

Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field

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

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

(Received 9 March 1982; accepted 1 December 1982)

The spectro-temporal receptive field (STRF) of an auditory neuron represents those characteristics of the sound stimulus in both the time and frequency domain that affect the firing probability of the neuron. The STRF is determined under stationary stimulus conditions for Gaussian wide-band noise. It has been demonstrated that for some neurons the response to that noise could to a considerable extent be derived from the STRF. In the present study the usefulness of the STRF is tested to predict responses to other stimuli such as noise with different frequency content and to species-specific vocalisations. It appears that the predicted response to vocalisations is at best in qualitative agreement with the actual response.

Key words: frog; torus semicircularis; species-specific vocalizations; neural response prediction.

Introduction

The spectro-temporal receptive field (STRF) has been introduced as a representation of the average stimulus properties in the spectro-temporal domain that influence the firing probability of the neuron [3,4]. The STRF is determined under stimulation with pseudorandom wide-band noise. In the accompanying paper [3] we investigated for this stimulus to what extent auditory neurons in the midbrain of the grass frog are quantitatively characterised by their STRFs. It appeared that for some neurons over half of their responses could be derived from the STRF.

The usefulness of the STRF would increase considerably if it could also be used to predict the response to other kinds of stimuli than the noise for which the STRF was determined. This was investigated in the present study. First, the response was predicted to noise different from the noise with which the STRF was determined. Second, response prediction was carried out for a selection of species-specific vocalisations.

Methods

In the accompanying paper [3] twelve neurons in the midbrain of the grass frog, *Rana temporaria* L., were quantitatively characterised by their STRFs which were

obtained under stimulation with Gaussian wide-band noise. In this study these STRFs were for several neurons used to predict the response to a different stimulus. For one of these neurons the STRF was used to predict the response to a segment of pseudorandom noise with a much shorter sequence length than that of the noise for which the STRF was determined. For two units the STRF was used to predict the response to pseudorandom noise with identical sequence length but with different bandwidth and/or average stimulus intensity. For seven neurons enough response data were available to allow the response prediction to species-specific vocalisations.

The stimuli

The STRFs of the neurons were determined under stimulation with pseudorandom wide-band noise generated by low-pass filtering a binary sequence of 1 048 575 steps at 500 Hz, 1500 Hz and 5000 Hz (-3 dB). These noise types will be indicated by < 500 , < 1500 and < 5000 Hz noise, respectively. Their overall intensity was 94 dB SPL.

The vocalisations used for the prediction were recorded under seminatural conditions. Adult grass frogs from Ireland were transported to our laboratory immediately after hibernation. Kept at about 13°C they readily started vocalising. Vocalisations of the male grass frog consist of sequences of tone pips separated by some tens of milliseconds [1,9]. Within one call the form of these tone pips remains conspicuously constant. Most energy is contained between 400 and 750 Hz. In most cases each tone pip also contains higher frequency components without a harmonic relation to the fundamental. The sonograms of some male vocalisations are shown in Fig. 4b, d. Female frogs only vocalise when a mating attempt is made while they are not receptive [2]. These 'release calls' have a tonal or harmonic structure and last about 100 ms. Some frequency modulation can often be observed and most energy is contained in frequency bands above 500 Hz, in many cases above 1 kHz. The sonograms of three female vocalisations are shown in Fig. 4c. The last one was evoked from a non-receptive female by pushing her behind the front legs with thumb and finger. In most cases these artificially evoked 'release calls' had a much more irregular, sometimes pulse-like structure. The shown call was one of the most regular ones. The peak equivalent SPLs of the vocalisations presented in Fig. 4b–d were 84, 86, 78, 86 and 89 dB, respectively.

The prediction procedure

Notations and use of symbols are derived from those in the accompanying paper [3]. The STRF was determined under stimulation with pseudorandom noise according to

$$R_{kp}(\tau) = \int_{-\infty}^{\infty} h_k(\sigma) R_{kk}(\tau - \sigma) d\sigma \quad (1)$$

The noise for which the STRF was determined, will be called *STRF-noise*. The *predicted response* to another stimulus is defined according to

$$\hat{p}(t) = p_o + \sum_{k=1}^K \int_{-\infty}^{\infty} h_k(\sigma) I_k(t - \sigma) d\sigma \quad (2)$$

