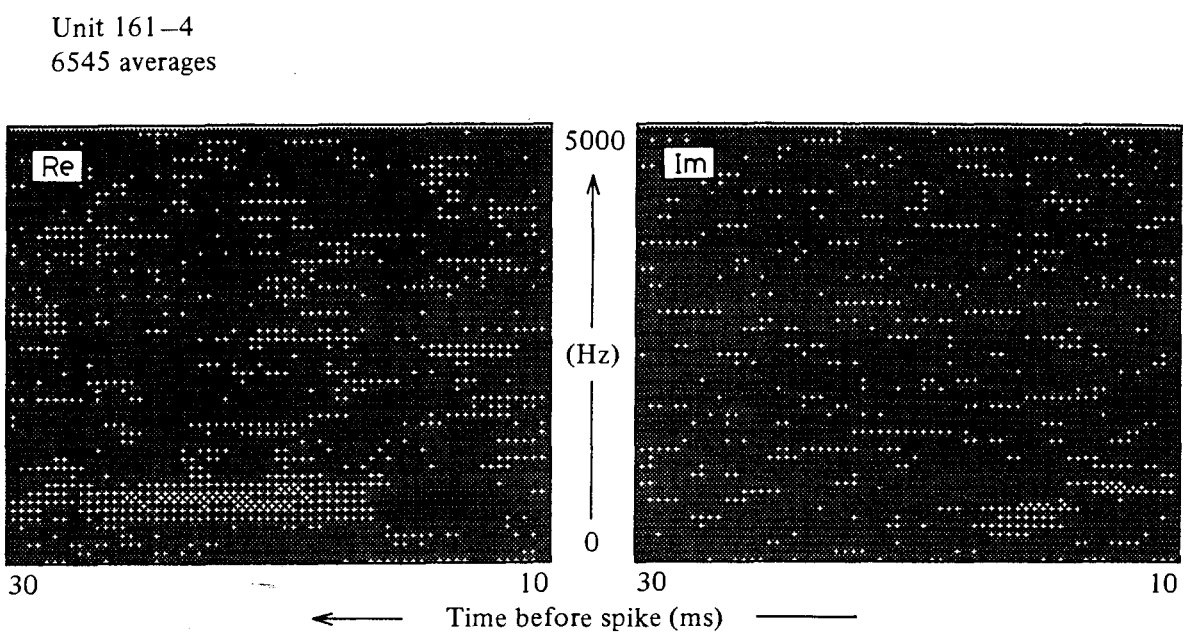


Fig. 20. Spectro-temporal receptive field for the same unit as in Fig. 19 as determined with the CoSTID technique. The real part shows a large similarity to the result obtained with the dynamic spectrum analyser. Note the linear frequency scale for the CoSTID as compared to the log-scale for the dynamic spectrum, note also the different time scale. (From Hermes *et al.* 1981.)

background, a dark region around 10 ms before the spike and a light region extending to about 30 ms before the spike representing respectively activation and post-activation suppression.

We may also process the spikes and the noise off-line and now form over a time window running from 10 to 30 ms preceding a spike the complex product $\Xi(\omega, t)$ for each individual pre-event stimulus and averaging them. The result is represented in Fig. 20. Take into account that the average CoSTID is represented on a linear frequency scale and the average pre-event sonogram on a logarithmic frequency scale and observe that there is a close correspondence between the real part of the average $\Xi(\omega, t)$ and the sonogram. Those parts of both representations that differ from the average level in a significant way constitute the *spectro-temporal receptive field* (Hermes *et al.* 1981). This STRF indicates the average characteristics of a second-order functional of the stimulus that 'causes' the spikes. We will survey various types of STRF as observed for neurons in the cochlear nucleus of the cat (Aertsen *et al.*, unpublished results) and tones semicircularis of the grassfrog (Hermes *et al.* 1981; Eggermont, Epping & Aertsen 1983*c*, *d*).

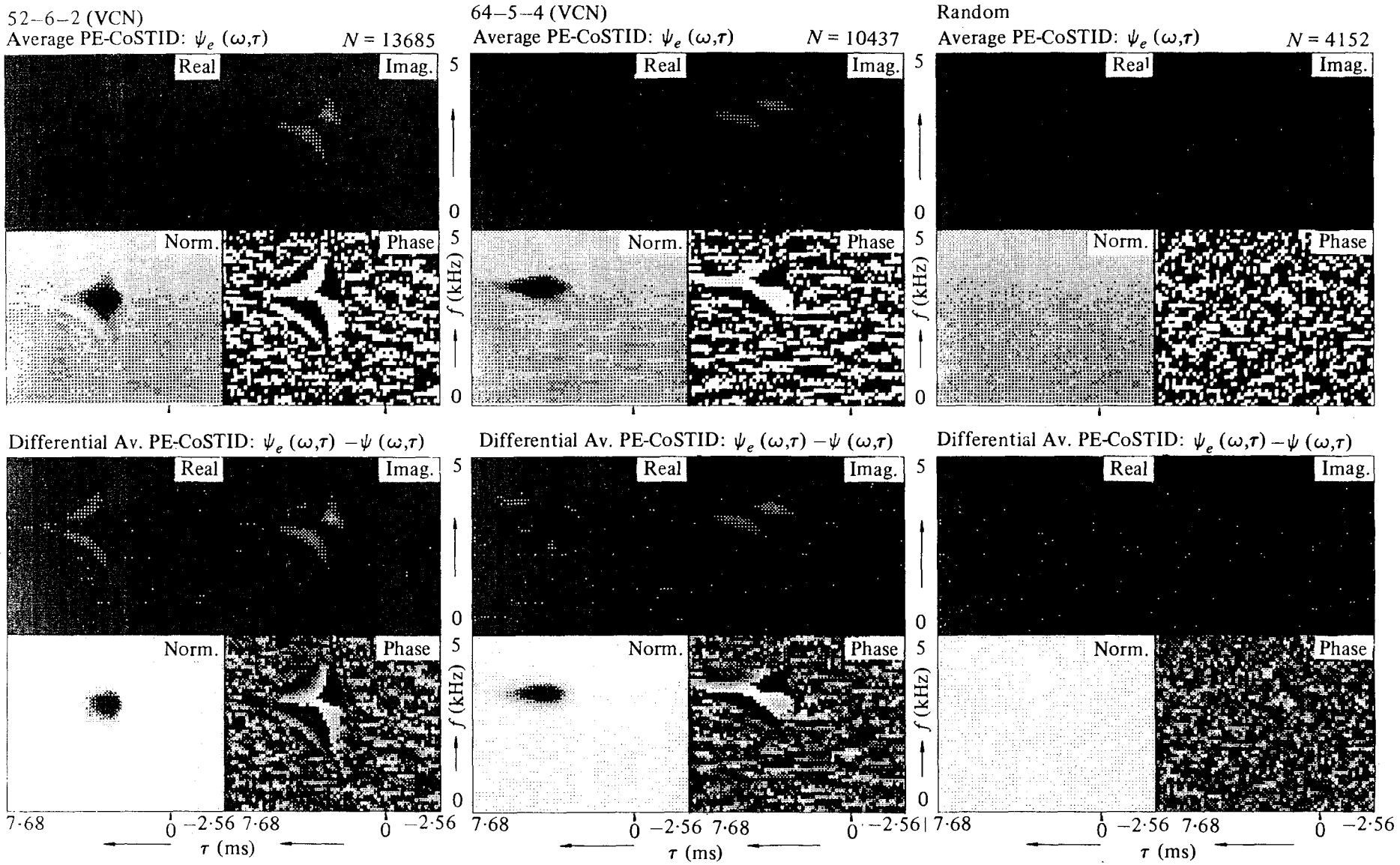
(*a*) *The cochlear nucleus complex of the cat*

In Fig. 21 we show for the two neurons from Fig. 18 the CoSTID both as real plus imaginary part as well as norm and phase representation. On the left-hand side a STRF is given for a neuron that showed also a clear reverse correlation (cf. Fig. 18) and had a best frequency of about 3 kHz and a latency of around 4 ms. In the middle column the STRF is given for a unit that showed no reverse correlation (cf. Fig. 18) although the BF was only slightly larger than for the other unit. Both units were from the ventral cochlear nucleus. The right-hand side represents the CoSTID for random triggers, this is an estimate of the background activity in the representation for the auditory units. The upper row shows the raw data, the lower row shows the corrected CoSTIDs, i.e. the spectro-temporal receptive fields. For unit 52–6–2 (left column) we observe suppression bands on both sides of and partly preceding the activation area in the representation of the real part. This spectro-temporal suppression area causes the depression in the average temporal intensity already signalled in connexion to Fig. 18. Unit 64–5–4 shows activation only (Johannesma & Aertsen, in preparation).

(*b*) *The torus semicircularis of the grassfrog*

Neurons in this area, a homologue of the inferior colliculus, show more complex STRF patterns than those in the cat's cochlear nucleus. Besides differences in temperature which affect the degree of phaselock (cf. section 8) there are also considerable differences in the inner ear between cats and frogs. The frog's inner ear consists of two separate receptor conglomerates, the amphibian papilla and the basilar papilla (Capranica, 1976). In the grassfrog, *Rana temporaria* L., the basilar papilla is tuned to a single frequency band around 1250–1700 Hz, thus units with BF from 1 to 2 kHz will originate from connexions with the basilar papilla. The amphibian papilla is a tonotopically organized structure roughly dividable in two regions. One gives rise to units with BF between 400 and 800 Hz, the other to BF's below 400 Hz. Two-tone suppression effects (Capranica & Moffat, 1980) are found for excitatory frequencies < 400 Hz and inhibitory frequencies in the 400–800 Hz region.

In Fig. 22 we represent four STRF's which exemplify, from top to bottom, activation only, activation followed by post-activation suppression, activation with lateral suppression and post-activation suppression, double tuning: two activation areas of which the lower


52–6–2 (VCN)
Average PE-CoSTID: $\psi_e(\omega,\tau)$
$N = 13685$
64–5–4 (VCN)
Average PE-CoSTID: $\psi_e(\omega,\tau)$
$N = 10437$
Random
Average PE-CoSTID: $\psi_e(\omega,\tau)$
$N = 4152$
Real
Imag.
Norm.
Phase
5
0
5
0
f (kHz)
Differential Av. PE-CoSTID: $\psi_e(\omega,\tau) - \psi(\omega,\tau)$
7·68
0
−2·56
7·68
0
−2·56
τ (ms)

one is accompanied by a suppression area. The upper two units are from the basilar papilla, the lower two units are from the amphibian papilla. Note that both the time window and the frequency range are different.

From this short survey it may become clear that, using the CoSTID approach with gaussian wide-band noise as a stimulus, one is able to derive for one complex-stimulus presentation both BF and $\Delta\omega$ as well as latency and Δt. In addition, the presence of lateral- and/or post-activation suppression, the presence of multiple tuning, the occurrence of inhibition and a measure for the information carrying capacity of the neuron (from the product $\Delta\omega \,.\, \Delta t$) is obtained. This overwhelmingly powerful characterization, however, can only be obtained off-line and at the expense of lengthy calculations. By sacrificing some sophistication this might be done in real time using a dynamic spectrum analyser (Aertsen & Johannesma, 1980) and thus obtain a guideline for efficient experimenting. In addition it must be noted that higher up in the CNS the number of units that responds in a stationary way to continuous wide band noise becomes smaller, so that this characterization cannot be applied uniformly. This, however, can be overcome by applying dynamic stimuli, combined with an appropriately adapted application of the CoSTID-analysis.

12. Reverse correlation function and spectro-temporal receptive field

After having shown the advantages of the STRF in characterizing the neural properties under stationary stimulus conditions it will be helpful to convert the reverse correlation function into an analogous representation. For this purpose we consider the reverse correlation function as the signal $x(t)$ and calculate the analytic signal as illustrated in Fig. 23. We may then form temporal intensity $\xi^*(t)\,\xi(t)$, the spectral intensity $\hat{\xi}^*(\omega)\,\hat{\xi}(\omega)$ and the CoSTID.

Fig. 21. Spectro-temporal receptive field representations for two neurons from the VCN in the cat. Shown are the same neurons as in Fig. 18. The upper part of the figure shows the average pre-event CoSTIDs, both in the real–imaginary part representation and in the norm and phase representation. Note that darker areas represent regions with more than average intensity density and lighter areas regions with less than average intensity density. The unit in the middle column shows a 'simple' receptive field, while the unit in the left column shows an indication of suppressive side-bands surrounding the excitation area. The right column shows the result for spontaneous activity correlated with the noise. The lower part of the figure shows the same results when the expected values obtained for random activity are subtracted. (From Johannesma & Aertsen, in preparation.)

