Auditory Discrimination of Amplitude Modulations Based on Metric Distances of Spike Trains

Sandra Wohlgemuth and Bernhard Ronacher
Department of Biology, Humboldt University, Berlin, Germany

Submitted 24 November 2006; accepted in final form 13 February 2007

Wohlgemuth S, Ronacher B. Auditory discrimination of amplitude modulations based on metric distances of spike trains. J Neurophysiol 97: 3082–3092, 2007. First published February 21, 2007; doi:10.1152/jn.01235.2006. Sound envelope cues play a crucial role for the recognition and discrimination of communication signals in diverse taxa, such as vertebrates and arthropods. Using a classification based on metric similarities of spike trains we investigate how well amplitude modulations (AMs) of sound signals can be distinguished at three levels of the locust’s auditory pathway: receptors and local and ascending neurons. The spike train metric has the advantage of providing information about the necessary evaluation time window and about the optimal temporal resolution of processing, thereby yielding clues to possible coding principles. It further allows one to disentangle the respective contributions of spike count and spike timing to the fidelity of discrimination. These results are compared with the traditional paradigm using modulation transfer functions. Spike trains of receptors and two primary-like local interneurons enable an excellent discrimination of different AM frequencies, up to about 150 Hz. In these neurons discriminability depends almost completely on the timing of spikes, which must be evaluated with a temporal resolution of <5 ms. Even short spike-train segments of 150 ms, equivalent to five to eight spikes, suffice for a high (70%) discrimination performance. For the third level of processing, the ascending interneurons, the overall discrimination accuracy is reduced. Spike count differences become more important for the discrimination whereas the exact timing of spikes contributes less. This shift in temporal resolution does not primarily depend on the investigated stimulus space. Rather it appears to reflect a transformation of how amplitude modulations are represented at more central stages of processing.

Introduction

Auditory systems are characterized by their exquisite sensitivity to temporal features of acoustic stimuli. Traditionally one distinguishes between the “fine structure” of a stimulus, determining its carrier frequency content, and the amplitude modulations (AMs), which change on a much slower timescale, thus constituting the “sound envelope.” Even in species whose hearing organs exhibit excellent (carrier) frequency-discrimination envelope cues are important in various perceptual tasks, such as for speech recognition in humans (Rosen 1992; Shannon et al. 1995; Smith et al. 2002) and song recognition in birds (Kroodsma and Miller 1996; Woolley et al. 2005, 2006). Evidently, envelope cues are of crucial significance for the recognition of communication signals in species with limited capacities for frequency analysis, such as frogs and insects (Alder and Rose 2000; Gerhardt and Huber 2002; Hennig et al. 2004; Pollack 1998, 2000; von Helversen and von Helversen 1997, 1998). Thus to uncover how amplitude modulations are processed and represented will be central for our general understanding of auditory systems (reviews: Joris et al. 2004; Langner 1992).

An established method of investigating the processing of amplitude modulations (AMs) by auditory systems is based on modulation transfer functions [rate (rMTFs) or temporal (tMTFs)]. These assays reveal the range of AM frequencies within which auditory neurons provide specific information about the stimulus envelope (e.g., Joris and Yin 1992; Joris et al. 2004; Krishna and Semple 2000). However, MTFs do not take into account the reproducibility of responses, which is an important factor that constrains neural encoding schemes. Moreover, these traditional procedures do not always provide an answer to the key question of how well different AM frequencies can indeed be discriminated on the basis of neural signals—the only information available to the CNS.

Discrimination may be poor, for example, if neurons exhibit low-pass or broad band-pass MTFs. To assess the fidelity of discrimination based on spike train comparison alone, here we apply a method to determine metric distances between spike trains, as introduced by van Rossum (2001). The spike train metric method has the advantage of providing information about the temporal precision with which spike trains should be evaluated, to obtain optimal discrimination (Machens et al. 2003). Thus it can help to infer the type of coding prevalent in neurons. We then compare these results with the description by the traditional tMTF and rMTF functions.

As an experimental system we chose the metathoracic auditory pathway of grasshoppers, a well-established model system for the processing of acoustic patterns (e.g., Gollisch et al. 2002; Machens et al. 2001a; Rokem et al. 2006; Stumpner and Ronacher 1991, 1994; Stumpner et al. 1991). This auditory system is conserved in evolution across species and is able to process a large variety of stimuli with a focus on amplitude modulations (Stumpner et al. 1991; von Helversen and von Helversen 1997). It is characterized by a hierarchical organization, consisting of three consecutive levels of processing: receptor neurons, local interneurons, and interneurons ascending to the brain, which are connected to receptors by only two or three steps of synaptic transmission. The classification of acoustic stimuli based on spike train similarities will be compared for these processing levels. This comparison can reveal how the representation of AM stimuli changes and what coding strategies may be used at different stages of the auditory...

METHODS

Animals and electrophysiology

All experiments were performed on adult locusts (Locusta migratoria), obtained from a commercial supplier. Legs, wings, head, and gut were removed and the animals were fixed with wax, dorsal side up onto a Peltier element, attached to a holder. The thorax was opened dorsally to expose the auditory nerve and the metathoracic ganglion, which houses the auditory pathway in question, was stabilized by a small NiCr platform. The whole torso was filled with locust Ringer solution (Pearson and Robertson 1981). During the experiments the preparation was kept at a constant temperature of 30 ± 2°C.