TABLE I
PROPERTIES OF STRF-NOISE AND PSTH-NOISE

	Bandwidth	Overall intensity	Sequence length
Unit 175-2			
STRF-noise	1500 Hz	94 dB SPL	1048575
PSTH-noise	1500 Hz	94 dB SPL	131071
Unit 166-9			
STRF-noise	1500 Hz	94 dB SPL	1048575
PSTH-noise	5000 Hz	94 dB SPL	1048575
Unit 169-3			
STRF-noise	500 Hz	94 dB SPL	1048575
PSTH-noise	1500 Hz	99 dB SPL	1048575

in which p_o and $I_k(t)$ are given the values they have for the stimulus to which the response was predicted. This predicted response is compared with the *actual response*, $p(t)$, the peristimulus time histogram (PSTH) to a sequence of noise or a species-specific vocalisation. When the prediction was carried out for pseudorandom noise, this noise will be called *PSTH-noise*. Properties of STRF-noise and PSTH-noise are summarised in Table I.

For one neuron, unit 175-2, the bandwidth and overall stimulus intensity of STRF-noise and PSTH-noise were the same, but the PSTH-noise was generated by a much shorter sequence than the STRF-noise. The correlation coefficient of $\hat{p}(t)$ and $p(t)$, $\rho_{\hat{p}p}$, will be used as a measure of similarity between $\hat{p}(t)$ and $p(t)$. This measure will be compared with that obtained for the characterisation under stimulation with the STRF-noise.

For two neurons, unit 166-9 and 169-3, the STRF-noise and the PSTH-noise had the same sequence length, but differed in bandwidth and/or overall stimulus intensity. Again, $\rho_{\hat{p}p}$ will be presented as a measure of similarity between $\hat{p}(t)$ and $p(t)$. Now, they will be compared with the values they had for the characterisation under stimulation with PSTH-noise.

For seven neurons the response was predicted to a number of species-specific vocalisations. Because the vocalisations were preceded by relatively long silent periods, a few hundred milliseconds, and spontaneous activity was insignificant, p_o was set to zero. No measure of similarity will be presented because this does not have any significance under these nonstationary stimulus conditions.

All results will be shown after applying a smoothing procedure that brought the spectral content of predicted and actual response in line with each other [3].

Results

The results of unit 175-2 are presented in Fig. 1. The STRF that showed that the neuron was maximally sensitive in the 630-Hz band was determined for < 1500-Hz

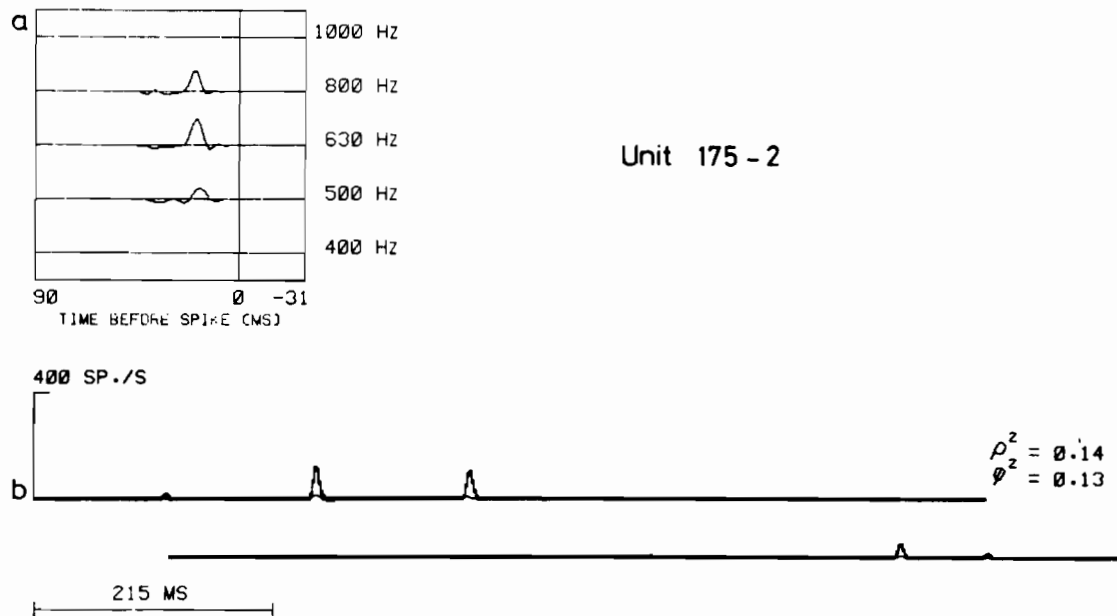


Fig. 1. Prediction of the response of unit 175-2 to a segment of noise generated by a shorter pseudorandom sequence than the noise with which the spectro-temporal receptive field (STRF) was determined. The STRF is shown in a. The thin line in b shows the predicted response, while the thick, staircase, plot shows the actual response.