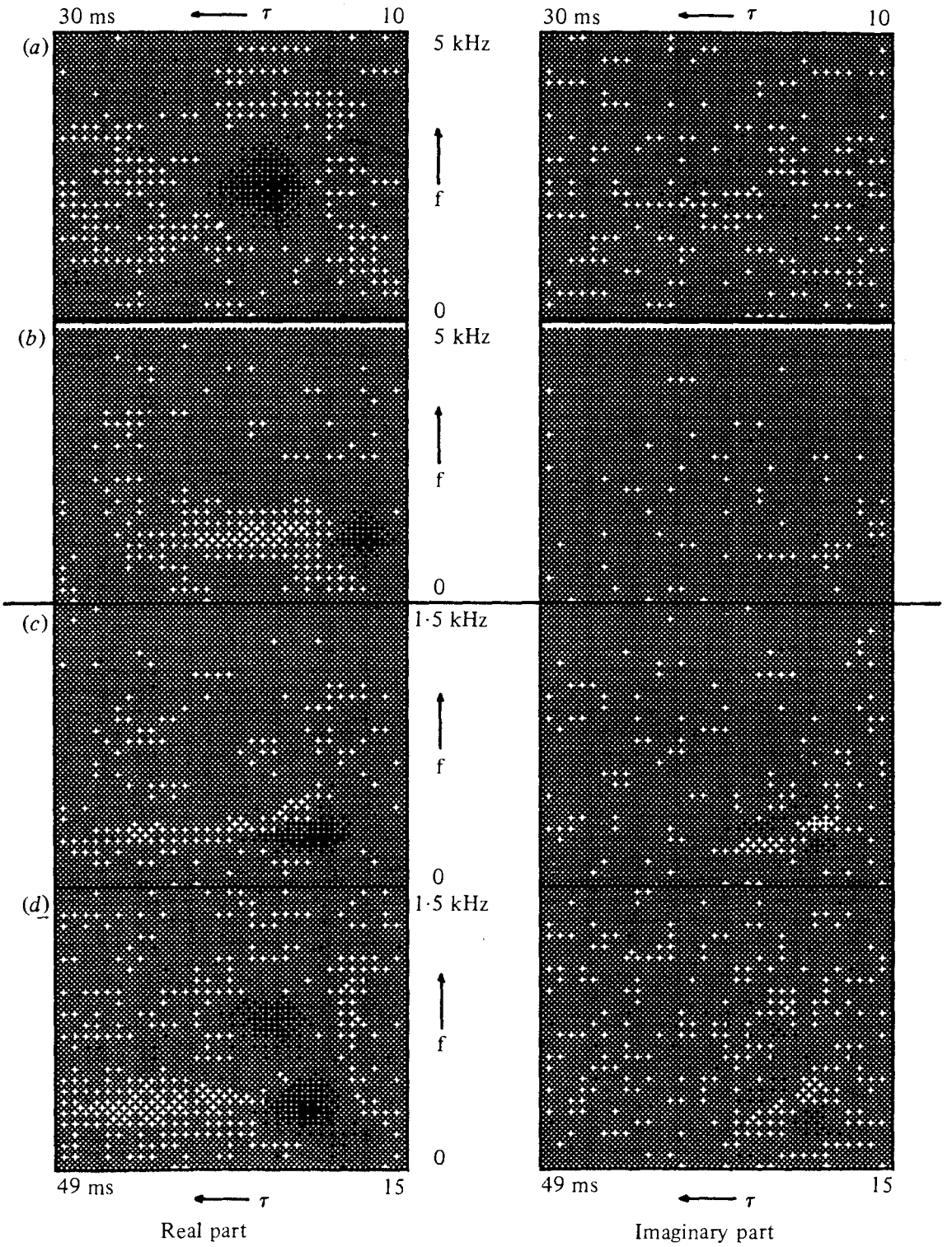


Fig. 22. Spectro-temporal receptive fields for four neurons from the torus semicircularis in the grassfrog as obtained from gaussian wide-band noise stimulation. The left-hand column shows the real part and the right-hand column the imaginary part of the complex valued CoSTIDs. For the neurons represented in the upper two rows a 5 kHz low-pass noise was used as stimulus and the time base runs from 10–30 ms before the spike. The neurons represented in the lower two rows were stimulated with a 1·5 kHz low-pass noise and the time base runs from 15 to 49 ms. In part (*a*) one observes a single excitation area (dark region); in part (*b*) an excitation area preceded by a suppression area (light region); in part (*c*) suppression is partially at the same time before the spike as activation (lateral suppression) and extends after the activation (post-activation suppression); in part (*d*) one observes two activation areas (double tuning) of which the lower one is accompanied by post-activation suppression.

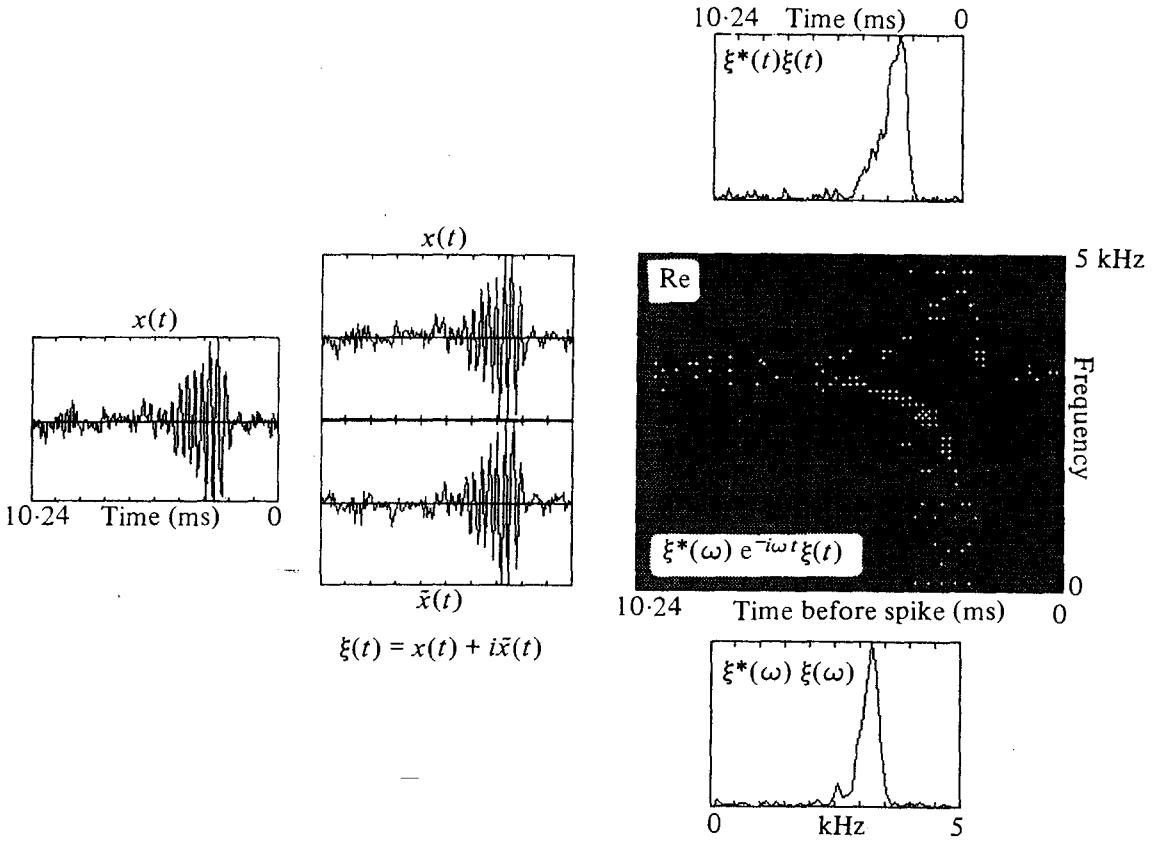


Fig. 23. Construction of the spectro-temporal receptive field on basis of the reverse correlation function for the VCN in the cat. From the reverse correlation function $x(t)$ the analytic signal $\xi(t)$ is constructed and its Fourier transformation $\xi(\omega)$ is determined. The temporal intensity $\xi^*(t)\,\xi(t)$, spectral intensity $\xi^*(\omega)\,\xi(\omega)$ and CoSTID $\xi^*(\omega)\,e^{-i\omega t}\,\xi(t)$ are calculated and shown. There is a best frequency of 3·2 kHz and a latency of 2·2 ms. From the CoSTID the impression arises of a lateral suppression for frequencies below the best frequency.

However, this CoSTID is based on an averaged signal: the reverse correlation function. The STRF is based on averaging CoSTIDs of elements of the pre-event stimulus ensemble. The difference between the two representations is a measure for the amount of phase-lock. Since the amount of phase-lock is predominantly determined by the frequency, the most straightforward measure will be based on $\hat{\xi}(\omega)$. If we indicate averaging by $\langle\ \rangle$ the amount of phase-lock $c(\omega)$ can be characterized by

$$c(\omega) = \frac{\langle\hat{\xi}^*(\omega)\rangle\langle\hat{\xi}(\omega)\rangle}{\langle\hat{\xi}^*(\omega)\,\hat{\xi}(\omega)\rangle}. \tag{12.1}$$

This can be written as

$$c(\omega) = \frac{\langle\hat{\xi}(\omega)\rangle^*\langle\hat{\xi}(\omega)\rangle}{\langle\hat{\xi}^*(\omega)\,\hat{\xi}(\omega)\rangle} = \frac{J_1(\omega)}{\langle J(\omega)\rangle}, \tag{12.2}$$

with $J_1(\omega)$ being the spctral intensity of the average pre-event stimulus (or reverse correlation function) and $\langle J(\omega)\rangle$ is the average spectral intensity of the PESE.

Since both $J_1(\omega)$ and $\langle J(\omega)\rangle$ are non-negative, $c(\omega) \geqslant 0$; together with the triangular inequality this leads to $0 \leqslant c(\omega) \leqslant 1$ for any ω. Note that $c(\omega)$ is still a function of frequency and in fact is a combined measure of spectral sensitivity and phase-locking.

This relation has been explored for neurons in the cochlear nucleus complex of the cat (Aertsen *et al.*, unpublished results). For 18 neurons responding in a sustained way to stationary noise, six were found to respond in a primary like way. Of these primary-like neurons, five showed a first-order reverse correlation function and for all five units $c(\omega)$ was close to 1. The 12 non-primary like neurons showed in six cases a reverse correlation function but the $c(\omega)$ was in all those cases much smaller than 1. The loss or absence of phase-lock is the most simple explanation to account for the differences between the STRF and the CoSTID of the reverse correlation function. However, the spectro-temporal receptive field in fact is an average of a second-order functional of the stimulus, thus differences with the CoSTID of the reverse correlation function can also arise as a result of a pronounced second (or higher even-) order non-linearity.

13. Correlation functions for tonal stimulus ensembles

Most readers will be familiar with the forward correlation procedures currently in use for the analysis of neural spike activity to stimulation with tonepips, tonebursts or continuous tones. When tonepips and tonebursts are used the tone-frequency is usually kept fixed and the stimuli are repeatedly presented. The event-density of the neural response is estimated by correlating the toneburst onset with the occurrence of spikes: the PST-histogram. When continuous tones are used the occurrence of spikes is correlated with the onset of the tone period: period histograms (e.g. Arthur, 1976). These procedures can be repeated for other frequencies and in this way the receptive field of the auditory neuron can be mapped sequentially (e.g. Gerstein, Butler & Erulkar, 1968; Van Gisbergen *et al.* 1975*a*; Webster & Aitkin, 1975). Random-frequency random-intensity tones have been employed in automated paradigms for obtaining single-fibre tuning curves (e.g. Evans, 1974) together with an indication of the firing rate of the nerve fibre within the response area.

A reverse correlation procedure relies to a large extent on this randomized stimulus presentation. By using a stimulus ensemble

consisting of tonepips with randomly selected frequency values but constant amplitude one may obtain a statistically structured stimulus ensemble having a flat spectrum. Optionally also the amplitude values may be randomly selected. For instance, a tonal sequence consisting of 16 ms duration tonepips selected from 255 frequency values and 127 amplitude values presented without silent intervals then requires a duration of 518 s (e.g. Aertsen & Johannesma, 1980). When silent intervals are included such sequences become prohibitively long and the amplitude variation must be reduced (e.g. Hermes *et al.* 1982).

Cross-correlation between spikes and the stimulus ensemble again will result in an averaged pre-event stimulus ensemble as discussed before for noise as a stimulus. Because a tonal stimulus ensemble consists of narrow-band stimuli we may parametrize each tonepip into its instantaneous frequency and its time envelope. Reverse correlation then is equivalent to averaging this time envelope for each frequency band. When, for example, 255 frequency values are grouped into 32 frequency bands this becomes practical. The resulting picture represents the average stimulus intensity (square of the tonepip envelope) as a function of tone frequency and as a function of time before the action potential and is called the average IFT (Aertsen *et al.* 1980). For random triggers, or spontaneous activity, the average IFT is flat and more or less noisy, depending on the number of averages carried out. For auditory neural units where the spikes are caused by certain features of the stimulus, the average IFT represents the spectro-temporal sensitivity (STS) of the neuron. This average IFT will differ from that obtained in case of spontaneous activity for those frequency bands for which the neuron is sensitive and at those times before the occurrence of the action potentials that reflect both the response latency and the integration time of the neuron. The STS is closely related to the STRF and can be converted into it by certain stimulus normalization procedures (cf. Section 15). The STS in itself is an important neural response characteristic (Hermes *et al.* 1982; Johannesma & Eggermont, 1983) and an example is shown in Fig. 24.

When the amplitude of the tonepips is fixed, the resulting three-dimensional picture may be considered as a family of contours related to the iso-intensity contours (cf. Section 4). The relation becomes more straightforward when the IFT representation is integrated over time to give an intensity–frequency contour. In case the system behaves in a linear way these IF contours should be of the same shape as the tuning curves. For the auditory system this will generally not

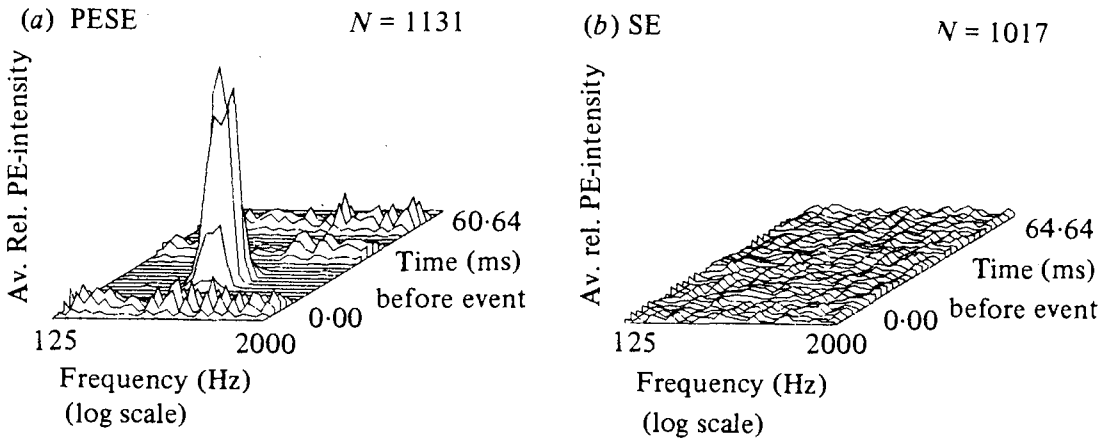


Fig. 24. Intensity–frequency–time (IFT) representation for a neuron from the torus semicircularis of the grassfrog. Shown on the left is the average intensity over the pre-event stimulus ensemble as a function of the tonepip frequency and time before the event. On the right-hand side the average intensity is shown over the entire stimulus ensemble for random triggers. On the left one observes a best frequency of around 500 Hz and a latency around 20 ms. (From Aertsen *et al.* 1980.)

be the case and a comparison of the results of the IFT analysis with the more familiar tuning curve analysis will show remarkable qualitative differences, as we will see shortly.