Auditory receptors and auditory interneurons were recorded intracellularly in the frontal auditory neuropil of the metathoracic ganglion. Neural responses were amplified (SEC-05LX; npi electronic, Tamm, Germany) and recorded by a data-acquisition board (PCI-MIO-16E-1; National Instruments, Munich, Germany) with a sampling rate of 20 kHz.

The tips of the glass microelectrodes (borosilicate, GC100F-10; Harvard Apparatus, Edenbridge, UK) were filled with a 3–5% solution of Lucifer yellow (Sigma–Aldrich, Taukirchen, Germany) in 0.5 M LiCl. After completion of the recordings the dye was injected into the recorded cell by applying hyperpolarizing current. After an experiment the thoracic ganglia were removed, fixed in 4% paraformaldehyde, dehydrated, and cleared in methylsalicylate. The stained cells were identified under a fluorescence microscope according to their characteristic morphology (Römer and Marquart 1984; Stumppner and Ronacher 1991). The experimental protocol complied with German law governing animal care.

Acoustic stimulation

All experiments were performed in a Faraday cage lined with foam prisms to attenuate echoes. Acoustic stimuli were broadcast by one of two speakers (D-28/2, Dynaudio, Skanderborg, Denmark), situated laterally at a distance of 30 cm from the preparation. The signals were amplified (Mercury 2000; Jensen, Pulpheim, Germany) and attenuated (PA5; Tucker-Davis Technologies, Gainesville, FL). All acoustic stimuli were stored digitally and delivered by custom-made software (PA5; Tucker-Davis Technologies, Gainesville, FL). Sound intensities were calibrated with a Bruel & Kjær (Nærum, Denmark) microphone (1/2-in.), and are given in decibels re 20 μPa (dB SPL).

Auditory stimulation

Once the distances between all spike trains of a certain cell were calculated matrices (see Fig. 2, middle column) a cluster algorithm was applied to quantify the classification success and thus the discrimination performance of the stimuli based on the spike trains. For reliable discrimination, spike trains elicited by the same stimulus should have smaller distances to each other than spike trains elicited by a different stimulus, resulting in a tight clustering of spike trains. Clustering was examined by randomly picking a spike train as a “template” for each stimulus. The remaining spike trains were classified by assigning them to that template to which they exhibited the smallest distance. This procedure was repeated for all possible template permutations, resulting in an average probability of classification, shown in the classification matrices (see Fig. 2, right column). The diagonal elements correspond to correctly classified spike trains, the off-diagonal elements to misclassified spike trains.

We checked whether the sample size of eight spike trains is sufficient for a reliable classification by dividing the data set in two and performing the analysis with this reduced set. This reduction to four spike trains per stimulus, however, yielded a discrimination
pattern and classification success that was highly similar to that of the complete data set, indicating that our sample size was adequate.

The classification success over the whole range of stimuli (Fig. 3) was calculated as the mean percentage of correct classifications over all spike trains. Because the classification success depends on the resolution parameter $\tau$ (see Fig. 6, A–C), for each cell its optimal $\tau$ value was used. These values ranged between 1 and 30 ms.

To obtain measures for the limits of correct discrimination that can be compared with MTF measures, corner frequencies were derived on the basis of the percentages of correct classifications for each modulation frequency (Fig. 5). Using the stimulus that elicited the maximum of correctly classified spike trains (see Fig. 5, A–C), the modulation frequency was determined for which the classification dropped to 90% of the maximum (see also Fig. 1C in Krishna and Semple 2000).

To investigate how the discrimination ability changes with the temporal resolution applied to the spike trains we plotted percentages of correctly classified spike trains as a function of the parameter $\tau$ within an evaluated time window of 500 ms (Fig. 6, A–C). From the maxima of these curves the optimal $\tau$ values were determined (Fig. 6D). If there was a broad constant maximum (e.g., at 100% classification; see Fig. S1A); blue curves for REC and BSN1) the midpoint $\tau$ value of this maximum was taken. In addition, a range of $\tau$ for near-optimal classification was determined representing the range of $\tau$ values yielding a performance 10% below the maximum of the curves (compare Fig. 6, D and E).

Modulation transfer functions

The responses to all AM stimuli as a function of modulation frequency were also characterized by rate modulation transfer functions (rMTFs) and temporal modulation transfer functions (tMTFs), based on vector strength, that describe the phase-locking of spikes to the stimulus envelope. Details of these procedures were given previously (for rMTF: Krishna and Semple 2000; for tMTF: Joris and Yin 1992; Krishna and Semple 2000). In short, for the rMTF (compare Fig. 1C) mean spike rates of the modulated part of the stimuli were computed (averaged over a time window of 500 ms). If the spike rate changed in the tested range of MF by $\geq 33\%$ (compare also Fig. 8C) a cutoff frequency was derived from the curves. This was the frequency at which the response fell to the minimum spike rate plus 10% of the difference to the maximum spike rate (Fig. 1C, left). If there were two distinct maxima of spike rate, the cutoff frequency according to the lower-frequency maximum was taken (following Krishna and Semple 2000); see point W1 in their Fig. 1E.