noise with a sequence length of 1 048 575 steps. The predicted response to a 1720.32 ms segment of < 1500 Hz noise with a much shorter sequence length of 131 071 steps is shown in Fig. 1b (thin line). The thick, staircase, plot in Fig. 1b is the actual response. The measure of similarity was 0.37. For the characterisation under stimulation with STRF-noise this was 0.35, which is about equal.

For unit 166-9 the results are presented in Fig. 2. This neuron was best sensitive to frequencies in the 1250 Hz band under stimulation with < 1500 Hz noise (Fig. 2a) as well as under stimulation with < 5000 Hz noise (Fig. 2b). Both kinds of noise had the same sequence lengths and overall stimulus intensities. The STRF for < 1500 Hz noise was used to predict the response to < 5000 Hz noise, the result of which is shown in Fig. 2c. The measure of similarity was 0.67. When the unit was characterised under < 5000 Hz noise, this was 0.70. The results of this characterisation are shown in Fig. 2d. It is observed that the predicted response based on the STRF of < 1500 Hz noise (Fig. 2c) and the response based on the STRF of < 5000 Hz noise differ mainly for a constant factor.

The third neuron, unit 169-3, was most sensitive in the 320 Hz band according to the STRF determined for < 500 Hz noise with an overall stimulus intensity of 94 dB SPL (Fig. 3a). The PSTH-noise was < 1500 Hz noise with an overall intensity of 99 dB SPL. So, in the frequency ranges below 500 Hz the average intensities per third octave differed only 0.23 dB. Under stimulation with this PSTH-noise this neuron, which showed lateral suppression [5], appeared to be most sensitive in the 250 Hz band (Fig. 3b). The measure of similarity of $p(t)$ and $\hat{p}(t)$ was 0.62 for this