Determination of the STS has to obey the restriction of 'stationarity' of the spike trains evoked by the stimulus-ensemble. In the particular case of 1 toneburst s^{-1} one cannot speak about stationarity in the ordinary sense. We will require, however, that a subsequent presentation of the whole stimulus sequence, which may last several minutes, produces the same number of spikes (within statistical limits). Usually (e.g. Eggermont *et al.* 1981) a complete stimulus ensemble consists of several sequences of tones and in most cases the responses to the first sequence have to be left out of the analysis because of a gradually decreasing firing rate. Aertsen *et al.* (1980) remark that 'application of the analysis to segments of these recordings suggest an increased spectro-temporal selectivity in the adapted stage as compared to the initial stage'. This suggests that the STS sharpens up in the stationary stage as compared to the transient stage of the neuronal activity. This was found to occur quite generally: non-adapting neurons in the torus semicircularis in general showed sharper receptive fields (in both frequency and time) combined with shorter (10–35 ms) latencies than the slowly adapting neurons with latencies of 40–70 ms or more. The existence of a subpopulation of more complex neurons in the torus semicircularis might therefore be indicated.

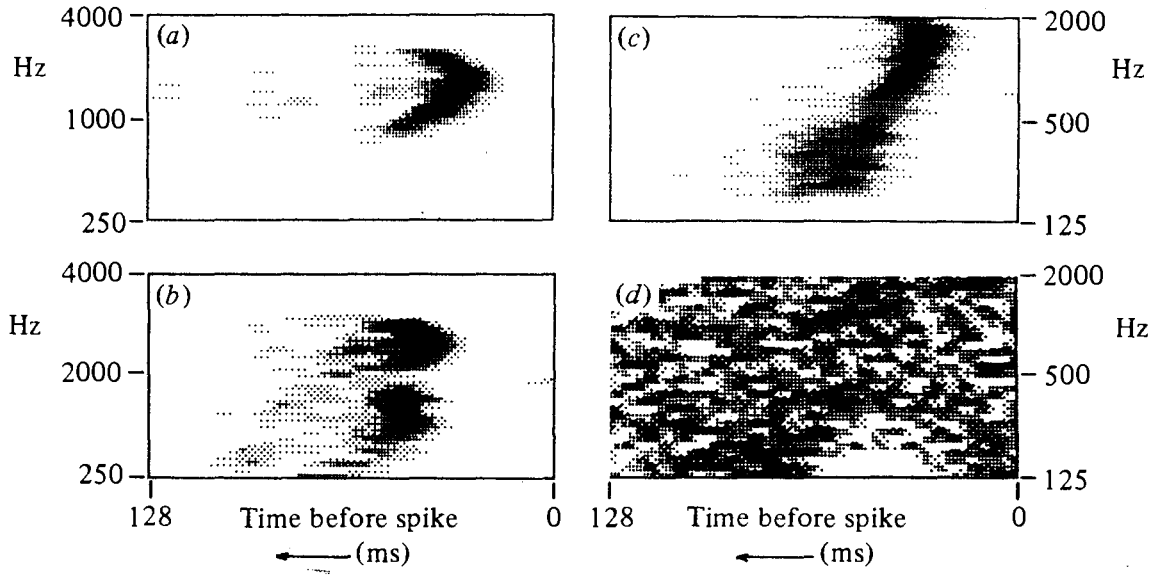


Fig. 25. Survey of different IFT types for neurons in the torus semicircularis of the grassfrog. (*a*) Excitation for a frequency region of 1–2·5 kHz (basilar papilla) with a latency depending strongly on the frequency. (*b*) Double tuning with one region (1–2 kHz) arising from the basilar papilla and the other region (250–900 Hz) with input from the amphibian papilla. (*c*) Broad tuning with a latency that depends on the frequency of stimulation. (*d*) Inhibition (125–250 Hz) caused by the tonal stimulus ensemble in this spontaneously active neuron; the latency is around 50 ms.

The use of tonal stimuli to determine the STS of units from the torus semicircularis will be illustrated by a few examples shown in Fig. 25. For non-spontaneously active units only excitation can be demonstrated, we distinguish single tuning (*a*) from a unit deriving its input from the basilar papilla, double tuning (*b*) with input from the basilar papilla as well as the amphibian papilla, and broad tuning (*c*) over a wide frequency range with a pronounced latency dependence on frequency. In case units are spontaneously active (*d*) one may demonstrate suppression; in this case frequencies below 250 Hz inhibited the unit with a latency of around 50 ms.

Hermes *et al.* (1982) found for 83 units in the torus semi-circularis of the lightly anaesthetized grassfrog that 69 units were simply tuned and 14 had double or broad tuning. In the same auditory nucleus of the same species, Walkowiak (1980) only found in 3 % of the cases a double tuning. Fuzeserry & Feng (1982) confirmed this for 130 units in the torus semi-circularis from *Rana pipiens*: they found double tuning in 4 % of the neurons investigated. Part of this discrepancy lies in the comparison of two vastly different methods: the results from Hermes *et al.* (1982) have to be compared with high level ($\simeq$ 89 dB SPL) iso-intensity curves or with a high-response-rate tuning curve. It appeared from the data reported by Fuzeserry and

Feng that some units with simple V-shaped tuning curves turned out to be bimodal for high rate criteria. The genuine double-tuning curves all had an inhibitory region in between. Whether the tuning curve or the STS represents the true frequency selectivity depends on the question one asks. For the inter-species communication a high-intensity measure might be more suited than a threshold criterion; the latter, however, is the more commonly used.

14. Is the spectro-temporal sensitivity separable?

The STS of auditory neurons as determined with tone sequences can be described by the IFT method, i.e. as $I(f, \tau)$ where I is the average signal intensity as a function of frequency and time before the event. One may wonder if this two-dimensional description is necessary or simply puts an unnecessary demand on available computer power and time. The question may be asked whether two separate descriptions such as the tuning curve and a PST-histogram at CF or to clicks would not be sufficient. This centres around the problem of separability: the independence of spectral and temporal properties of $I(f, \tau)$. Basically this was recognized already in 1968 by Gerstein *et al.*, where they state, in a study on toneburst stimulation of the cat cochlear nucleus, that 'the firing of a unit can be approximately a function of frequency alone multiplied by a function of time alone, or in more complicated cases it can be a function of both frequency and time'.

If we could interpret $I(f, \tau)$, after proper normalization, as a joint probability density function for finding specific (f, τ)-pairs in the stimulus preceding the event, and having marginal density functions defined by

$$I_1(f) = \int_0^\infty I(f, \tau)\, \mathrm{d}\tau \tag{14.1}$$

and

$$I_2(\tau) = \int_0^\infty I(f, \tau)\, \mathrm{d}f \tag{14.2}$$

then the concept of independence of f and τ would imply

$$I(f, \tau) = I_1(f) \,.\, I_2(\tau). \tag{14.3}$$

Conversely, if equation (14.3) holds, then the independence of the random variables follows and separability is thus established. One may interpet this as the absence of a spectro-temporal structure in the pre-event stimulus ensemble.

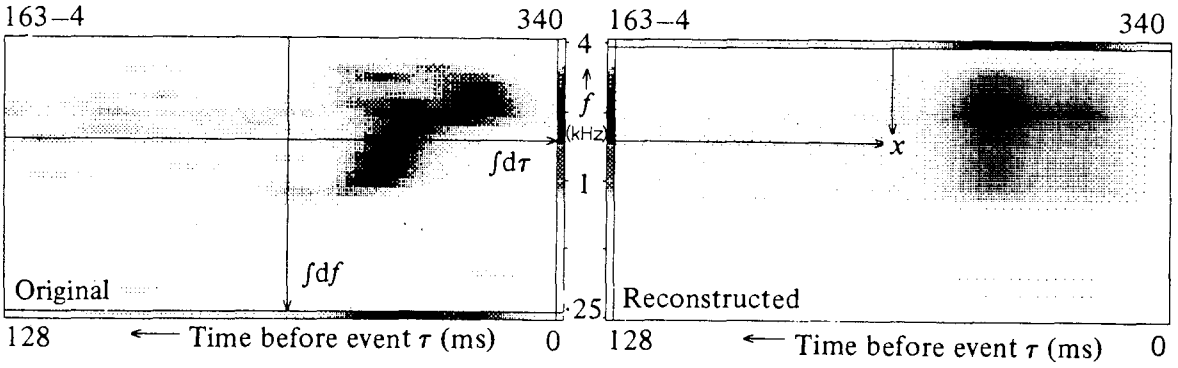


Fig. 26. Separability of the IFT. In the left-hand part an IFT is shown for a torus semicircularis neuron that is activated by basilar papilla input. The vertical column on the right indicates this frequency region and is obtained by integration over time. The horizontal bar below the IFT represents integration over frequency and gives the latency distribution. On the right-hand part of the figure the IFT is reconstructed under the assumption of separability by multiplication of the frequency spectrum (spectral intensity) and the latency distribution (temporal intensity). Since the right-hand part of the figure does not resemble the left-hand part, the assumption of separability is violated. (From Eggermont *et al.* 1981.)

Eggermont *et al.* (1981) have investigated separability by studying the validity of equation (14.3) (cf. Fernald & Gerstein, 1972). This is illustrated in Fig. 26, where on the left hand $I(f, \tau)$ is represented for neuron 163–4 from the torus semicircularis, together with both marginal densities. Assuming independence one may now construct a joint function

$$\tilde{I}(f, \tau) = I_1(f) \,.\, I_2(\tau) \qquad (14.4)$$

The similarity between $I(f, \tau)$ and $\tilde{I}(f, \tau)$ is a measure for separability. A necessary and sufficient condition for separability is that $I(f, \tau)$ equals $\tilde{I}(f, \tau)$. In Fig. 26 $I(f, \tau)$ is shown on the left-hand side and $\tilde{I}(f, \tau)$ on the right-hand side; it can be concluded that this $I(f, \tau)$ is inseparable. This was tested for 83 neurons from the torus semicircularis of the lightly anaesthetized grassfrog and related to recording site and spike waveform. It appeared that $I(f, \tau)$ in non-adapting neurons and in the incoming fibre population were separable. The 23% of the neurons that appeared to have inseparable parts were located dominantly in the caudoventral parts of the torus semicircularis. This region shows multiradiated cells capable of integrating input from a large number of units. This indicates that a spectro-temporal characterization is not only mandatory in about a quarter of the neurons in the torus semicircularis but, by means of the separability concept, also offers a new way to understand information processing in the central nervous system.

15. The use of reverse correlation with natural stimuli

Natural stimuli can be used to study the spectro-temporal sensitivity of auditory neurons much in the same way as noise or tonal stimuli. When using the reverse correlation approach with noise and tonal stimuli we stressed the statistical structure of the stimulus ensemble as well as the whiteness of the spectrum. An ensemble of natural stimuli, the acoustic biotope (Aertsen & Johannesma, 1980; Smolders, Aertsen & Johannesma, 1979) including vocalizations of many species, however, has considerable internal structure and in general will not have a flat spectrum. Therefore correction procedures must be applied to obtain an unbiased STS.

Specific reasons for the use of natural stimuli in neurophysiological research are based on the notion that these stimuli are meaningful to the animal under investigation (at least if it is not anaesthetized) and that, for example, auditory system of lower vertebrates has evolved specifically to process these stimuli (e.g. Capranica, 1976). Furthermore, natural stimuli may be used to test if there is any gain in understanding the nervous system using these natural stimuli as compared to the use of tonebursts or noise. This issue has been addressed quite globally by Smolders *et al.* (1979), who compared responses to natural and tonal stimuli from single units in the cat cochlear nucleus and medical geniculate body. While the responses of most cochlear nucleus units to natural stimuli could be understood on basis of the STS as determined with tonebursts, this was no longer possible for the MGB.

Aertsen *et al.* (1981) addressed themselves to analyse methodically the responses from single neurons in the torus semicircularis of the grassfrog to a long ($\simeq$ 8 min) ensemble of natural stimuli. The reverse correlation approach was used: the average pre-event stimulus spectrogram was determined using a dynamic spectrum analyser. Because of the non-whiteness of the stimulus spectrum an equalization procedure was carried out to obtain a less-biased estimate of the STS. For the units studied the authors concluded that 'it appeared that the results of natural stimulation of the stationary units, on the whole, agree[d] 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.'

A spectro-temporal sensitivity determined on basis of natural

stimuli may remotely show some similarity to that obtained with tonal stimuli or noise (Johannesma & Eggermont, 1983). While unstructured stimulus ensembles allow for a straightforward stimulus correction and are then readily interpreted, the results obtained with highly structured natural stimuli (including species-specific vocalizations) do not allow much of a conclusion. It must be remarked that the reverse correlation method using this type of natural stimulus ensemble is not suited to find out if neurons are highly specialized, e.g. respond only to very specific parts of vocalizations such as found in the CNS of birds (Leppelsack & Vogt, 1976). Due to the intricate time structure of vocalizations the average pre-event spectrogram made for such neurons in case of stimulation with a natural stimulus ensemble will nearly always contain most if not all of the complete vocalization to which features they respond. A possible way, in principle, to overcome this problem would be to generate artifically a 'pseudo-natural' stimulus ensemble, containing sufficient statistical variations and/or distortions of the natural sounds to create a rich pool of possible stimulus configurations that is wider than the neurons selective properties. What the reverse correlation type of approach using the present natural stimulus ensemble yields is a general technique to investigate the important concept of stimulus invariance of the STS or STRF as determined for noise, tones and natural stimuli (cf. Section 17).