For the computation of the tMTF, period histograms (divided into 18 bins, i.e., $2^{18}$) were obtained over the whole range of 4-s amplitude modulations of each stimulus. From the period histograms values of the vector strength (VS) were calculated according to

$$VS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \cos \alpha_i} + \left(\frac{1}{n} \sum_{i=1}^{n} \sin \alpha_i\right) \quad (J)$$

where $\alpha_i$ is the timing of spike $i$ defined as phase of the modulation waveform and $n$ is the number of spikes. The vector strength can vary from a minimum of zero to a maximum of 1, the latter indicating a “perfect” phase locking, i.e., all spikes falling in the same bin. From the resulting tMTF curves corner frequencies were obtained (see Fig. 1C; compare also with the rMTF analysis). The corner frequency was determined as the frequency where the curve dropped to the 90% point from the maximal vector strength (Fig. 1C, right). In all cases, linear interpolation was used if required. Only the cutoff frequencies derived from the rMTF curves and corner frequencies of the tMTF curves are shown here (Fig. 8).

$^{1}$ The online version of this article contains supplemental data.

FIG. 1. Summary of the stimulus characteristics and the response measures used. A: sample amplitude modulation (AM) stimuli and spike traces of a local bisegmental interneuron (BSN1). B: spike trains were convolved with an alpha function to compute metric distances. Width of the filter is given by $\tau$. C: sample rate modulation transfer function (rMTF) and temporal modulation transfer function (tMTF) (BSN1 neuron) and definition of corner and cutoff frequencies. Two criteria used for the analysis are indicated in gray (cutoff frequency of the rMTF and corner frequency of the tMTF).

RESULTS

Patterns of spike train distances change from receptors to interneurons

The central question of this study is how well different AM stimuli can be discriminated by a nervous system, based on the sole information provided by the spike trains from auditory neurons. Discrimination will be impeded by the fact that spike trains exhibit some variability and usually differ even if an identical stimulus is presented several times. Thus a basic requirement for correct classification is that any spike trains elicited by repeated presentations of the same stimulus should be more similar compared with other spike trains elicited by a different stimulus. Our analysis therefore relies on an evaluation of spike train (dis)similarities, as introduced by van Rossum (2001) (see also Machens et al. 2003). Spike train distances within a stimulus reflect the response variability, whereas the distances across stimulus yield cues for discrimination, provided that across-stimulus distances are larger than within-stimulus distances.

The procedure shall be exemplified with the data of four neurons that are representative for the first three processing stages in the auditory pathway of the locust: receptor neurons, local interneurons, and ascending interneurons (Fig. 2). On the left, spike trains are shown as raster plots for the nine stimuli, each with eight stimulus presentations, whereas in the middle column the distances of each spike train to all other spike trains are shown in a color-coded two-dimensional matrix. Comparing a spike train to itself yields a distance of zero (as visible in the purple points on the diagonal). The distance matrix of the
A receptor cell shows distinct squares of dark blue (i.e., small distance values) aligned along the diagonal, up to a modulation frequency of 125 Hz (Fig. 2A). These $8 \times 8$ squares reveal a property of the spike trains that is crucial for the reliable classification of AM stimuli—that is, that the spike trains evoked by repeated presentations of a certain AM frequency should exhibit high similarity. For the receptor neuron the distances to all other spike trains appear to be rather homogeneously distributed over the whole matrix. The matrices of the three other neurons, however, are radically different (Fig. 2, B–D). The local bisegmental neuron (BSN1) exhibits a large region of uniformly small metric distances and therefore poor discrimination, for all spike trains produced by frequencies $>83$ Hz (zone of dark blue on the bottom right). The spike trains produced by 10, 20, and 40 Hz are very different from those produced by all other frequencies. The matrix of the ascending neuron AN3 (Fig. 2C) resembled that of BSN1, although the maximum distances were somewhat smaller. Again, a large homogeneous region of small distance values existed at $>40$ Hz. The pattern of spike train distances of the ascending neuron AN4 was different from that of the other three cells, exhibiting two symmetry axes. As a result of the almost complete cessation of spiking activity at modulation frequencies between 83 and 167 Hz (see raster plot on the left), all distances in the center of the matrix became very small. The responses to 10- and 20-Hz stimuli exhibited highest distances to these midfrequencies that elicited virtually no spikes, whereas the spike trains exhibited increased similarity to still higher frequencies as a result of the reappearing spike responses (compare Fig. 2D).

**Stimulus classification based on spike train distances**

As a next step the spike train distances were converted to a classification matrix, expressed as percentage of correctly classified stimuli, to obtain a measure for the discrimination performance (Fig. 2, right column). For each stimulus type one spike train was (randomly) selected, which served as a template for this stimulus. The distance data were then subjected to a cluster algorithm that assigned each spike train to that template to which it exhibited the smallest distance value. This was repeated for different template spike trains, resulting in percentages of correctly classified spike trains (values on the diagonal) and of misclassified spike trains (off-diagonal values).