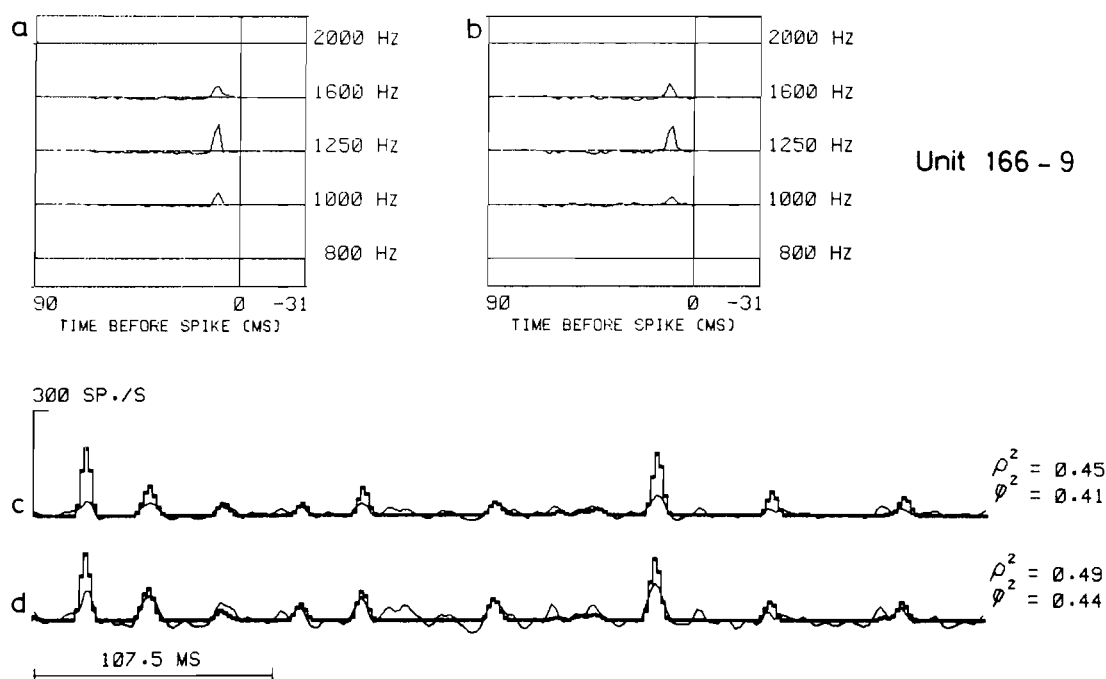


Fig. 2. The results of the response prediction and the characterisation of unit 166-9. The STRF under stimulation with < 1500 Hz noise is shown in a; the STRF under stimulation with < 5000 Hz noise is shown in b. The STRF for < 1500 Hz noise was used to predict the response to < 5000 Hz noise, the results of which are shown in c. The thin line shows the predicted response, the thick line the actual response. The results of the characterisation for < 5000 Hz noise are shown in d.

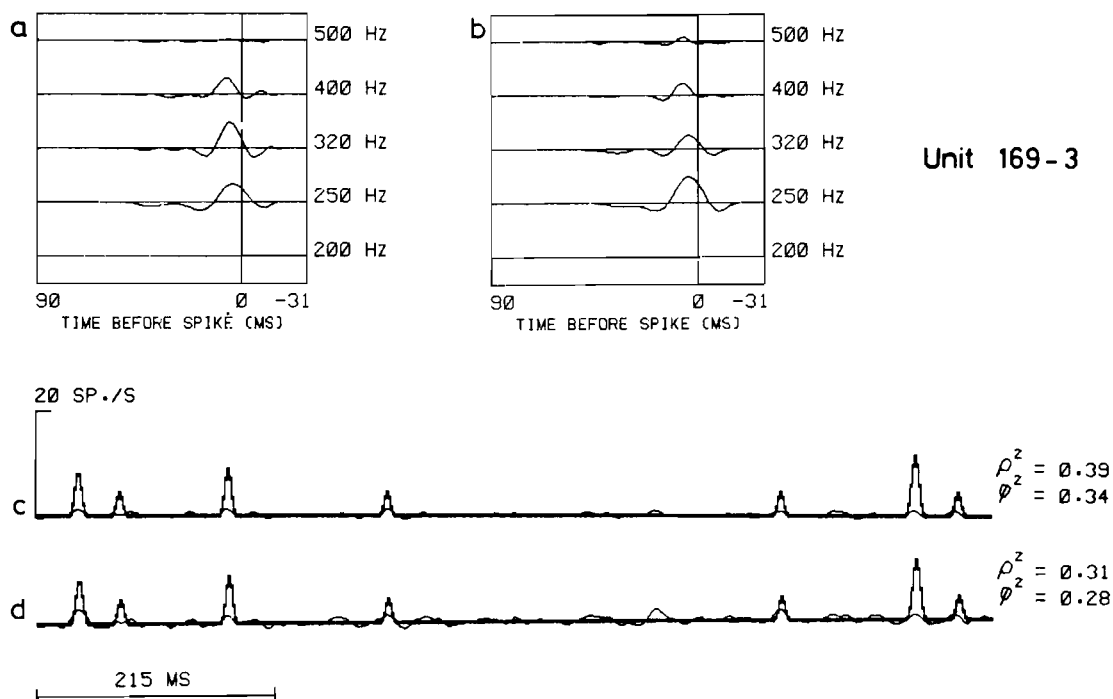


Fig. 3. The results of the response prediction and the characterisation of unit 169-3. The STRF under stimulation with < 500 Hz noise is shown in a; the STRF under stimulation with < 1500 Hz noise is shown in b. The STRF for < 500 Hz noise was used to predict the response to < 1500 Hz noise, the results of which are shown in c. The thin line shows the predicted response, the thick line the actual response. The results of the characterisation for < 1500 Hz noise are shown in d.

TABLE II

AVERAGE FIRING RATES AND MEASURES OF SIMILARITY FOR THE PREDICTIONS COMPARED TO THOSE FOUND FOR THE CHARACTERISATION

	P_o	$P_{\hat{p}p}$	
Unit 175-2			
Prediction	0.87	0.37	
Characterisation ¹	0.84	0.35	¹ under STRF-noise
Unit 166-9			
Prediction	12.3	0.67	
Characterisation ²	8.9	0.70	² under PSTH-noise
Unit 169-3			
Prediction	0.26	0.62	
Characterisation ³	0.21	0.62	³ under PSTH-noise

prediction, while it was the same for the characterisation under the noise that was used as PSTH-noise (Fig. 3c, d).