16. Functional description of non-linear systems

The auditory system behaves in a highly non-linear way, i.e. the characterization of the system depends on stimulus intensity as well as on the type of stimulus. We have seen that, under the assumption that the auditory periphery can be modelled as a static non-linearity sandwiched between two linear filters (cf. Fig. 8), reverse correlation will result in the impulse response of the cascaded linear filters (equation 5.3). However, this impulse response does not describe the behaviour of the overall system. What we need is an extension of the methods to characterize linear systems that we have dealt with. Such a method can be offered by the functional representation of deterministic, constant parameter, finite-memory non-linear systems. Generally, in case of a non-linear system the output $y(t)$ is related in a functional way to the input $x(t)$:

$$y(t) = (Sx)(t). \tag{16.1}$$

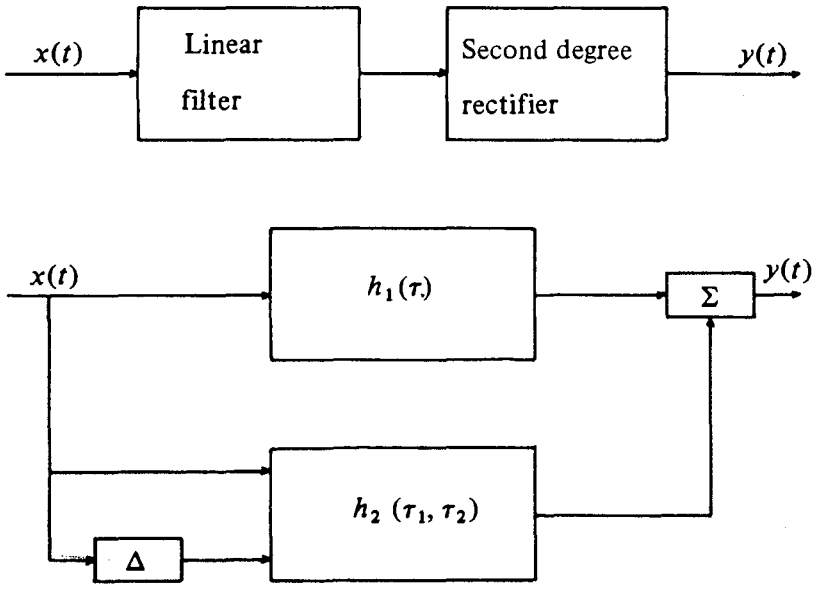


Fig. 27. A Volterra representation of a cascaded second-order non-linear system is in the form of two parallel filters, $h_1(\tau)$ and $h_2(\tau_1, \tau_2)$.

Under certain smoothness conditions S can be represented by a functional expansion of homogeneous polynomials:

$$S(x) = \sum_{n=0}^{\infty} K_n(x), \tag{16.2}$$

which is in fact a Taylor-like functional power series. This is known as the Volterra series expansion (Barrett, 1963),

$$y(t) = \sum_{m=0}^{\infty} (V_n x)(t), \tag{16.3}$$

in which

$$(V_n x)(t) = \int d\tau_1 \ldots \int d\tau_n v_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} x(t-\tau_i) \tag{16.4}$$

is the nth-order Volterra functional describing *completely* the contribution of the nth-order system non-linearity to the output $y(t)$. For example, for a system comprising a linear part and a quadratic part one has to find the first- and second-order Volterra kernels $v_1(\tau_1)$ and $v_2(\tau_1, \tau_2)$ to describe the system completely. A simple example would be formed by a linear filter followed by a second-degree rectifier. The Volterra series expansion then transforms this system in a linear part *parallel* to a quadratic-order part, the outputs of which are then added (Fig. 27). If the system is to a large extent unknown, as the auditory system still is, experimental ways to determine the Volterra kernels

$v_n(\tau_1, \ldots, \tau_n)$ are not available. It is possible, however, to rearrange the Volterra functionals into a new series.

$$y(t) = \sum_{n=0}^{\infty} (W_n x)(t), \tag{16.5}$$

where the W_n formed by a linear combination of V_n, V_{n-2}, V_{n-4}, etc., are called Wiener functionals. The Wiener functionals have the advantage that they are mutually orthogonal with respect to a Gaussian white noise input. Unlike the V_n, the W_n depend on the noise characteristics, e.g. the spectrum and the power level P. The form of W_n is the same as the nth Hermite polynomial $He_n(x)$ (cf. Johnson, 1980):

$$He_n(x) = x^n - \binom{n}{2}\frac{P}{2}x^{n-2} + 1\cdot 3\binom{n}{4}\left(\frac{P}{2}\right)^2 x^{n-4} - \ldots, \tag{16.6}$$

from which it can be seen that all odd-order Wiener functionals will contain a linear term, or in other words contribute to the first-order Volterra functional. Therefore the first-order Wiener functional does not, in general, describe the linear part of the system. When the system is not higher than second order the first- and second-order Volterra- and Wiener kernels are identical. Wiener kernels $w_n(\tau_1, \ldots, \tau_n)$ have the great advantage that they can be obtained *independently* from each other out of input–output cross-correlation (Lee & Schetzen, 1965). For a third-order system the first three Wiener kernels are estimated as

$$\tilde{w}_1(\tau) = \frac{1}{T}\int_0^T \mathrm{d}t\, x(t-\tau)\, y(t), \tag{16.7}$$

$$\tilde{w}_2(\tau_1, \tau_2) = \frac{1}{T}\int_0^T \mathrm{d}t\, x(t-\tau_1)\, x(t-\tau_2)\, y(t), \tag{16.8}$$

$$\tilde{w}_3(\tau_1, \tau_2, \tau_3) = \frac{1}{T}\int_0^T \mathrm{d}t\, x(t-\tau_1)\, x(t-\tau_2)\, x(t-\tau_3)\, y(t). \tag{16.9}$$

They can be computed relatively easy and describe the system completely. Volterra kernels can be computed from the Wiener kernels according to

$$v_n(\tau_1, \ldots, \tau_n) = \sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{(n+2\nu)!}{n!\nu!}\left(\frac{P}{2}\right)^{\nu} \times \int \mathrm{d}\sigma_1 \ldots \int \mathrm{d}\sigma_{\nu}\, w_{n+2\nu}(\tau_1, \ldots, \tau_n, \sigma_1, \sigma_1, \ldots, \sigma_{\nu}, \sigma_{\nu}) \tag{16.10}$$

(cf. Aertsen & Johannesma, 1981).

The Volterra kernels for the third-order system therefore are

$$v_1(\tau) = w_1(\tau) - 3P\int d\sigma\, w_3(\tau, \sigma, \sigma), \qquad (16.11)$$

$$v_2(\tau_1, \tau_2) = w_2(\tau_1, \tau_2), \qquad (16.12)$$

$$v_3(\tau_1, \tau_2, \tau_3) = w_3(\tau_1, \tau_2, \tau_3), \qquad (16.13)$$

and describe the contributions of respectively the linear part, the quadratic part and the third-order non-linearity to the systems output. In this particular case the second- and third-order Wiener kernels completely describe the second- and third-order non-linearity, but the first-order Wiener kernel does not completely determine $v_1(\tau)$.

From the definition of $w_1(\tau)$ we conclude that the first-order Wiener kernel is equal to the reverse correlation function (cf. equation 5.1). In Section 5 it was shown that for auditory-nerve fibres in rat and guinea pig and to a lesser degree in cat the reverse correlation is intensity-dependent. This can now easily be understood from equation (16.11). Since $v_1(\tau)$ is a system characteristic and therefore does not depend on stimulus level, $w_1(\tau)$ will depend on the noise power level because of the factor P before the term containing $w_3(\tau, \sigma, \sigma)$.

The auditory periphery shows its non-linearity among others in the production of cubic difference tones such as $2f_1 - f_2$, which are quite strong. This suggests that $v_3(\tau_1, \tau_2, \tau_3)$ and therefore w_3 will be quite large. It is therefore somewhat surprising that although $2f_1 - f_2$ is quite strong in cat auditory nerve fibres (Goldstein & Kiang, 1968) the reverse correlation function is much less intensity-dependent (De Boer, 1969) than for rodents (Møller, 1977; Harrison & Evans, 1982). A few intensity studies for cochlear nucleus units in the cat did not point to a pronounced intensity dependence of the reverse correlation function either (Grashuis, 1974).

The spectro-temporal receptive field as we have seen in Section 11 is, by virtue of two arguments, ω and τ, a second-order system characteristic. Since it is determined by a second-order cross-correlation it is formally identical to a single Fourier transform of $w_2(\tau_1, \tau_2)$; besides that the STRF is based upon the analytic signal $\xi(t)$ and the Wiener kernels on the real signal $x(t)$. The STRF only describes the second-order non-linearity of the system completely if the order of the system is not higher than three plus optional higher odd-order non-linearities.

17. Stimulus effects on the spectro-temporal receptive field

Stimuli may differ in their temporal and spectral structure. For constant-parameter linear systems one is able to define unique functions which characterize the system completely: the impulse response or its Fourier transform the transfer function. Combination of these two representations may result in a spectro-temporal characterization (cf. Section 12). This characterization will be the same whether clicks, noise or tonal stimuli are used or any other type of stimulus having a sufficiently wide spectrum. Given the impulse response one is able to predict the response of the constant parameter linear system to, for example, a frog vocalization.

Neural units can be characterized by their spectro-temporal receptive field as determined for noise stimulation or by their spectro-temporal sensitivity when using a tonal stimulus ensemble or an ensemble of natural stimuli. A stimulus invariant STRF would require that the STS after proper normalization procedures could be transformed into this STRF. In that case the neuron behaves the same whether stimulated with noise, tones or natural stimuli, and knowing the STRF would allow one to predict accurately the response to, for example, frog vocalizations provided that the higher-order system kernels are of minor importance. Since neural units can be found that do not respond in a stationary way to noise, or even cease firing when continuous noise is presented, but obviously respond very well to tones or vocalizations (e.g. Johannesma & Eggermont, 1983) one knows *a priori* that stimulus invariance of the STRF does not hold for all units.

Attempts to characterize auditory nerve fibre responses to complex stimuli on the basis of a linear approximation derived from tuning properties (e.g. Evans & Wilson, 1973; Goldstein, Baer & Kiang, 1971; Kiang & Moxon, 1974) or their reverse correlation function (Møller, 1977; De Boer, & De Jongh, 1978) have not been very successful. Although Kiang & Moxon (1972) have shown that the responses of cat auditory nerve fibres to human speech sounds can be grossly predicted from the tuning curves and the characteristics of the stimulus, it has not been possible to infer analytically the form of the responses to speech sounds from the responses to simple stimuli (Johnson, 1980).

Sachs & Young (1980) have investigated the coding of vowels as a function of intensity for auditory nerve fibres in the cat in terms

of average rate and cochlear place (i.e. characteristic frequency) versus synchrony across the neural population. They concluded that 'although a representation of speech in terms of rate and place does not seem to be as robust as the temporal representation, we have previously considered the compelling reasons to retain a representation which includes place as one parameter. In light of these arguments we should re-emphasize that the average temporal representation (for different fibres) is a combination of both place and temporal information'.

A spectro-temporal characterization of auditory nerve fibres and neural units in other auditory nuclei might therefore be a better basis for this prediction than the linear one. Successful prediction requires that the spectro-temporal characteristic and other (higher-) order system characteristics are stimulus invariants. The important problem of stimulus invariance of the spectro-temporal receptive field and its determination from the STS has been addressed by Aertsen & Johannesma, 1981*a*, *b*) and more heuristically by Johannesma & Eggermont (1983). The alternative approach is to find out whether response prediction is possible for neurons (e.g. in the torus semicircularis in the grassfrog) on basis of the STRF as determined for noise (Eggermont, Aertsen & Johannesma, 1983*a*, *b*).

We will first discuss stimulus invariance of the STRF, test for the completeness of the characterization on basis of the STRF, and thereafter consider the many problems that arise in response prediction procedures.

(*a*) *Stimulus invariance of the STRF*

The STRF has been defined as the 'sensitivity of auditory neurons with respect to the spectro-temporal intensity density of acoustic stimuli' (Aertsen *et al.* 1980). The term has also been used as the difference between the average second-order functional of the pre-event stimulus ensemble and the average second-order functional of the stimulus ensemble (Hermes *et al.* 1981). When the average second-order functional of the SE is not subtracted the outcome has been termed STS, especially in use for tonal stimulus ensembles (Hermes *et al.* 1982) with the IFT approach (cf. Section 13) or as an equivalent to the average pre-event spectrogram (Aertsen & Johannesma, 1981*b*). This manifold of meanings attributed to second-order system characteristics is not very elucidating but it reflects the interaction between experimental procedures and theoretical retrospection when realizing that each specific second-order correlation method seemed to illuminate other properties.

It would be helpful to reserve the name STRF for a stimulus invariant system characteristic. In the previous section we have shown that the second-order Volterra kernel is such a stimulus invariant characteristic. One could therefore postulate the single Fourier transform $\hat{v}_2(\omega, \tau)$ of the second-order Volterra kernel $v_2(\tau_1, \tau_2)$ as the STRF (Aertsen & Johannesma, 1981 *a*). The system to which this applies then has to be time-invariant and to have finite memory, while its firing probability is a smooth function of the input. We can therefore define an STRF provided that the requirements for convergence of the Volterra expansion are fulfilled. Defining an STRF, however, is a completely different subject than measuring it. The estimation of $\hat{v}_2(\omega, \tau)$ in general requires that all even-order Wiener kernels can be estimated (in case of stimulation with gaussian white noise) thus, implicitly assuming the system to be of finite order. For the auditory system a reasonable limitation of the systems non-linearity would be considering it as of third degree. In that case $\hat{v}_2(\omega, \tau)$ equals the single Fourier transform of $w_2(\tau_1, \tau_2)$ and agrees with the definition of the STRF as used by Hermes *et al.* (1981). Thus in that case the STRF will be stimulus invariant and can be measured using gaussian noise.