**FIG. 2.** Examples of spike trains (left column), metric spike train distances (middle column), and classification matrices (right column) for neurons at 3 processing levels. A: an auditory receptor neuron. B: a local interneuron (BSN1). C and D: ascending interneurons (AN3, AN4). Left column: spike raster plots of responses from the 4 cells to the AM stimuli. For each stimulus 8 spike trains are shown. Middle column: distance values computed between the consecutively numbered spike trains (top to bottom) of the 4 cells shown in the left column (i.e., spike trains 1–8 belong to the 10-Hz stimulus, 9–16 to the 20-Hz stimulus, and so on). Distance matrices were calculated with optimal $\tau$ values for each cell (cf. Fig. 6D; A: $\tau = 2$ ms; B: $\tau = 4$ ms; C: $\tau = 5$ ms; D: $\tau = 20$ ms). Right column: resulting classification matrices. Rows and columns of the matrices represent the denoted stimuli. Matrix entries give the percentage of spike trains that were elicited by the stimulus a (horizontal axis) and classified as belonging to the stimulus b (vertical axis). Diagonal elements show the percentages of correctly classified spike trains (that is: a = b).
in Fig. 2, right column; see METHODS). The receptor responses yielded a perfect assignment of spike trains to the respective stimulus class \( H_11349 \) Hz. As expected from the distance data of the middle column, for the local and ascending neurons, BSN1 and AN3, the range of modulation frequencies allowing for a correct assignment of spike trains was reduced, leading to many misclassifications for frequencies \( H_11022 \) Hz. The classification pattern of the AN4 neuron was different again. Only the two lowest modulation frequencies could be clearly discriminated from all others on the basis of the spike trains, whereas in the midfrequency range no reasonable discrimination was possible because of the almost complete cessation of spiking. The spike trains produced by high modulation frequencies were again similar to those of low frequencies, resulting in an X-like pattern of misclassifications characteristic for this neuron type.

Note that local neurons are restricted to the thoracic ganglia and transmit information from receptors to ascending neurons that send their axons to the brain. Thus in this system receptor neurons and ascending neurons are separated by only two (or perhaps three) steps of synaptic transmission (cf. Boyan 1992, 1999). Among ascending neurons one finds a general increase in spike train distances and remarkably complex spike train distance patterns. In addition, the discrimination performance as measured by the probability of correct classification diminishes at higher processing levels.

Spike rates and the speed of discrimination

A basic requirement for on-line processing of acoustic signals is speed. In many situations animals must respond quickly to external stimuli and thus the minimal time needed for correct classification may be crucial. It is therefore a central question how the fidelity of discrimination does depend on the length of the spike train segment evaluated and thus on the number of spikes available. To answer these questions we computed the percentage of correctly classified spike trains as a function of the duration of the evaluation time window (Fig. 3A), or of the number of spikes available (Fig. 3B), for the three processing levels. Two trends were evident in the data: first, most curves showed a rather steep slope during the first 100 to 150 ms and then saturated. Second, the maximum classification performance decreased considerably from receptors to ascending neurons. Most receptors attained a level of correct classifications between 70 and 85\% and similarly high values were found for two local neurons with primary-like responses, TN1 and SN1. Note that 80\% correct classification indicates a quite remarkable discrimination ability, considering that our stimulus regime contained a large proportion of high-modulation frequencies (four of nine: 167, 250, 333, and 500 Hz). Interestingly, most specimens of another local neuron with phasic–tonic response pattern, BSN1, attained a much smaller classification success rate and, in this respect, appear more similar to ascending neurons, for which the classification level rarely exceeded 50 to 60\%. Among the ascending neurons, AN3 and AN11 types performed best, whereas the highly directional neurons, AN1 and AN2, performed worst.

There was a clear trend that classification performance decreased among ascending neurons. However, this could simply result from their lower average spike rates: Vogel et al. (2005) reported that ascending neurons exhibit distinctly smaller average spike rates than local neurons or receptor neurons. We controlled for such an influence by converting the \( x \)-axis of Fig. 3A into number of spikes evaluated. This led to a somewhat tighter clustering of the curves, in particular for the ascending neurons (Fig. 3B). Again there was a steep increase in classification performance up to five to ten spikes. In other

\[ \text{FIG. 3. Discrimination performance as the function of evaluated spike train length (A) and as the function of the number of evaluated spikes (B). For each cell means are computed over the correctly classified spike trains.} \]