The results of these predictions are summarised in Table II, where they are compared with the figures obtained for the characterisation.

Response prediction to species-specific vocalisations

The results of predictions to species-specific vocalisations were very variable and did not allow a comprehensive survey. Correlation coefficients are measures of doubtful significance under these nonstationary conditions, and were very low. The results of one unit will be dealt with in detail. This neuron, unit 161-4, had its best frequency in the third-octave band of 630 Hz and a clear postactivation suppression, as shown by its STRF (Fig. 4a).

Predicted and actual responses are presented for the start of a male vocalisation in Fig. 4b. Inspection of the response records shows that the neuron fired once to each tone pip of the vocalisation even at the beginning where the intensity of the tone pips fell below the resolution capacity of the sonogram. When we ignore the negative parts of the prediction, resulting from postactivation suppression, the prediction is too low initially but reaches about the same magnitude later on. During the whole presented segment the predicted response lags behind the PSTH. Fig. 4c presents predicted and actual responses to three female vocalisations. The first vocalisation contains most of its energy in the 800 Hz band and the neuron generally responded with one spike to each presentation with an occasional second spike at the end of the vocalisation. The first spike occurred at the moment when the predicted response attained a significant value. The second spike occurred at the end of the postactivation suppression. The second female vocalisation consists of a first part with most energy in the 800 Hz band followed by a second part with energy in the 630 Hz band. The neuron responded in most cases with two spikes. The first one occurred at each presentation, while the second one sometimes failed to occur. The prediction shows two parts with a high firing probability that lagged behind the