A similar procedure can be used in case statistically structured tonal ensembles are used. However, the normalized spectro-temporal sensitivity measured in this way (Aertsen & Johannesma, 1981 *b*) may differ from the STRF as determined for noise stimulation. The tonal stimulus ensembles as described in Section 13 can *never* result in a complete functional description of non-linear systems since only one tonepip at a time is present. Thus any complex interaction between simultaneously present frequencies (as in noise) such as two-tone interaction or lateral suppression will remain undetected. Therefore statistically structured double and multiple tone combinations could be introduced, ultimately resulting in a sum-of-sinusoid method as used in vision research (Victor & Knight, 1979; Spekreijse & Reits, 1982). Also, for more complex stimulus ensembles, like the acoustic biotope, a formal way to arrive at a stimulus invariant STRF was attempted. Heuristically, the method runs as follows (Aertsen & Johannesma, 1981 *b*): first of all the difference spectrogram is determined by subtracting the average dynamic spectrogram of the overall stimulus ensemble from the average dynamic spectrogram of the pre-event stimulus ensemble. Secondly, for each individual channel of the dynamic spectrum analyser an equalization procedure is carried out to account for the overall spectral differences between the stimulus ensemble and gaussian white noise. Thirdly, a correction for

the inherent temporal structure of the natural sounds is performed by dividing in the frequency domain the equalized difference spectrogram by the appropriately shifted spectrum of the natural sound ensemble. Using this procedure, normalized tonal spectro-temporal sensitivities could be compared with the normalized STS as determined with natural sounds for neural units from the midbrain of the grassfrog. The results pointed to considerable differences in both estimates of the STRF of the neural unit. Several factors could account for these differences: the low complexity of the tonal ensemble as compared to the natural sound ensemble, the different degree of adaptation of the auditory system in both cases, and the assumption that there are no higher-order even non-linearities than order two. Nevertheless, the procedures outlined above might lead to tentative ideas about stimulus invariance of the second-order neural characteristic, the STRF.

If a stimulus invariant characteristic can be obtained, as might be the case for short latency units in the torus semicircularis of the grassfrog (Johannesma & Eggermont, 1983), the next item to be investigated centres around the question: is the neural characterization complete?

The question of completeness is also interesting when the STRF is not stimulus invariant, i.e. it is interesting when staying within the class of stimuli it was determined with. Obviously the 'prediction' in that case has to remain within that class.

(*b*) *Completeness of the STRF*

Since for neurons in the torus semicircularis of the grassfrog the first-order neural characteristic (the reverse correlation function or first-order Wiener kernel) can only be obtained for best frequencies below 350 Hz and even then is quite noisy, the only practically useful characteristic is the STRF. We will examine in how far the STRF characterizes the neural properties.

For twelve neurons Eggermont *et al.* (1983*a*) investigated whether the STRF as determined for wide-band noise as a stimulus can be used to calculate the response probability to a given segment of the noise in comparison to an experimentally determined PST-histogram. The calculated response will be called the STRF-based response. The procedure basically consists of a convolution of the STRF of a neuron with the dynamic spectrum of the segment of noise. Thus, the degree of matching between the dynamic spectrum and the STRF determines the firing probability for the neuron.

For easiness of computation both the STRF and the dynamic spectrum were determined with the dynamic spectrum analyser (cf. Section 9). If the power in the one-third-octave band around frequency f_k is denoted by $P_k(t)$ and the average level by P_{k0}, the measured PST-histogram by $p(t)$ and the STRF of the neuron by $\{h_k(t)\}$, then the STRF of the neuron is determined by

$$\hat{h}_k(\omega) = \frac{\hat{R}_{kp}(\omega)}{\hat{R}_{kk}(\omega)} \quad (k = 1,\ldots,K), \tag{17.1}$$

where

$$\hat{R}_{kp}(\omega) = \hat{I}_k^*(\omega)\,\hat{p}(\omega), \tag{17.2}$$

$$\hat{R}_{kk}(\omega) = \hat{I}_k^*(\omega)\,\hat{I}_k(\omega), \tag{17.3}$$

$$\hat{R}_{kl}(\omega) = 0 \quad (k \neq l) \tag{17.4}$$

and

$$\hat{I}_k(\omega) = \hat{P}_k(\omega) - P_{k0}. \tag{17.5}$$

Assume that $p(t)$ can be decomposed in three mutually orthogonal components

$$p(t) = p_0 + p_1(t) + \epsilon(t), \tag{17.6}$$

where p_0 is the average firing rate to the segment of noise, $p_1(t)$ is the component in the PST histogram that is linearly related to $\{h_k(t)\}$, and $\epsilon(t)$ is any activity due to higher-order neural characteristics.

In order to test if $\{h_k(t)\}$ is a complete characteristic, i.e. all higher-order characteristics are zero, the STRF-based response

$$\tilde{p}(t) = p_0 + p_1(t) \tag{17.7}$$

is calculated with $p_1(t)$ following from

$$\hat{p}_1(\omega) = \sum_{k=1}^{K} \hat{h}_k(\omega)\,\hat{I}'_k(\omega), \tag{17.8}$$

where $\hat{I}'_k(\omega)$ relates to the segment of noise for which the PST histogram is obtained but defined according to equation (17.5).

For twelve neurons $\tilde{p}(t)$ was compared to $p(t)$ and the similarity expressed by the correlation coefficient. The correlation coefficient was generally not very high (0·2–0·7) and stayed below 0·5 in nine neurons. This indicates that a linear characterization on basis of a second-order neural characteristic is insufficient. An example of a comparison between $\tilde{p}(t)$ and $p(t)$ is shown in Fig. 28(*a*).

By assuming in cascade with the present model either a linear or a quadratic rectifier with its threshold at zero intensity level, the results could be improved slightly: correlation coefficients were between 0·25 and 0·8, but only in one neuron below 0·5 (Fig. 28*b*, *c*).

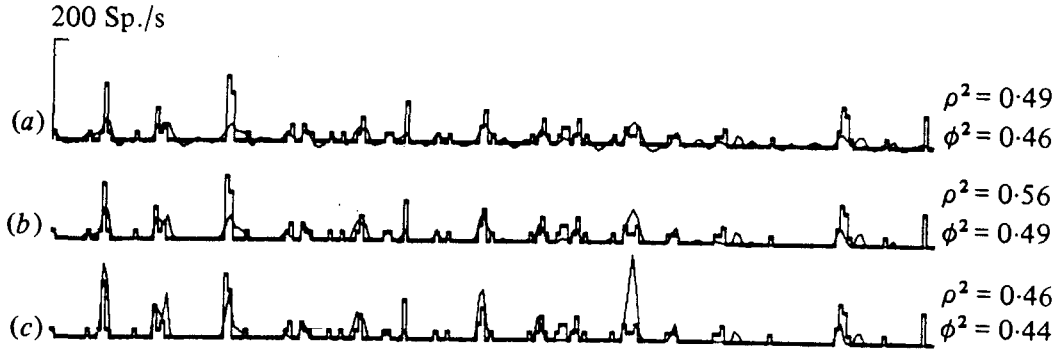


Fig. 28. For a unit from the torus semicircularis of the grassfrog the PST histogram to a 430 ms segment of pseudorandom noise is shown (thick staircase line) together with a linear reconstruction on basis of the STRF (*a*), with additional assumption of a linear rectifier stage (*b*) or a quadratic rectifier (*c*) shown as this continuous lines. The best correlation ($\rho^2 = 0{\cdot}56$) is obtained for the linear rectifier (from Eggermont *et al.* 1983).

This indicates that non-linear contributions (higher than second order) are quite important and besides that implies that the STRF is an incomplete characteristic in the majority of cases.

(*c*) *Predictability on basis of the STRF*

In the original sense (Erulker, Nelson & Bryan, 1968), predictability was based on the notion that 'the response of a single neuron to brief, spectrally and temporally simple, stimuli might be adequate to account for responses to stimuli in which complexity was increased along one or a few dimensions'. We may explore if the STRF of a neuron, after having obtained evidence about stimulus invariance as well as completeness, can be used to predict the neurons response to, for example, species specific vocalizations. Eggermont *et al.* (1983*b*) found that this in general is not the case. Again the application of a neural characteristic obtained in a stationary response situation to inherent non-stationary stimuli requires that a measure of the dynamic properties of the neuron has to be taken into account. Since it may be assumed that in first order the dynamics of the neuron (cf. Section 7) are independent of the spectro-temporal properties one could try to multiply the predicted response with an inverse-adaptation function. Vice versa one may obtain information about the neural dynamics from a comparison of the PST histogram to a vocalization with its STRF-based, predicted, response. An example of a prediction is shown in Fig. 29.

The findings of Eggermont *et al.* (1983*b*) only to a small extent corroborate Symmes (1981) remark that 'predictions of other sorts, such as from random noise responses to wide band vocalization

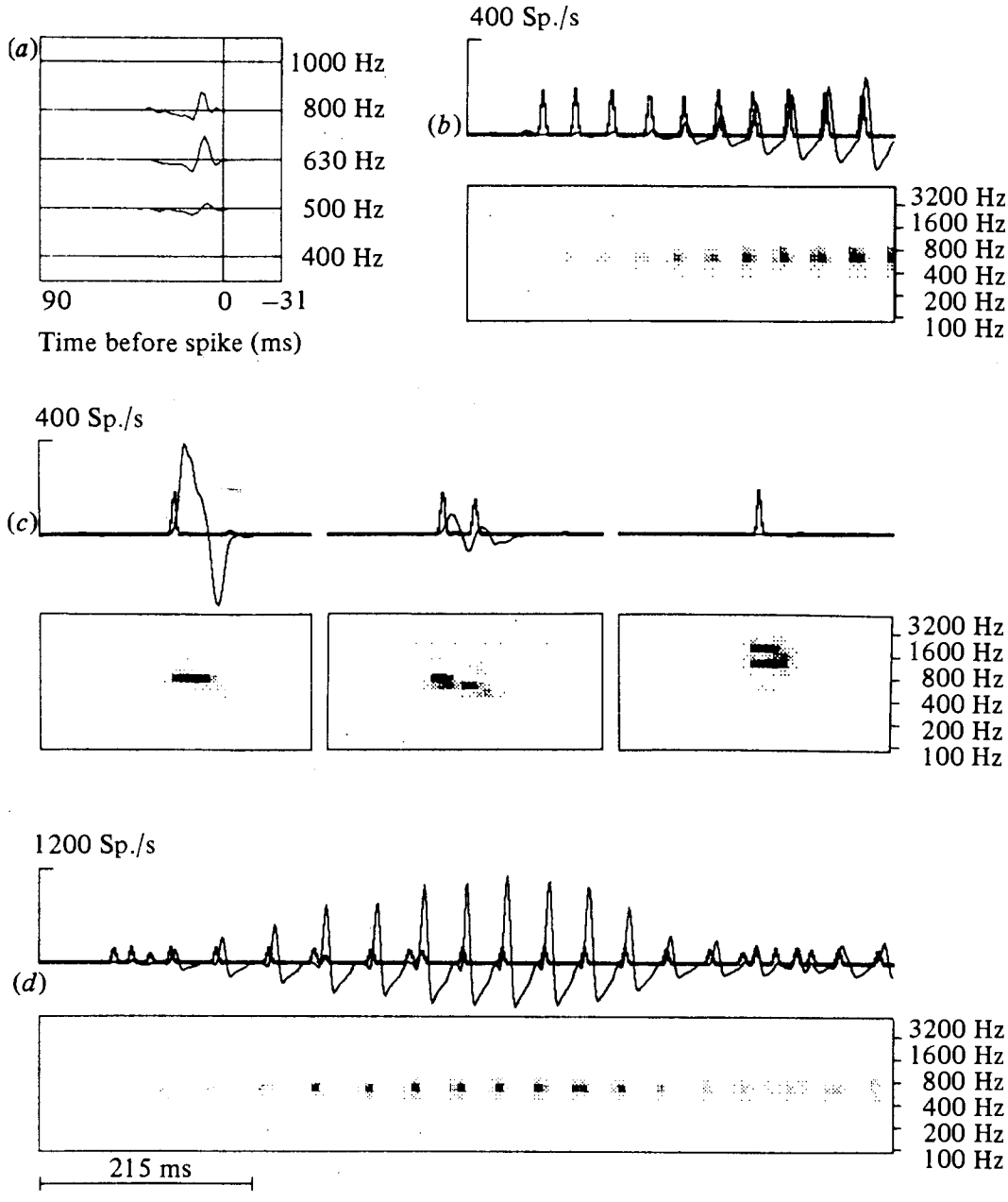


Fig. 29. Prediction of the response to natural sounds on basis of the STRF as obtained for noise. A linear prediction is carried out. In (*a*) the average pre-event sonogram for the noise stimulus is shown in three one-third-octave wide-frequency bands, this forms an estimate of the STRF. (*b*) The PST histogram and prediction are shown to a male vocalization (cf. the sonogram); while the prediction increases with increasing sound level in the vocalization the actual PST histogram is constant. (*c*) Predictions are shown to three female vocalizations (sonograms shown) and in (*d*) a prediction to another male vocalization is shown. Observe that the instants of firing are reasonably well predicted but that the amount of firing does not bear a relation to the amount predicted. (From Eggermont *et al.* 1983*b*.)

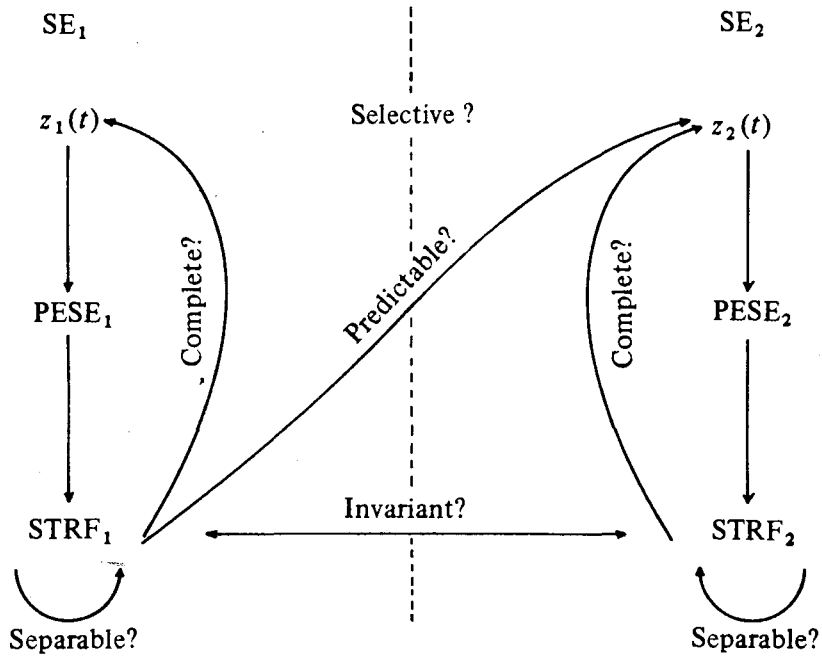


Fig. 30. Schematic representation of properties of the spectro-temporal receptive field of a single neuron. To different stimulus ensembles SE_1 and SE_2 different spike trains $z_1(t)$ and $z_2(t)$ are observed. From this basis two estimates of the spectro-temporal receptive field $STRF_1$ and $STRF_2$ are obtained. Both can be tested on separability and completeness (capability to estimate the original firing pattern). A comparison between both STRF's allows conclusions about stimulus invariance, and the use of $STRF_1$ to predict spike activity $z_j(t)$ $(i \neq j)$ gives further information about neuronal properties such as selectivity.

responses lack a clear logical foundation'. The logical foundation of response prediction is certainly present; all one has to be sure of is stimulus invariance and completeness of the neural characteristic besides full knowledge of the adaptation properties of the neuron.