\[ \text{FIG. 4. Comparison of classification success between processing levels. A: maximal classification success. B: classification success possible by taking into account 8 spikes. C: length of spike train segment necessary to achieve a classification of 40\% correct. D: test for correlations between classification success and mean spike rate. Spike rate was determined as the maximal rate achieved. For receptors (REC) and local neurons (LN) there was no significant correlation, whereas for ascending neurons (AN) there was a significant correlation (r = 0.59, P < 0.05).} \]
words, a spike train of 150- to 200-ms length, corresponding to a number of five to ten spikes, suffices for a rather reliable classification of AM stimuli, at least for receptors and local neurons. To further extend the evaluation time window beyond 200 ms would yield only little improvement.Remarkably, the difference in classification performance between processing levels cannot be explained by differing spike rates alone: even at the same spike count the probability of correct classification was distinctly worse for the ascending neurons than for receptors and local neurons (Fig. 4, A and B). As a consequence, there was a clear tendency that larger time windows were necessary for ascending neurons to achieve a moderate 40% probability of correct classification (Fig. 4C). We then looked for possible correlations between spike rate and classification success separately for cells of the different processing levels (Fig. 4D). For receptor and local neurons classification performance did not significantly depend on spike rate, whereas for the ascending neurons a weakly significant positive correlation was observed ($r = 0.59$, $P < 0.05$).

**Range of distinguishable modulation frequencies**

In the previous section (Figs. 3 and 4) we examined the general discrimination performance of neurons of the three processing levels. We now take a more detailed look at the influence that different modulation frequencies exert on the classification success, to enable a direct comparison with the results of the modulation transfer function paradigm. From the classification matrices of all recorded neurons (see Fig. 2, right column) we computed the percentages of correct classifications for each modulation frequency separately (Fig. 5). The spike trains of receptor neurons and the two primary-like local neurons TN1 and SN1 allowed for a perfect or near-perfect classification up to modulation frequencies of 125 Hz, or even 167 Hz (Fig. 5, A and B). The high overall classification performance (Fig. 3) reflects this behavior. For these neurons, the intratype variability was also small. This was different with another local neuron, BSN1. Some specimens of this neuron enabled the identification of stimuli at 125 Hz, whereas others approached chance level (11%) already at 83 Hz (see large SDs in Fig. 5B). This trend was even stronger with the ascending neurons. None of these allowed for a perfect classification at 125 Hz and most neurons already showed a steep decline of correct classifications between 20 and 83 Hz (Fig. 5C), which is the main reason for the drop in their overall classification success visible in Figs. 3 and 4. From the individual curves corner frequencies were determined, which are characterized by a value 10% below the maximum of the curves (similar as for the tMTF; see Fig. 1 and METHODS). The data indicate that for the ascending neurons and BSN1 there was considerable...
variability in the corner frequencies within one cell type (Fig. 5D). The corner frequencies of receptors TN1 and SN1 were tightly clustered at >125 Hz, indicating good discrimination up to nearly 150 Hz. Their distribution of corner frequencies (diamonds and open circles in Fig. 5D) had no overlap to that of ascending neurons, which were all <100 Hz, whereas BSN1 (crosses) occupied an intermediate position.

As a general picture, the responses of ascending neurons enabled a near-perfect discrimination of modulation frequencies only in a stimulus range that was much more restricted to low frequencies than that of receptors, TN1, and SN1. The two neurons that best encode sound direction (AN1, AN2; see Stumpner and Ronacher 1994) performed worst in the classification task and also for these neurons a very high spike count variability was previously found by Vogel et al. (2005). The specimens of the local BSN1 neuron were intermediate between receptor neurons and ascending neurons. The relatively large range of corner frequencies found for BSN1 probably can be explained by the fact that this neuron type exists in two (or perhaps even three) copies on each side (twin neurons; see Römer and Marquart 1984) that exhibit somewhat different response properties (Stumpner 1989). Ascending neurons seem to perform a relatively coarse classification, allowing discrimination only between broader subsets of AM frequencies.

Discrimination performance and temporal resolution

The discrimination results mentioned earlier were obtained using for each neuron its specific optimal \( \tau \)-value. This parameter determines the width of the filter function by which each spike was replaced to compute the metric distances (see Fig. 1B and METHODS). Thus \( \tau \) describes the temporal precision with which the spikes are evaluated for the distance metric. A very large \( \tau \) (>200 ms) has the consequence that the timing of spikes is virtually neglected, for which reason the metric distance between spike trains reflects mainly spike count differences (van Rossum 2001; see also Machens et al. 2003). Very small \( \tau \) values imply that the vast majority of spikes increase the distance, except for spikes of the two spike trains that are truly coincident. Changing the value of \( \tau \) thus influences the distances between spike trains and, consequently, also the classification success of stimuli. How the classification success varied with the resolution parameter \( \tau \) is shown in Fig. 6, A–C for the different neuron types recorded in this study. From the curves of individual specimens both the optimal \( \tau \)-value (i.e., at the peak of the curves) and the range of \( \tau \)-values enabling a near-optimal classification (90%; see inset) were assessed (Fig. 6, D and E). For receptors, the optimal \( \tau \) was <5 ms and the same was true for all local neurons, except one BSN1 (Fig. 6D). In sharp contrast, the ascending neurons (except one AN3) exhibited optimal \( \tau \) ≥5 ms; for AN12, AN4, AN2, and most AN1 the optimal \( \tau \) was >10 ms. As in other respects, the receptors and TN1, SN1 exhibited very similar curves. Although the optimal \( \tau \) values did not differ much between receptors and BSN1, the curves of the former extended even into the submillisecond range (Fig. 6, A and B). This can also be seen in Fig. 6E in which the ranges of \( \tau \) are shown that allowed a near-optimal classification, to avoid local biases in measuring the optima (see inset in Fig. 6).