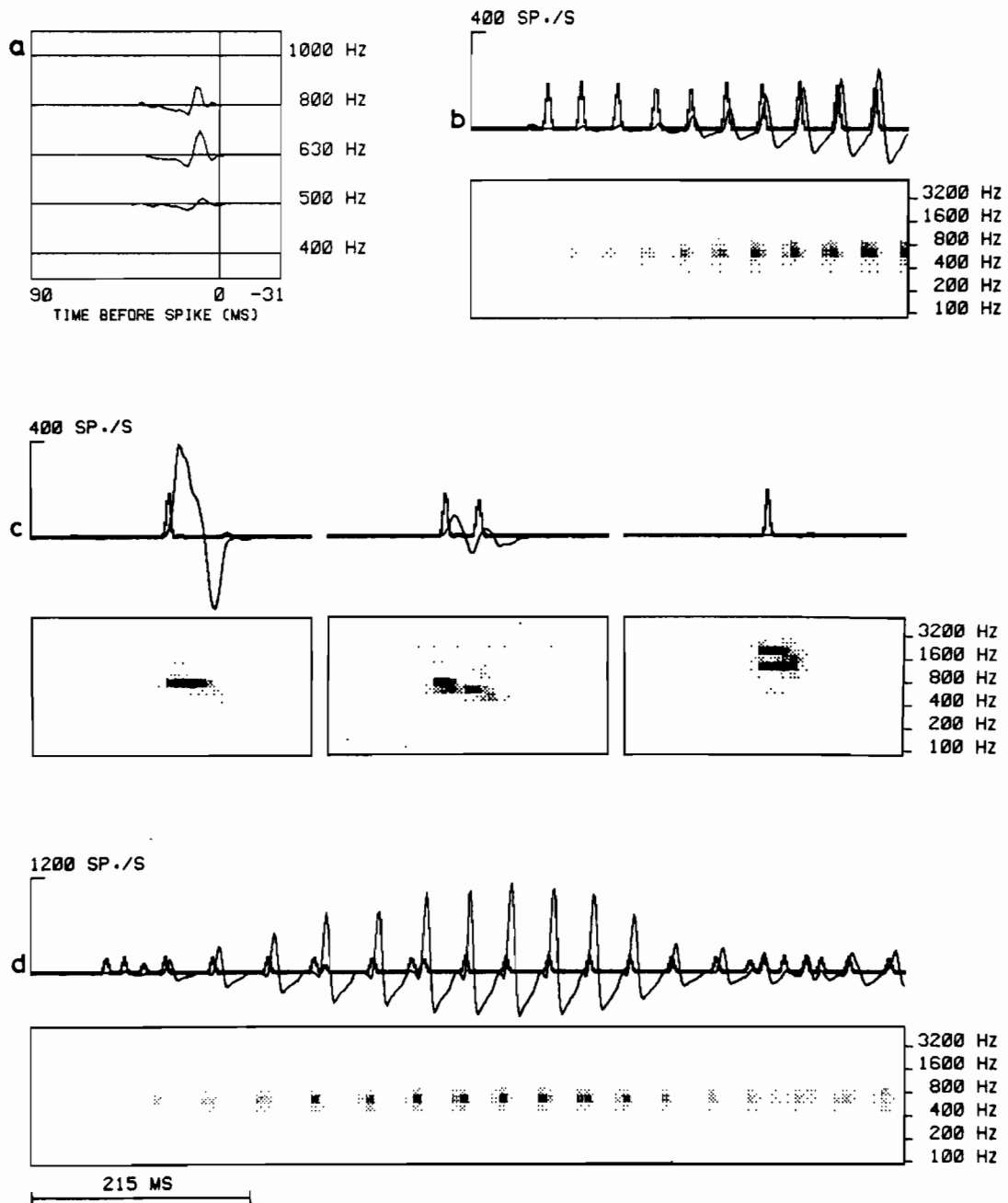


Fig. 4. The results of the prediction of the responses to species-specific vocalisations for unit 161-4. The STRF is shown in a. In b the predicted response (thin line) and actual response (thick line) are shown for a male vocalisation the sonogram of which is shown underneath. In c predicted and actual responses are shown for three female vocalisations. In d predicted and actual responses are shown for another male vocalisation. Note the different scaling in d.

PSTH. The third female vocalisation had most of its energy outside the STRF resulting in a very low predicted response. Nevertheless, the neuron responded with one spike to each presentation of this vocalisation.

Fig. 4d finally shows the results for another male vocalisation. The predicted response becomes much too high and lagged behind the actual response. At some instances double firings occurred, some indication for which can also be seen in the predicted response.

The results of the other neurons were in most cases similar. At the start of a male vocalisation the predicted response was much too low and lagged behind the actual response. Later on the predicted and actual responses could attain comparable values but in many cases the predicted response remained much lower than the actual response especially when the vocalisation contained most of its energy outside the STRF. When the energy of the vocalisation overlapped with the STRF, prediction was better at the end of the vocalisation where, for example, the lag between predicted and actual response became less. All tested units, none of which had best frequencies below the 500 Hz band, responded to at least some of the female vocalisations even when the spectral content was outside their STRF *. In most cases only one spike was produced at each presentation with an occasional second spike. Only one neuron with hardly any postactivation suppression (STRF shown in Fig. 2) responded in a more or less sustained way to the female vocalisations in accordance with the predicted response. Another unit, however, which did not show any significant postactivation suppression either, fired only once at each presentation while the predicted response lasted much longer. One neuron responded with one or more spikes at the presentation of all vocalisations after which it stopped firing. In the case of male vocalisations only the latter peaks in the PSTH showed some relation with the predicted response.

Discussion

General

The STRF, which forms the basis of the response predictions, was determined under stationary stimulus conditions. Its derivation makes an essential use of this stationarity, because otherwise neither the autocovariance function of the input nor the input-output correlation function are properly defined functions. Furthermore, the STRF is a stimulus-dependent functional, as illustrated by unit 169-3 in this study (Fig. 3). Only with respect to the stimulus for which the STRF was determined the STRF-based response provides an optimal estimate in mean-square-error sense when the neuron is described according to Eqn. 2. When the STRF is used to predict the response to stimuli with different statistical properties, this is no longer realised.

The stationarity of the conditions under which the STRF was determined implies that the average temporal intensities of the various frequency bands are time-inde-

* Unit 169-3 (Fig. 3) showing lateral suppression did not respond to vocalisations at all, probably because these vocalisations had their main energy in frequency bands higher than the activation region of the STRF, where their effect, if any, was suppressive. Because of the lack of spontaneous activity this could not be verified.

pendent. The temporal intensity of the frequency band to which the neuron is sensitive gently fluctuates around a well defined average and never becomes very low nor very high for a considerable period of time. This means that the *adaptation* state of the neuron will remain within a small range. Under stimulation with noise of a different average intensity this range will be different. Since adaptation can strongly influence the sensitivity of the neuron, the predicted responses will deviate accordingly. Because of this adaptation action potentials produced during the first sequence of the pseudorandom noise were excluded from the calculation of the STRF.