The presently outlined method can be seen as an extension to the so-called tuning-curve predictability (e.g. Goldstein *et al.* 1971; Symmes, 1981). Predictability, implying invariance and completeness, is related to selectivity. Symmes (1981) notes that 'selective cells are cells which respond to a limited range of complex acoustic signals and have a level tolerant specificity'. This of course implies that their responses cannot be predicted and their spectro-temporal receptive fields will not be stimulus-invariant. An overview of the various measures on the STRF is presented in Fig. 30. Briefly it indicates that separability and completeness of the STRF only require the application of one stimulus ensemble. Invariance and predictability relate the responses to one stimulus ensemble to those of a completely different stimulus ensemble. Selectivity implies that responses can be obtained to only a few stimuli, for these neurons invariance and predictability are excluded.

(d) Stimulus reconstruction by a population of neurons

The prediction procedure outlined above forms a test for the invariance and completeness of the neural characteristic. Given a set of neurons with predictable response properties one may try to estimate how well these neurons as a functional group represent the actual stimulus. In other words, can one reconstruct from the response patterns of the neurons the stimulus that evoked them? Thereby it is tested whether there is a simple, complete coding of the stimulus properties on a given neural level. This problem has been addressed by Johannesma (1981) and Gielen & Johannesma (to be published). They simulated a population of 200 neurons from the auditory nerve of the cat with characteristic frequencies spaced equidistantly on a log scale from 200 Hz to 5 kHz using a neuron model. The first-order neural characteristics, or reverse correlation function, were obtained for gaussian wide-band noise as a stimulus. Then a test stimulus was presented and for each spike elicited the neuron's reverse correlation function was substituted. This was done for each neuron and after summation over all neurons an estimate of the presented stimulus, a sensory interpretation of the neural activity, was obtained. This reconstructed stimulus was then compared to the original test stimuli presented: a click, a tonepip, a sinewave and a cat vocalization (Fig. 31).

Qualitatively good results were obtained, but one wonders at what level in the auditory system this simple coding method breaks down, i.e. at what level the stimulus invariance of the neural characteristic disappears. Incidentally, the reconstruction method requires phase-lock and in this respect resembles the synchrony coding for speech sounds as proposed by Sachs & Young (1980). The method can be generalized to the construction of a so-called etho-acoustical space of an animal by forming the combination of the neuro-acoustic spaces associated with individual neurons (Johannesma & Aertsen, 1982). This combination is thereby determined by the coherent activity pattern of an interacting population of auditory neurons.

18. Estimating the form of the non-linearity: polynomial correlation

The Wiener–Volterra approach for the identification of non-linear systems requires the estimation of a large number of system kernels if the non-linearity is not simply of low order as we assumed. In

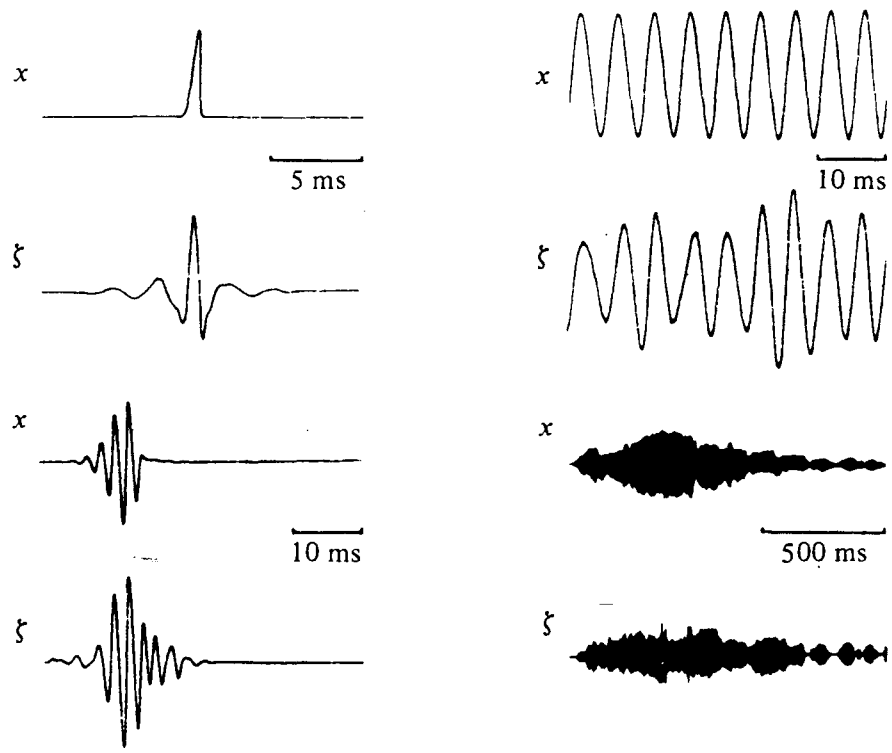


Fig. 31. Example of presented stimulus $x(t)$ and reconstructed stimulus $\xi(t)$ based on the neural activity pattern of a set of 200 independent model neurons. Results are shown for an impulse, a frequency modulated tone at 211–236 Hz, an amplitude modulated tone (tonepip) at 500 Hz and part of the vocalization of a cat. (From Johannesma, 1981.)

practice it is easy to calculate the first- and second-order Wiener kernels, it will be possible to calculate the third-order Wiener kernel, but unmanageable to calculate higher ones. When, as in the case of the auditory system, there is at least a third-order term, an approximation of the neural characteristics on basis of the first- and second-order Wiener kernels does not necessarily give better results than taking only the first-order kernel (Johnson, 1980). In addition one may ask whether the Wiener–Volterra approach for a highly adaptive system is able at all to reveal the character of the non-linearity. The drawback of model-free, very general, methods for system identification is clear: any prior information about the system (cf. Section 2) has to be identified again.

Model-based approaches, of course, are limited to that particular model. If the assumption turns out to be basically wrong, then any conclusion based thereupon can only be correct by accident. Although the basilar membrane hair-cell system works as an active non-linear filter (e.g. Sellick *et al.* 1982) it can often be approximated rather well by a sandwiched model consisting of two linear filters with an algebraic non-linearity in between (cf. Fig. 8). Such a model is also

called a BPNL network and was originally devised to account for the originally assumed broad basilar membrane tuning followed by a much sharper second filter (Pfeiffer, 1970; De Boer, 1976*b*). It is also possible to interpret it as a sharp cochlear filter followed by a non-linearity and a low-pass (synaptic) filter (cf. Section 5). It has been shown (De Boer, 1976*b*; Johannesma, 1971, 1972) that for such a model the (first-order) input–output cross-correlation function does not reveal the presence of the non-linearity in its shape, but only in its size, due to the algebraic nature thereof.

As shown in Section 5, the input–output cross-correlation is proportional to the convolution of the impulse response of both linear filters (when using gaussian wide-band noise as a stimulus) and does not show the properties of the individual filters either.

If we consider the input to the BPNL network (Fig. 8) to be gaussian noise, $x(t)$, then the output of the first filter, $u(t)$, is also gaussian noise. The non-linearity gives an output $v(t) = \phi\{u(t)\}$. Since $u(t)$ is gaussian, $v(t)$ can be written as an expansion in polynomials that are orthogonal with respect to a gaussian signal. These polynomials are the Hermite polynomials $\mathrm{He}_n(x)$ based on a weighting function $e^{-x^2/2}$, corresponding to a gaussian probability density function with $\mu_x = 0$ and $\sigma_x^2 = 1$ (cf. Section 16):

$$\mathrm{He}_n(x) \exp\left(-\frac{x^2}{2}\right) = (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n} \exp\left(-\frac{x^2}{2}\right), \tag{18.1}$$

where $x(t)$ is assumed normally distributed. The first five polynomials read

$$\begin{aligned} &\mathrm{He}_0(x) = 1, \quad \mathrm{He}_1(x) = x, \quad \mathrm{He}_2(x) = x^2 - 1, \\ &\mathrm{He}_3(x) = x^3 - 3x, \quad \mathrm{He}_4(x) = x^4 - 6x^2 + 3. \end{aligned} \tag{18.2}$$

The orthogonality relation is a particular one since the inner product now involves a weighting function:

$$\int \mathrm{He}_r(x)\, \mathrm{He}_s(x) \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = r!\, \delta_{rs}. \tag{18.3}$$

The output $v(t)$ of the non-linearity can therefore be written as a linear combination of these Hermite polynomials in such a way that any new polynomial added reduces the mean square error in the approximation of the non-linearity (De Boer, 1979; Lammers & De Boer, 1979):

$$v(t) = \sum_{n=0}^{\infty} \frac{a_n}{n!} \mathrm{He}_n(u') \tag{18.4}$$

with $u' = [u(t) - \bar{u}]/\sigma_u$ is standard normally distributed.

A special property of Hermite polynomials which are orthogonal to two gaussian stimulus ensembles $\chi_1(t)$ and $\chi_2(t)$ which both are standard normally distributed is that (Billings & Fakhouri, 1978)

$$E[\mathrm{He}_n(\chi_1)\,\mathrm{He}_m(\chi_2)] = n!\,\delta_{nm}\rho_{12}^n, \tag{18.5}$$

where ρ_{12} is the correlation coefficient between $\chi_1(t)$ and $\chi_2(t)$. Using this property to calculate the cross-correlation between $v(t)$ and $\mathrm{He}_n(x)$,

$$R_{\mathrm{He}_n(x)\,v}(\tau) = E[v(t)\,\mathrm{He}_n(x(t-\tau))], \tag{18.6}$$

results by substitution of equation (18.4) in

$$R_{\mathrm{He}_n(x)\,v}(\tau) = a_n\rho_{xu'}^n(\tau) = \frac{a_n}{\sigma_u^n}R_{xu}^n(\tau). \tag{18.7}$$

Note that

$$\frac{a_n}{\sigma_u^n} = E\left[\frac{\mathrm{d}^n v(t)}{\mathrm{d}u^n}\right] = K_n. \tag{18.8}$$

To obtain the desired cross-correlation function $R_{\mathrm{He}_n(x)y}(\tau)$ we convolute $K_n R_{xu}^n(\tau)$ with the impulse response of the second linear filter $k(\tau)$:

$$R_{\mathrm{He}_n(x)\,y}(\tau) = K_n R_{xu}^n(\tau) * k(\tau). \tag{18.9}$$

If $x(t)$ is gaussian noise, then $R_{xu}(\tau)$ is proportional to $h(\tau)$, the impulse response of the first linear filter (cf. Section 3). Since the non-linearity is determined by the various K_n we theoretically should evaluate all cross-correlations $R_{\mathrm{He}_n(x)\,y}(\tau)$. Fortunately one needs only to determine two cross-correlations that are different from zero (De Boer, 1979), e.g.

$$R_{\mathrm{He}_1(x)\,y}(\tau) = K_1 h(\tau) * k(\tau) \tag{18.10}$$

and

$$R_{\mathrm{He}_2(x)\,y}(\tau) = K_2 h^2(\tau) * k(\tau). \tag{18.11}$$

Note that these are first-order cross-correlations: the first one is the classical reverse correlation (equations 5.3 and 5.5) and the second one compares to intensity correlation. In case there is no first-order correlation function one needs to calculate a higher-order term.

In Fig. 32 we show four cross-correlations $R_{\mathrm{He}_n(x)\,z}(\tau)$ ($n = 1, 2, 3, 4$) calculated on basis of the spike trains of two cochlear nucleus units and one for random triggers. The selected units are the same ones as in Figs. 18 and 21. One of the units did not show a reverse correlation function, both show clearly discernible higher-order polynomial correlation functions (Aertsen *et al.*, to be published). These polynomial correlation functions can be seen as proportional to the diagonal functions of the corresponding Wiener kernels.