An obvious objection to our approach is that the optimal resolution, and thus discrimination, may be strongly influenced by the stimulus ensemble used. Because our stimuli consisted of a large proportion of high-modulation frequencies, this might have biased the optima of the resolution parameter \( \tau \). In the data analysis we checked for such influences of the stimulus space by restricting our stimulus ensemble in three ways: exclusion of the three highest-modulation frequencies; restricting the stimuli to 83, 125, and 167 Hz; and exclusion of the six highest frequencies. The first two restrictions had only minor effects on the optimal \( \tau \) (see Fig. S1 in supplementary information). If the stimulus ensemble was reduced to 10–40 Hz, the range of optimal \( \tau \) became broader, extending up to about 50 ms, but still encompassed the \( \tau \) of the nine-stimulus ensemble. However, this extension toward larger \( \tau \) values does not necessarily imply a shift of the optimal \( \tau \) (or a change in the coding principle). Rather it indicates that it was simply not detrimental to increase the \( \tau \) as long as the spikes were distributed in large, well-separated clusters (see raster plots in Fig. 2).

Contribution of spike count and spike timing to discrimination

The metric distance measure offers a significant advantage: the width \( \tau \) of the filter function allows separation of the respective contributions of spike timing and of spike count differences to the discrimination of stimuli. This characteristic is visible in the curves of, say, AN3 or AN4 (Fig. 6). With very small \( \tau \) values (<0.5 ms) the stimulus classification dropped to chance level (11%), whereas for large \( \tau \) (>200 ms) performance was low—roughly 30 or 40%—but consistently above chance level. The proportion of correct classification at \( \tau = 1,000 \text{ ms} \) can be attributed exclusively to differences in spike count, whereas all information about spike times is ignored. Consequently, the difference between the percentage of correct classification at 1,000 ms and the peak value of the curves reflects the increment in classification performance that can be achieved by additionally taking into account temporal information about spike times (see double arrow in inset diagram in Fig. 6). These increment values based on spike times are plotted in Fig. 7A (gray bars), together with the discrimination success that can be achieved on the basis of spike count differences alone (black bars). The highest benefit that can be attributed to temporal information was attained by the primary-like TN1 and SN1 neurons, around 65%, followed by the receptors (roughly 60%). For these neurons the classification success depended nearly exclusively on spike timing, whereas spike count differences contributed only marginally (roughly 5% above chance level, Fig. 7A). The situation was clearly different with ascending neurons. For these the contribution of spike timing to discrimination dropped to nearly 20% and roughly equaled the contribution of spike count differences. The other local neuron, BSN1, resembled ascending neurons, with an increased spike count contribution and a reduced contribution of spike timing (only roughly 30%). The curves of Fig. 6 suggest a twofold trend across processing levels: toward an increased importance of spike count and toward reduced temporal resolution, i.e., larger optimal \( \tau \) values. There was indeed a strong negative correlation between the increment in discrimination attributable to spike times and the value of the optimal \( \tau \) (Fig. 7B; \( r = -0.85 \)), which suggests a general trend within the auditory pathway. The negative correlation re-
mained significant if receptors and the primary-like TN1, SN1 were excluded from the analysis ($r = -0.63, P < 0.01, n = 23$).

Comparison of discrimination performance with modulation transfer functions

A further goal of the present study was a quantitative comparison of the spike train metric paradigm with the traditional MTF methods. To directly relate results obtained with the metric method and the standard MTF paradigm, we derived corner frequencies from the curves of Fig. 5, which indicate the dynamic range (in this graph receptors and local neurons are included with their small rate differences). Taken over all neuron types there existed a strong positive correlation between the size of the optimal $\tau$ and the maximal spike rate difference in the rMTF curves (compare Fig. 1C; $r = 0.73, P < 0.0001$). However, excluding the data from receptors and TN1, SN1, thus reducing the data set to ascending neurons and BSN1, the significance disappeared ($r = 0.39, P \approx 0.06$). A comparison of the optimal $\tau$ with the corner frequencies derived from the tMTF curves yielded a strong inverse correlation (Fig. 8D; $r = -0.77, P < 0.0001$). Thus in general the corner frequency obtained from the tMTF is a good predictor for the temporal resolution with which an optimal classification of AM stimuli can be achieved.

What could be the cause of these strong correlations between $\tau$ and the MTF data? A hint comes from Figs. 7 and 8C. The cluster of small $\tau$ values at low spike rate differences (Fig. 8C) confirms that any information about the stimulus envelope must be encoded in the timing of spikes if different stimuli do not elicit reliable differences in spike count (see also Fig. 7A).

from the rMTF and metric corner frequencies ($r = 0.21; P > 0.10$; data not shown). In contrast, the comparison between metric and rMTF-based corner frequencies (Fig. 8B) yielded a rather close correspondence between the two measures for most cells (with the possible exception of BSN1). In particular, the data of receptors and TN1, SN1 resulted in very similar corner frequencies in both data evaluations.