This adaptation state will vary much more strongly under stimulation with signals as species-specific vocalisations. Because of the long silent intervals between them the neurons will be almost completely unadapted at the start of the vocalisation. Furthermore, the energy of the vocalisations of the grass frog is concentrated in one or a few frequency bands, whereas the energy of wide-band noise is distributed over all frequency bands within the bandwidth of the noise. Consequently, the peak intensity in the main band of a vocalisation will attain much higher values than for noise, provided their total energy has the same order of magnitude.

Prediction of responses to stationary noise

If two kinds of noise differ only in sequence length and have the same bandwidth and overall stimulus intensity, one must obtain equal STRFs. This explains why for unit 175-2 the measures of similarity for the characterisation were similar to those for the prediction.

The significance of adaptation is indicated by a comparison of the predicted response to < 5000 Hz noise (Fig. 2c) which is based upon the STRF belonging to < 1500 Hz noise, and the STRF-based response for < 5000 Hz noise (Fig. 2d), which is based upon the STRF belonging to the < 5000 Hz noise itself. They appeared to be nearly identical except for a constant factor, the predicted response being lower than the STRF-based response. This shows that the STRFs for < 1500 Hz noise and < 5000 Hz noise differ only by a constant factor. When the STRF of < 1500 Hz noise is used for the prediction to < 5000 Hz noise too low a response is predicted. This can be explained by inferring that under stimulation with < 1500 Hz noise the neuron is more adapted than under stimulation with < 5000 Hz noise, which is likely because the average intensity in the main frequency band of the STRF, the 1250 Hz band, is 2.3 dB (a factor 1.7) higher than for < 5000 Hz noise. As a consequence the prediction to < 5000 Hz, based on the STRF of the neuron in a more adapted state, will come out too low. In this case adaptation has the character of a level-dependent gain control.

For unit 169-3 with lateral suppression the situation is more complicated. The STRF-noise had about the same average intensities in the bands below 500 Hz as the PSTH-noise but they were lower in the upper frequency bands of the STRF. Also here the predicted response was too low. The neuron might 'adapt' to the stronger stimulation of the higher frequency bands, which suppress the activation of the lower frequency bands, by showing a stronger activation in regions away from the suppression region.

Prediction of responses to species-specific vocalisations

Under highly nonstationary stimulation with signals as species-specific vocalisations phenomena such as described for the predictions to stationary noise will come out much more pronounced. At the start of a vocalisation, when the neuron is nearly unadapted, the predicted response that is based on properties of the neuron in adapted state, will be much too low. This can also explain why in most cases the predicted response lagged behind the actual response, since latency to tones with long silent intervals (1 s) is about the same amount shorter than to tones with no silent intervals in between [5]. Furthermore, the temporal intensity of the main frequency bands of species-specific vocalisations attains much higher values than in the case of noise. Many auditory neurons in unadapted state show much broader frequency sensitivities to tones of high intensity than to stationary noise [5]. Also in mammals broader sensitivities for tones of high intensities are very often found. So this may correspond to the finding that many neurons fired to vocalisations with a spectral content outside their STRFs. The lack of adaptation and the high peak intensities of the vocalisations can also explain a phenomenon hardly observed under stimulation with stationary noise, viz. saturation. In the smoothed PSTHs presented in this paper this is attained at 200 spikes/s, in which case the neuron always fired at the instance when the stimulus was presented. Under stimulation with pseudorandom noise this value was only very occasionally attained. Adaptation to the average stimulus intensity of the noise apparently prevents a large-scale saturation. In the PSTHs to species-specific vocalisations saturation at the onset was only absent when the spectral content of the vocalisation was far outside the STRF.

So, these considerations can qualitatively explain many aspects of these response predictions. To explain them quantitatively the dynamic properties of adaptation have to be investigated thoroughly. In contrast to auditory nerve fibres of mammals [7], a study in the peripheral auditory system of anurans [6] showed that the time constants of short-term adaptation can vary over a very wide range. More centrally the situation is bound to be even more complex.