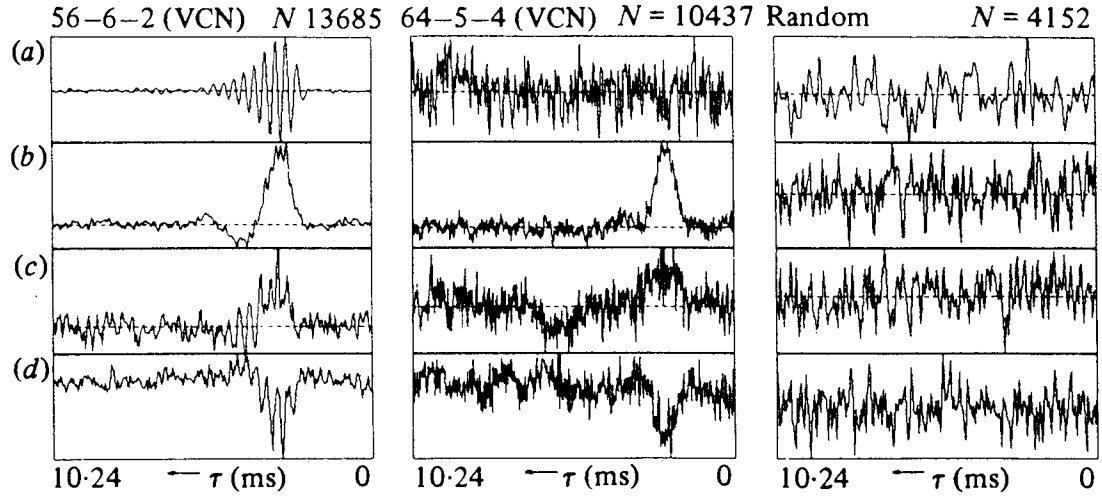


Fig. 32. Polynomial correlation for two neurons from the VCN in the cat. Shown are correlations of the spike train with polynomial functions of the pre-event stimulus. In (*a*) correlation is with $x(t)$, the signal itself resulting in the reverse correlation function in case there is phase-lock. In (*b*) correlation is with $x^2(t)-1$, resulting in a type of temporal intensity (cf. Fig. 18 for the same neurons). (*c*) Correlation with $x^3(t)-3x(t)$, which is quite clear for the unit in the left-hand column, which also has a reverse correlation function, but very close to background fluctations for the other neuron. In (*d*) the result of correlating with $x^4(t)-6x^2(t)+3$ is shown; this again is clearly present in both neurons. Note that the various polynomials are orthogonal, therefore very little leakage of lower-order correlations is expected to the higher-order ones provided that the statistical properties of the noise are adequate. In the right-hand column control results for spontaneous activity of one of the units are shown.

Billings & Falhouri (1978) have devised a method to determine $h(\tau)$ and $k(\tau)$ from any two known cross-correlations $R_{\mathrm{He}_n(x)\,y}(\tau)$ and $R_{\mathrm{He}_m(x)\,y}(\tau)$. Recall that for a proportional pulse generator (cf. equation 5.6)

$$R_{\mathrm{He}_n(x)\,y}(\tau) = R_{\mathrm{He}_n(x)\,z}(\tau). \qquad (18.12)$$

A small problem arises when the low-pass filter $k(\tau)$ completely overlaps the band-pass filter $h(\tau)$ as is the case for auditory neurons with low characteristic frequency. In that case $h(\tau) * k(\tau)$ is approximately equal to $h(\tau)$. Then the non-linearity that is estimated need not be the static non-linearity of the basilar membrane hair-cell system but actually that of the pulse-generating neuron. It is therefore more plausible to study those neurons that have high CF and therefore either incomplete phase-lock or no phase-lock at all. In the latter case one expects

$$R_{\mathrm{He}_1(x)\,y}(\tau) = R_{\mathrm{He}_3(x)\,y}(\tau) = R_{\mathrm{He}_{\mathrm{odd}}(x)\,y}(\tau) \simeq 0, \qquad (18.13)$$

as can be seen in the case for the non-phase-lock neuron shown in Fig. 32 (note the scale). The results based on this method are still preliminary but suggest that the method might work. The basic

problem is that of approximating an inherently non-linear filter by a sandwich model and lies in the assumption that this accounts for most if not all of the manifestations of the non-linearity. As such it presents a good example of both the advantages and the disadvantages of a model-based approach as indicated earlier.

19. Future outlook and conclusions

In the previous pages we have reviewed the results obtained with the various reverse correlation techniques and have shown that for a highly non-linear system as the auditory system, so far not a single measurement specifies the system completely. Reverse correlation methods are part of Wiener-like identification techniques and as such aim to describe the system completely on the basis of a single gaussian white-noise stimulus application by the (finite) set of kernels computed therefrom. So far only first- and second-order Wiener kernels (or equivalently the CoSTID) have been determined and on this basis alone successful predictions of the response to different inputs proved far from reached. This, however, seems inherent to Wiener kernels determined by the cross-correlation methods and the functional expansion thus obtained does not necessarily converge for any given input function (Poggio, 1981). Therefore the test for completeness (Section 17*b*) proved to be far better than the prediction to species-specific vocalizations (Section 17*c*). In principle, one expects reasonable predictions for stimuli that resemble the gaussian white-noise stimulus for which the kernels are determined.

While prediction on the basis of the CoSTID alone, in the case of nearly zero or zero-reverse correlation functions, seems to work reasonably if applied to noise stimuli, in the case where there is a clear reverse correlation function the prediction on the basis thereof is only slightly less than when the second-order Wiener kernel is taken into account (Wickesberg *et al.*, in preparation). Computational problems so far seem to exclude routine calculation of higher-order kernels; thus predictions will remain largely incomplete. Some advantage in that respect may be obtained from the polynomial correlation in which only the main diagonal terms of the various Wiener kernals are computed. In principle the method results in estimating the shape of the non-linearity by evaluation of the various K_n terms (equation 18.8).

One of the simplifications that are underlying most of the interpretations is that the peripheral auditory system can be approximated

by a sandwich system (cf. Fig. 8). This is possible only if the system is analytic or smooth enough to represent it by a Volterra-like form and when the system can be described with a discrete time-parameter (Palm, 1978). If the system can be characterized by a Volterra representation then the Wiener expansion is also valid.

Wiener-like methods depend on the mean value and power of the gaussian noise-input signal. Identification of the auditory system therefore will generally require kernel estimates for its various adaptation states. Wiener kernels determined for a given noise level and corresponding adaptation state are therefore of little value in predicting responses to other stimuli (e.g. vocalizations) which cause the system parameters to be different.

A method is needed to describe the time evolution of the CoSTID or Wiener kernels for a given noise level as a function of time after noise onset. Experiments using 4 s noise followed by 8 s silence applied to the grassfrog auditory system are currently in progress, and algorithms to compute the STRF as a function of time after noise onset are in development in our laboratory.

A further problem that we encountered is that the cross-correlation methods did not seem to reveal the non-linearity. In principle the bispectrum (the double Fourier transform of the second-order Wiener kernel) should reveal the contribution of frequencies ω_1 and ω_2 which are simultaneously present in the input signal (i.e. the noise) to the $\omega_1+\omega_2$ and $\omega_1-\omega_2$ components in the output signal. Wickesberg *et al.* (in preparation), however, did not observe clear manifestations of the difference tone $\omega_1-\omega_2$ in the bispectrum. We must, however, keep in mind that distortion products as $\omega_2-\omega_1$ and $2\omega_1-\omega_2$ have been described for double-tone stimulation and it is not completely clear if they will arise also for wide-band stimulation. First of all, for double tone stimulation, the tone frequencies are so selected that neither of them excites the neural unit but the difference tones do. In case of wide-band noise stimulation there will always be quite some activity at the CF of the neural unit, and the unit might well be near saturation for noise levels for which one expects a sufficient level for the combination tones. In this case the unit cannot respond any more to the much weaker distortion products One can also say that a wide-band noise stimulus linearizes. A way out of this problem might be to use high-pass noise that indirectly, or only very weakly, excites the neural unit. However, single auditory nerve-fibre data by Evans & Elberling (1982), who used high-pass noise masking, do not substantiate this idea.

Having indicated the obvious limitations of the white-noise approach (cf. Johnson, 1980), the extension of the reverse correlation technique to other stimulus ensembles, such as random tones or natural sounds, allows a comparison of spectro-temporal sensitivities of auditory neurons for a wide class of stimulus conditions. The concepts of invariance of the spectro-temporal sensitivity and predictability will be useful for revealing specific neural properties. Invariance assumes that a given kernel, or equivalently the STRF, is the same whatever stimulus is applied. Predictability combines invariance and completeness of the neural characterization. Predictability excludes selectivity and may be one of the crucial properties to test in the search for feature-extracting or feature-detecting neurons. Highly selective neurons will show a level tolerance in their response properties, therefore the demonstration of predictability at one stimulus level will generally be sufficient to exclude this selectivity.

In conclusion, the various reverse correlation methods give the same linearized description of the auditory nervous system as conventional methods using clicks and tones do. In addition, the reverse correlation technique offers methods to test how far the neuron is or can be characterized. The problem of phase-lock can be circumvented by the use of a second-order polynomial correlation or by the IFT or dynamic spectrum methods. It appears that auditory neurons behave quite differently under wide-band noise stimulation than for tonal stimulation, which is a consequence of their non-linear properties. Therefore reverse correlation methods for noise, tones as well as natural sounds at the moment offer the widest possible characterization for a given neuron. The method is, however, to be combined with all prior knowledge about the system. Only model-based approaches will reveal the basic biophysical mechanisms and circuitry of the auditory system.

References

Aertsen, A. M. H. J. & Johannesma, P. I. M. (1980). Spectro-temporal receptive field of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli. *Biol. Cybernet.* **38**, 223–234.

Aertsen, A. M. H. J. & Johannesma, P. I. M. (1981*a*). The spectro-temporal receptive field. A functional characterization of auditory neurons. *Biol. Cybernet.* **42**, 133–143.

Aertsen, A. M. H. J. & Johannesma, P. I. M. (1981*b*). A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol. Cybernet* **42**, 145–156.

Aertsen, A. M. H. J., Johannesma, P. I. M. & Hermes, D. J. (1980).

Spectro-temporal receptive fields of auditory neurons in the grassfrog. II. Analysis of the stimulus-event relation for tonal stimuli. *Biol. Cybernet.* **38**, 235–248.

AERTSEN, A. M. H. J., OLDERS, J. H. J. & JOHANNESMA, P. I. M. (1981). Spectro-temporal receptive fields of auditory neurons in the grassfrog. III. Analysis of the stimulus-event relation for natural stimuli. *Biol. Cybernet.* **39**, 195–209.

ANDERSON, D. J., ROSE, J. E., HIND, J. E. & BRUGGE, J. F. (1971). Temporal position of discharges in single auditory nerve fibres within the cycle of a sine-wave stimulus: Frequency and intensity effects. *J. acoust. Soc. Am.* **49**, 1131–1139.

ARTHUR, R. M. (1976). Harmonic analysis of two-tone discharge patterns in cochlear nerve fibers. *Biol. Cybernet.* **22**, 21–31.

BARRETT, J. F. (1963). The use of functionals in the analysis of non-linear physical systems. *J. Electron. Control.* **15**, 567–615.

BIBIKOV, N. & GORODETSKAYA, O. (1981). Coding of amplitude-modulated tones in the midbrain region of the frog. In *Neuronal Mechanisms of Hearing* (ed. J. Syka and L. Aitkin), pp. 347–352. New York: Plenum.

BILLINGS, S. A. & FAKHOURI, S. Y. (1978). Identification of a class of non linear systems using correlation analysis. *Proc. IEE* **125**, 691–697.

BOER, E. DE (1967). Correlation studies applied to the frequency resolution of the cochlea. *J. Aud. Res.* **7**, 209–217.

BOER, E. DE (1968). Reverse correlation. I. A heuristic introduction to the technique of triggered correlation with application to the analysis of compound systems. *Proc. K. ned. Akad. Wet.* C **71**, 472–486.

BOER, E. DE (1969). Reverse correlation. II. Initiation of nerve impulses in the inner ear. *Proc. K. ned. Akad. Wet.* C **72**, 129–151.

BOER, E. DE (1973). On the principle of specific coding. *J. Dynamic Systems, Meas. Contr.* **95** G, 265–273.

BOER, E. DE (1976*a*). On the residue and auditory pitch perception. In *Handbook of Sensory Physiology*, vol. v/3 (ed. W. D. Keidel and W. D. Neff), pp. 479–583. Berlin: Springer.

BOER, E. DE (1976*b*). Cross correlation function of a bandpass nonlinear network. *Proc. IEEE* **64**, 1443–1444.

BOER, E. DE (1979). Polynomial correlation. *Proc. IEEE* **67**, 317–318.

BOER, E. DE & JONGH, H. R. DE (1978). On cochlear encoding: potentialities and limitations of the reverse correlation technique. *J. acoust. Soc. Am.* **63**, 115–135.

BOER, E. DE & KUYPER, P. (1968). Triggered correlation. *IEEE Trans. Biomed. Eng.* BME-15, 169–179.

CAPRANICA, R. R. (1976). Morphology and physiology of the auditory system. In *Frog Neurobiology: A Handbook* (ed. R. Llinás and W. Precht), pp. 551–573. Berlin: Springer.

CAPRANICA, R. R. & MOFFAT, A. J. M. (1980). Non linear properties of the peripheral auditory system. In *Comparative Studies of Hearing in Vertebrates* (ed. A. N. Popper and R. R. Fay), pp. 139–165. Berlin: Springer.

ECKHORN, R. & PÖPEL, B. (1979). Generation of gaussian noise with improved quasi-white properties. *Biol. Cybernet.* **32**, 243–248.

EGGERMONT, J. J. (1973). Analog modelling of cochlear adaptation. *Kybernetik* **14**, 117–126.

EGGERMONT, J. J. (1975). Cochlear adaptation: a theoretical description. *Biol. Cybernet.* **19**, 181–190.

EGGERMONT, J. J., AERTSEN, A. M. H. J., HERMES, D. J. & JOHANNESMA, P. I. M. (1981). Spectro-temporal characterization of auditory neurons: redundant or necessary. *Hear. Res.* **5**, 109–121.

EGGERMONT, J. J., AERTSEN, A. M. H. J. & JOHANNESMA, P. I. M. (1983*a*). Quantitative characterization procedure for auditory neurons based on the spectro-temporal receptive field. *Hear. Res.* **10**, 167–190.

EGGERMONT, J. J., AERTSEN, A. M. H. J. & JOHANNESMA, P. I. M. (1983*b*). Prediction of responses of auditory neurons in the midbrain of the grassfrog based on the spectro-temporal receptive field. *Hear. Res.* **10**, 191–202

EGGERMONT, J. J., EPPING, W. J. M. & AERTSEN, A. M. H. J. (1983*c*). Binaural hearing and neural interaction. In *Hearing – Physiological Bases and Psychophysics* (ed. R. Klinke and R. Hartmann), pp. 237–242. Berlin: Springer.

EGGERMONT, J. J., EPPING, W. J. M. & AERTSEN, A. M. H. J. (1983*d*). Stimulus dependent neural correlations in the auditory mid-brain of the grassfrog (*Rana temporaria* L.). *Biol. Cybernet.* **47**, 103–117.