Figure 8C compares the temporal precision parameter $\tau$ (see Fig. 6) with the difference between maximal and minimal spike rates obtained in the rMTF functions as an indicator of the dynamic range (in this graph receptors and local neurons are included with their small rate differences). Taken over all neuron types there existed a strong positive correlation between the size of the optimal $\tau$ and the maximal spike rate difference in the rMTF curves (compare Fig. 1C; $r = 0.73, P < 0.0001$). However, excluding the data from receptors and TN1, SN1, thus reducing the data set to ascending neurons and BSN1, the significance disappeared ($r = 0.39, P \approx 0.06$). A comparison of the optimal $\tau$ with the corner frequencies derived from the tMTF curves yielded a strong inverse correlation (Fig. 8D; $r = -0.77, P < 0.0001$). Thus in general the corner frequency obtained from the tMTF is a good predictor for the temporal resolution with which an optimal classification of AM stimuli can be achieved.

What could be the cause of these strong correlations between $\tau$ and the MTF data? A hint comes from Figs. 7 and 8C. The cluster of small $\tau$ values at low spike rate differences (Fig. 8C) confirms that any information about the stimulus envelope must be encoded in the timing of spikes if different stimuli do not elicit reliable differences in spike count (see also Fig. 7A).
The reverse, however, is not true: a neuron could encode information in the timing of spikes and in spike count differences. However, the shift toward larger $\tau$ values in ascending neurons (Fig. 8C) and the data of Fig. 7A show that among ascending neurons the spike timing information contributed less. Thus it must be asked why among ascending neurons is the temporal information about spike times obviously not very useful. A likely reason for this is a less-precise locking of spikes to the stimulus envelope (Fig. 2). Again the local BSN1 neuron, with rather small optimal $\tau$ values in spite of larger spike count differences, seems to occupy an intermediate position between primary-like and ascending neurons.

**DISCUSSION**

This study explored how envelope cues of acoustic stimuli are processed and represented in an auditory system of moderate complexity, thereby focusing on three main goals: First, how well can these stimuli be discriminated on the basis of spike train similarities. Second, what is the optimal temporal resolution for this discrimination task and how much does spike timing information contribute to the discrimination of amplitude modulated stimuli. Third, to compare the advantages and limitations of this method, with those of the “classical” paradigm of modulation transfer functions.

**Representation of amplitude modulations at consecutive stages of the auditory pathway**

In the auditory system of grasshoppers the ascending neurons are connected to receptor neurons by local neurons. Thus we have only few, in some cases only two to three, steps of synaptic processing that convert receptor spike trains to the responses of ascending neurons (Boyan 1992, 1999; Marquart 1985; Stumpner et al. 1991).

The responses of receptor neurons and the primary-like local neurons (TN1, SN1) allowed a remarkable discrimination, based on the classification of single spike trains, up to stimulus frequencies of about 150 Hz. Very short spike train segments and small numbers of spikes sufficed to attain this high classification success. Importantly, an extension of the evaluation time window $>200$ ms (corresponding to roughly 10 spikes) would yield only small improvements, thus implying a severe law of diminishing returns (Fig. 3). A short evaluation time window is in accord with behavioral observations in another grasshopper species, which recognizes conspecific signals even if these are shortened to 160- to 250-ms duration, in a one-shot manner (Ronacher and Krahe 1998). In contrast to receptors and TN1, ascending neurons fail to resolve AMs at much lower frequencies—yielding near-chance performance for frequencies $\approx 83$ Hz (Fig. 5). Interestingly, another local neuron, BSN1, that shares several features with ascending neurons, was previously shown to be an important distributor of information onto ascending neurons (Boyan 1992).

**Contributions of spike timing and spike count to discrimination**

The classification procedure applied here searched for the optimal discrimination by varying the width $\tau$ of the filter function used to determine spike train (dis)similarities (Figs. 1 and 6; see also Machens et al. 2003). A specific power of the metric method resides in the possibility of separating the respective contributions of spike count, as determined in a large “evaluation window,” and spike timing to the discrimination performance (Fig. 7). The comparison between the classification performance at optimal $\tau$ and at $\tau = 1,000$ ms revealed that the increment in classification success that can be attributed to the timing of spikes became small for most ascending neurons—in sharp contrast to the receptor and primary-like neurons for which the classification success depended essentially on spike timing, whereas spike count differences contributed very little (Fig. 7A). It should be emphasized, however, that the separation of spike count and spike time contributions makes sense only if the optimal $\tau$ is distinctly $<1,000$ ms, as was indeed the case in our system.