Furthermore, in the accompanying paper [3] it was argued that complex interactions between different parts of the STRF may explain deviations of the STRF-based response from the actual response. When a neuron has, for example, postactivation suppression it is possible that a high intensity in the frequency band to which the neuron is sensitive must necessarily be preceded by a low intensity if the neuron is to fire. This necessity does not reveal itself in the STRF. This condition is pre-eminently realised at many instances in species-specific vocalisations. In the vocalisations of the grass frog this can be observed at the start of a female vocalisation and at the start of all tone pips that constitute a male call. In addition to adaptation, also such a dynamic nonlinearity may explain why the responses at these instances are much stronger than predicted. As another consequence, the neuron then only fires once as long as the intensity of the sound signal within the frequency-sensitivity range of the neuron remains high. This may explain why unit 161-4 fired once to female vocalisations with ongoing intensities, whereas it fired twice to a female vocalisation with a low intensity in the middle (Fig. 4c). A phenomenon like this can be described by assuming inhibition by a nonadapting neuron with the same

frequency sensitivity but with a longer latency. It is in this respect remarkable that unit 166-9 (STRF shown in Fig. 2) which hardly showed any postactivation suppression, kept firing as long as the predicted response remained high. Why unit 175-2 (STRF shown in Fig. 1) fired only once to a female vocalisation, despite its lack of postactivation suppression, must be due to another mechanism. The presence of complex interactions between different parts of the STRF is also indicated by the different STRFs of unit 169-3 with lateral suppression (Fig. 3) found under stimulation with two different kinds of noise.

Summarising, two explanations are presented for the deviations between the predicted response to a species-specific vocalisation and the actual response. The first consists of a level-dependent gain control with a relatively long time constant. Under stimulation with pseudorandom noise this gain control induces a constant adaptation state in the neuron, whereas under stimulation with species-specific vocalisations the neuron is much more sensitive especially at the start. The second explanation relates to conditions which do exist in noise but not to the same extent as in species-specific vocalisations. Because of this the contribution of these dynamic nonlinearities to the response is much stronger in species-specific vocalisations than in pseudorandom noise.

It must be emphasised that these arguments do not only apply to species-specific vocalisations but also to other tonal stimuli. They also show, that neurons respond to properties of species-specific vocalisations that are almost completely absent in pseudorandom noise. Therefore, phenomena come forward that are obscured under stationary stimulation with noise which is also mentioned by Symmes [8]. In this way the basis is explored of the differential response to tonal sounds such as species-specific vocalisations. This clarifies in which direction the selective information processing of species-specific vocalisations higher up in the central nervous system must be investigated.

Acknowledgements

This investigation was supported by the Netherlands Organisation for the Advancement of Pure Research (ZWO). Dik Hermes performed all the calculations. The paper is based upon part of his Ph.D. Thesis 'Spectro-temporal characterisation of auditory neurons in the torus semicircularis of the grass frog, *Rana temporaria* L.' (University of Nijmegen). The manuscript was critically read by Ton Vendrik. Koos Braks assisted in the recording of species-specific vocalisations. Wim van Deelen and Jan Bruijns were of great help in the various aspects of data analysis. Marianne de Leng prepared the manuscript.

References

- 1 Brzoska, J., Walkowiak, W. and Schneider, H. (1977): Acoustic communication in the grass frog (*Rana t. temporaria* L.): Calls, auditory thresholds and behavioral responses. *J. Comp. Physiol.* 118, 173–186.

- 2 Diakov, C. (1977): Initiation and inhibition of the release croak of *Rana pipiens*. *Physiol. Behav.* 19, 607–610.
- 3 Eggermont, J.J., Aertsen, A.M.H.J. and Johannesma, P.I.M. (1983): Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. *Hearing Res.* 10, 167–190.
- 4 Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M. and Eggermont, J.J. (1981): Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with noise stimuli. *Hearing Res.* 5, 145–178.
- 5 Hermes, D.J., Eggermont, J.J., Aertsen, A.M.H.J. and Johannesma, P.I.M. (1982): Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with tonal stimuli. *Hearing Res.* 6, 103–126.
- 6 Megela, A.L. and Capranica, R.R. (1981): Response patterns to tone burst in peripheral auditory system of anurans. *J. Neurophysiol.* 46, 465–478.
- 7 Smith, R.L. (1977): Short-term adaptation in single auditory-nerve fibers: Some poststimulatory effects. *J. Neurophysiol.* 40, 1098–1112.
- 8 Symmes, D. (1981): On the use of natural stimuli in neurophysiological studies of audition. *Hearing Res.* 4, 203–214.
- 9 Van Gelder, J.J., Evers, P.M.G. and Maagnus, G.J.M. (1978): Calling and associated behaviour of the common frog, *Rana temporaria*, during breeding activity. *J. Anim. Ecol.* 47, 667–676.