ERULKAR, S. D., NELSON, P. G. & BRYAN, J. S. (1968). Exprimental and theoretical approaches to neural processing in the central auditory pathway. In *Contributions to Sensory Physiology*, vol. 3 (ed. W. D. Neff), pp. 149–189. New York: Academic Press.

EVANS, E. F. (1974). Auditory frequency selectivity and the cochlear nerve. In *Facts and Models in Hearing* (ed. E. Zwicker and E. Terhardt), pp. 118–129. Berlin: Springer.

EVANS, E. F. (1977). Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus. In *Psychophysics and Physiology of Hearing* (ed. E. F. Evans and J. P. Wilson), pp. 185–192. London: Academic Press.

EVANS, E. F. & ELBERLING, C. (1982). Location-specific components of the gross cochlear action potential. *Audiology* **21**, 204–227.

EVANS, E. F. & WILSON, J. P. (1973). The frequency selectivity of the cochlea. In *Basic Mechanisms in Hearing* (ed. A. R. Møller), pp. 519–551. New York: Academic Press.

FENGLER, R. (1980). Reverse correlation: Anwendung und Ergebnisse des Verfahrens am peripheren Auditorischen System des Brillen Kaiman. Thesis, Freien Universität, Berlin.

FERNALD, R. D. & GERSTEIN, G. L. (1972). Response of cat cochlear nucleus neurones to frequency and amplitude modulated tones. *Brain Res.* **45**, 417–435.

FUZESSERY, Z. M. & FENG, A. S. (1982). Frequency selectivity in the anuran auditory midbrain: single unit responses to single and multiple tone stimulation. *J. comp. Physiol.* **146**, 471–484.

GABOR, D. (1946). Theory of communication. *J. IEE* **93**, 429–457.

GERSTEIN, G. L., BUTLER, R. A. & ERULKAR, S. D. (1968). Excitation and inhibition in cochlear nucleus. I. Tone-burst stimulation. *J. Neurophysiol.* **31**, 526–536.

GERSTEIN, G. L. & KIANG, N. Y. S. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophys. J.* **1**, 15–28.

GISBERGEN, J. A. M. VAN (1974). Characterization of responses to tone and noise stimuli of neurons in the cats cochlear nuclei. Thesis, Nijmegen.

GISBERGEN, J. A. M. VAN, GRASHUIS, J. L., JOHANNESMA, P. I. M. & VENDRIK, A. J. H. (1975*a*). Neurons in the cochlear nucleus investigated with tone and noise stimuli. *Expl Brain Res.* **23**, 387–406.

GISBERGEN, J. A. M. VAN, GRASHUIS, J. L., JOHANNESMA, P. I. M. & VENDRIK, A. J. H. (1975*b*). Statistical analysis and interpretation of the initial response of cochlear nucleus neurons to tone bursts. *Expl Brain Res.* **23**, 407–423.

GOBLICK, T. J. & PFEIFFER, R. R. (1969). Time-domain measurements of cochlear nonlinearities using combination click stimuli. *J. acoust. Soc. Am.* **46**, 924–938.

GOLDSTEIN, J. L., BAER, TH. & KIANG, N. Y. S. (1971). A theoretical treatment of latency, group delay, and tuning characteristics for auditory nerve responses to clicks and tones. In *Physiology of the Auditory System* (ed. M. B. Sachs), pp. 133–141. Baltimore: National Educational Consultants.

GOLDSTEIN, J. L. & KIANG, N. Y. S. (1968). Neural correlates of the aural combination tone $2f_1 - f_2$. *Proc. IEEE* **56**, 981–992.

GRASHUIS, J. L. (1974). The pre-event stimulus ensemble. An analysis of the stimulus response relation for complex stimuli applied to auditory neurons. Thesis, Nijmegen.

HARRIS, D. M. & DALLOS, P. (1979). Forward masking of auditory nerve fiber responses. *J. Neurophysiol.* **42**, 1083–1107.

HARRISON, R. V. & EVANS, E. F. (1982). Reverse correlation study of cochlear filtering in normal and pathological guinea pig ears. *Hear. Res.* **6**, 303–314.

HERMES, D. J., AERTSEN, A. M. H. J., JOHANNESMA, P. I. M. & EGGERMONT, J. J. (1981). Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grassfrog (*Rana temporaria* L.) investigated with noise stimuli. *Hear. Res.* **5**, 145–179.

HERMES, D. J., EGGERMONT, J. J., AERTSEN, A. M. H. J. & JOHANNESMA, P. I. M. (1982). Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grassfrog (*Rana temporaria* L.) investigated with tonal stimuli. *Hear. Res.* **6**, 103–126.

HEUSDEN, E. VAN & SMOORENBURG, G. F. (1983). Response from AVCN units in the cat before and after inducement of an acute noise trauma. *Hear. Res.* **11**, 295–326.

JOHANNESMA, P. I. M. (1971). Dynamical aspects of the transmission of stochastic neural signals. In *Proc. First European Biophysics Congress* (ed. E. Broda, A. Locker and H. Springer-Lederer), pp. 329–333. Vienna: Verlag der Wiener Medizinischen Akademie.

JOHANNESMA, P. I. M. (1972). The pre-response stimulus ensemble of neurons in the cochlear nucleus. In *Proc. of the IPO Symp. on Hearing Theory* (ed. B. L. Cardozo), pp. 58–69. Eindhoven: IPO.

JOHANNESMA, P. I. M. (1980). Functional identification of auditory neurons based on stimulus event correlation. In *Psychophysical, Physiological and Behavioral Studies in Hearing* (ed. G. van den Brink and F. A. Bilsen), pp. 77–84. Delft University Press.

JOHANNESMA, P. I. M. (1981). Neural representation of sensory interpetation of neural activity. In *Adv. Physiol. Sci.* vol. 30 (ed. G. Székely, E. Lábos and S. Damjanovich), pp. 103–125. Budapest: Akadémiai Kiadó.

JOHANNESMA, P. I. M. & AERTSEN, A. M. H. J. (1982). Statistical and dimensional analysis of the neural representation of the acoustic biotope of the frog. *J. Medical Systems* **6**, 399–421.

JOHANNESMA, P., AERTSEN, A., CRANEN, B. & ERNING, L. VAN (1981). The phonochrome: a coherent spectro-temporal representation of sound. *Hear. Res.* **5**, 123–145.

JOHANNESMA, P. & EGGERMONT, J. (1983). Receptive fields of auditory neurons in the midbrain of the frog as functional elements of acoustic communication. In *Advances in Vertebrate Neuroethology* (ed. J. P. Ewert, R. R. Capranica and D. J. Ingle) pp. 901–910. New York: Plenum.

JOHANNESMA, P. I. M., GISBERGEN, J. A. M. VAN, GRASHUIS, J. L. & VENDRIK, A. J. H. (1973). Forward and backward analysis of stimulus response relations of single cells in the cochlear nucleus. In *Proc. IVth Int. Biophys. Congr.* pp. 508–523. Moscow.

JOHNSON, D. H. (1980). Applicability of white-noise nonlinear system analysis to the peripheral auditory system. *J. acoust. Soc. Am.* **68**, 876–884.

JONGH, H. R. DE (1978). Modelling the peripheral auditory system. Thesis, Amsterdam.

KEMP, D. T. & CHUM, R. (1980). Properties of the generator of stimulated acoustic emissions. *Hear. Res.* **2**, 213–232.

KIANG, N. Y. S. & MOXON, E. C. (1972). Physiological considerations in artificial stimulation of the inner ear. *Ann. Otol. Rhinol. Laryngol.* **81**, 714–730.

KIANG, N. Y. S. & MOXON, E. C. (1974). Tails of tuning curves of auditory nerve fibers. *J. acoust. Soc. Am.* **55**, 620–630.

KIANG, N. Y. S., WATANABE, T., THOMAS, E. C. & CLARK, L. F. (1965). *Discharge Patterns of Single Fibers in the Cats Auditory Nerve.* Cambridge, Mass.: MIT Press.

KLINKE, R. & PAUSE, M. (1980). Discharge properties of primary auditory fibers in Caiman *Crocodillus*: comparisons and contrasts to the mammalian auditory nerve. *Expl Brain Res.* **38**, 137–150.

LAMMERS, H. C. & BOER, E. DE (1979). Regression function of a bandpass nonlinear (BPNL) network. *Proc. IEEE* **67**, 432–434.

LEE, Y. W. & SCHETZEN, M. (1965). Measurement of the Wiener kernels of a non-linear system by cross-correlation. *Int. J. Control* **2**, 237–254.

LEPPELSACK, H. J. & VOGT, M. (1976). Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *J. comp. Physiol.* **107**, 263–274.

LIBERMAN, M. C. (1978). Auditory nerve response from cats raised in a low-noise chamber. *J. acoust. Soc. Am.* **63**, 442–445.

MARMARELIS, P. Z. & MARMARELIS, V. Z. (1978). *Analysis of Physiological Systems. The White Noise Approach.* New York: Plenum.

MØLLER, A. R. (1973). Statistical evaluation of the dynamic properties of cochlear nucleus units using stimuli modulated with pseudorandom noise. *Brain Res.* **57**, 443–456.

MØLLER, A. R. (1975). Latency of unit-responses in cochlear nucleus determined in two different ways. *J. Neurophysiol.* **38**, 812–821.

MØLLER, A. R. (1976*a*). Dynamic properties of the response of single neurones in the cochlear nucleus of the rat. *J. Physiol.* **259**, 63–82.

MØLLER, A. R. (1976*b*). Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus. *Acta physiol. Scand.* **98**, 157–167.

MØLLER, A. R. (1977). Frequency selectivity of single auditory nerve fibers in response to broad band noise stimuli. *J. acoust. Soc.* **62**, 135–142.

MØLLER, A. R. (1978). Frequency selectivity of the peripheral auditory analyzer studies using broad band noise. *Acta physiol. Scand.* **104**, 24–32.

NARINS, P. M. (1983). Frequency selectivity in the inner ear of anuran amphibians. In *Hearing-Physiological Bases and Psychophysics* (ed. R. Klinke and R. Hartmann), pp. 70–75. Berlin: Springer.

PALM, G. (1978). On representation and approximation of non-linear systems. *Biol. Cybernet.* **31**, 119–124.

PFEIFFER, R. R. (1970). A model for two-tone inhibition of single cochlear nerve fibers. *J. acoust. Soc. Am.* **48**, 1373–1378.

PFEIFFER, R. R. & KIM, D. O. (1972). Response patterns of single cochlear nerve fibres to click stimuli: descriptions for cat. *J. acoust. Soc. Am.* **52**, 1669–1677.

POGGIO, T. (1981). Wiener like identification techniques. In *Theoretical Approaches in Neurobiology* (ed. W. E. Reichardt and T. Poggio), pp. 60–63. Cambridge: MIT Press.

POTTER, R. K., KOPP, G. A. & GREEN, H. C. (1947). *Visible Speech.* New York: Van Nostrand.

RIHACZEK, A. W. (1968). Signal energy distribution in time and frequency. *IEEE Trans. Inf. Theory* **14**, 369–374.

ROSE, J. E., BRUGGE, J. F., ANDERSON, D. J. & HIND, J. E. (1967). Phase locked responses to low frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* **30**, 769–793.

ROSE, J. E., HIND, J.E., ANDERSON, D. J. & BRUGGE, J. F. (1971). Some effects of stimulus intensity in response of auditory nerve fibers in the squirrel monkey. *J. Neurophysiol.* **34**, 685–699.

SACHS, M. B. & YOUNG, E. D. (1980). Effects of non-linearities on speech encoding in the auditory nerve. *J. acoust. Soc. Am.* **68**, 858–875.

SEGUNDO, J. P. (1970). Communications and coding by nerve cells. In *The*

Neurosciences, Second Study Proram (ed. F. O. Schmitt), pp. 569–586. New York: Rockefeller University Press.

SELLICK, P. M., PATUZZI, R. & JOHNSTONE, B. M. (1982). Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. *J. acoust. Soc. Am.* **72**, 131–141.

SELLICK, P. M. & RUSSELL, I. J. (1979). Two tone suppression in cochlear hair cells. *Hear. Res.* **1**, 227–236.

SMITH, R. L. (1979). Adaptation, saturation and physiological masking in single auditory nerve fibers. *J. acoust. Soc. Am.* **65**, 166–178.

SMOLDERS, J. W. T., AERTSEN, A. M. H. J. & JOHANNESMA, P. I. M. (1979). Neural representation of the acoustic biotope, a comparison of the response of auditory neurons to tonal and natural stimuli in the cat. *Biol. Cybernet.* **35**, 11–20.

SPEKREIJSE, H. & REITS, D. (1982). Sequential analysis of the visual evoked potential system in man; nonlinear analysis in a sandwich system. *Ann. N.Y. Acad. Sci.* **388**, 72–97.

SWERUP, C. (1978). On the choice of noise for the analysis of the peripheral auditory system. *Biol. Cybernet.* **29**, 97–104.

SYMMES, D. (1981). On the use of natural stimuli in neurophysiological studies of audition. *Hear. Res.* **4**, 203–214.

VICTOR, J. D. & KNIGHT, B. W. (1979). Nonlinear analysis with an arbitrary stimulus ensemble. *Q. appl. Math.* **37**, 113–136.

WALKOWIAK, W. (1980). The coding of auditory signals in the torus semicircularis of the fire-bellied toad and the grassfrog: responses to simple stimuli and conspecific calls. *J. comp. Physiol.* **138**, 131–148.

WEBSTER, W. R. & AITKIN, L. M. (1975). Central auditory processing. In *Handbook of Psychobiology* (ed. M. S. Gazzaniga and C. Blakemore), pp. 325–364. New York: Academic Press.

WICKESBERG, R. E., DICKSON, J. W., GIBSON, M. M. & GEISLER, C. D. Wiener kernel analysis of the responses of anteroventral cochlear nucleus neurons in the cat. (Manuscript.)

WILSON, J. P. & EVANS, E. F. (1975). Systematic error in some methods of reverse correlation. *J. acoust. Soc. Am.* **57**, 215–216.

YOUNG, E. D. & SACHS, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *J. acoust. Soc. Am.* **66**, 1381–1403.