The optimal $\tau$ values show with what temporal resolution such an evaluation of spike timing should take place to extract the maximum information about the stimulus envelopes. The very small optimal $\tau$ values found for receptors and TN1 (Fig. 6D), mostly $<3$ ms, do not only demonstrate a high temporal precision of their spike trains (see also Rokem et al. 2006), combined with a high trial-to-trial reliability, but also that the spike trains of these neurons should be processed with outstanding temporal resolution to extract all information. If we interpret the filter function of Fig. 1B as the shape of postsynaptic potentials in downstream neurons, half-widths of 1–3 ms incur questions about the physiological plausibility of such a hypothetical evaluation procedure. However, this resolution problem is appreciably relieved if we consider the ranges of $\tau$ yielding 90% performance (Fig. 6E): without substantial loss of classification performance, these $\tau$ values now extend into the range of 5 to 10 ms and thus are consistent with observed durations of postsynaptic potentials (Boyan 1999; Franz and Ronacher 2002; Marquart 1985). For many ascending neurons the optimal $\tau$ values were shifted to 10–20 ms (see Fig. 6, D and E).

The classification procedure based on the spike train metric may be particularly suited to determine the characteristic timescale of a coding process. As Theunissen and Miller (1995) pointed out, the duration of the “encoding time window” associated with a neural code must be in accord with the rate of dynamic stimulus variations, and thereby probably reflects an (evolutionary or current) adaptation to the timescales of amplitude fluctuations in naturally relevant sound stimuli. Our results—in particular the controls with reduced stimulus ensembles shown in the supplementary Fig. S1—indicate that the optimum of the resolution parameter $\tau$ does not depend directly on the stimulus ensemble, but rather seems to reflect a specific property of neurons.

Obviously two factors may influence the observed values of the optimal $\tau$ in Fig. 6: 1) spike count differences between AM frequencies and 2) the temporal precision of spike trains, which constrains the upper limits of useful temporal resolution. If no spike count differences are available, any correct classification must rely on the timing of spikes, as was found for receptors and primary-like local neurons with their all-pass rate-MTFs. Their $\tau$ curves extend into the submillisecond range, indicating an outstanding temporal precision of spiking responses (Fig. 6; see also Rokem et al. 2006). If distinct spike count differences exist, as is true for many ascending neurons, the $\tau$ curves may extend to larger values to match this aspect (see Fig. 6). However, the shift of the left-hand flank to reduced temporal
resolution (and the corresponding shift in optimal $\tau$) obviously does reflect a decrease in spike timing precision. The spike raster plots of Fig. 2 do indeed suggest an increase of spike jitter between auditory receptors, BSN1, and the ascending neurons. This impression is supported by quantitative data by Vogel et al. (2005). In the locust auditory pathway these authors report a substantial increase of spike train variability between receptors and ascending neurons. Thus the optimal temporal resolution visible in the curves of Fig. 6 appears to reflect a compromise between averaging out spike time jitter on one hand and the loss of information about the stimulus’ temporal structure on the other (see also Narayan et al. 2006).

There are several general trends common to our observations in the auditory pathway of locusts and the processing of amplitude-modulated sound in vertebrates. Peripheral neurons of vertebrates encode envelope information by synchronizing their activity to the modulation waveforms, whereas neurons at higher levels become progressively tuned in terms of average firing rate, rather than synchronizing to the envelope (for a comprehensive review see Joris et al. 2004)—compare with Figs. 6 and 7. Such a transformation from temporal-to-rate coding may be a general feature of auditory systems (Joris et al. 2004; Lu and Wang 2004; Narayan et al. 2005, 2006; the term “temporal coding” is used here in the sense proposed by Theunissen and Miller 1995). Also at higher processing stages in vertebrates one observes a decrease in the highest AM frequencies that can influence the neural responses (Joris et al. 1999). The optimal frequency obtained from the tMTF curves (Fig. 8).

Comparison with the MTF paradigm

Modulation transfer functions are a classical method to describe the response of a neuron or a sensory system to the temporal pattern of amplitude modulations. However, the MTF curves are based on period histograms that summarize the responses to many stimulus periods. As a consequence, MTF procedures tend to disregard the trial-to-trial variability of neuronal responses, which is, however, crucial for recognition tasks. The spike train metric, in contrast, is susceptible to response variability and thus enables a more realistic approach to sensory processing. Indeed, the results presented in Fig. 8A indicate that the rate-MTF method may miss important aspects relevant for discrimination of different modulation frequencies. In contrast, the corner frequencies derived from tMTF and spike train metrics correlate rather well, but still the tMTF describes the locking of spikes to the envelope and not primarily the discriminability of different modulation frequencies. Here we show for the first time that the discrimination performance of interneurons correlates well with the corner frequency obtained from the tMTF curves (Fig. 8).

Taken together, the classification method based on spike train metrics applied here offers three important advantages, as compared with the tMTF method: 1) it is sensitive to the trial-to-trial variations of spike trains; 2) it yields additional information about the optimal resolution parameter $\tau$, as a potential clue to the encoding strategies present at various stages in the nervous system; and 3) it allows one to separate the contributions of spike count differences and spike timing to the principal discriminability of amplitude-modulated acoustic stimuli.

ACKNOWLEDGMENTS

We are indebted to C. Machens for providing programs for the discrimination based on spike train metrics and to R. Krahe and M. Hennig for critically reading and improving the manuscript. A. Vogel and G. Weschke kindly provided data of the AN12 neuron.

GRANTS

This work was supported by Deutsche Forschungsgemeinschaft Grant SFB 618 to B. Ronacher.

REFERENCES


