Brief Communications

Timescale-Invariant Representation of Acoustic Communication Signals by a Bursting Neuron

Felix Creutzig,1* Sandra Wohlgemuth,1* Andreas Stumpner,2 Jan Benda,1 Bernhard Ronacher,1 and Andreas V. M. Herz2

1Department of Biology and Bernstein Center for Computational Neuroscience Berlin, Humboldt-Universität zu Berlin, 10099 Berlin, Germany,
2Georg-August Universität Göttingen, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Neurobiologie, 37073 Göttingen,
Germany, and 3Department of Biology and Bernstein Center for Computational Neuroscience München, Ludwig-Maximilians-Universität München, 82152
Martinsried, Germany

Acoustic communication often involves complex sound motifs in which the relative durations of individual elements, but not their
absolute durations, convey meaning. Decoding such signals requires an explicit or implicit calculation of the ratios between time
intervals. Using grasshopper communication as a model, we demonstrate how this seemingly difficult computation can be solved in real
time by a small set of auditory neurons. One of these cells, an ascending interneuron, generates bursts of action potentials in response to
the rhythmic syllable–pause structure of grasshopper calls. Our data show that these bursts are preferentially triggered at syllable onset;
the number of spikes within the burst is linearly correlated with the duration of the preceding pause. Integrating the number of spikes
over a fixed time window therefore leads to a total spike count that reflects the characteristic syllable-to-pause ratio of the species while
being invariant to playing back the call faster or slower. Such a timescale-invariant recognition is essential under natural conditions,
because grasshoppers do not thermoregulate; the call of a sender sitting in the shade will be slower than that of a grasshopper in the sun.
Our results show that timescale-invariant stimulus recognition can be implemented at the single-cell level without directly calculating the
ratio between pulse and interpulse durations.

Key words: auditory system; acoustic communication; temporal pattern; spike train; burst code; invariant object recognition

Introduction

Object recognition relies on the extraction of stimulus attributes that are invariant under natural variations of the sensory input. Such stimulus variations include the size, orientation, or position of a visual object (Sáry et al., 1993; Ito et al., 1995; Rolls, 2000), the strength of an odor (Stopfer et al., 2003), and the intensity, frequency composition, or duration of a sound signal (Shannon et al., 1995; Moore, 1997; Bendor and Wang, 2005; Benda and Henning, 2008). A particular challenge arises when timescale-invariant features of an acoustic signal are to be extracted because this computation generally involves ratios of temporal quantities. For example, to calculate the relative duration of two sound patterns, the respective durations of both components need to be measured and their ratio calculated.

Many animals use acoustic communication signals to find conspecific mates and judge their reproductive fitness (Gerhardt and Huber, 2002). Temporal parameters of songs are species-specific and decisive for song recognition in most insects that use acoustic signals (Hennig et al., 2004). For ectothermic species, such as grasshoppers, this poses a severe problem because their calling songs are faster or slower depending on the ambient temperature. Sound signals that are rhythmically structured into syllables and pauses (see Fig. 1A) offer a solution if the receiver uses the temperature-independent syllable-to-pause ratio for species identification, as suggested by behavioral experiments with artificial stimuli (see Fig. 1B). Indeed, the syllable-to-pause ratio varies greatly between different species, facilitating the recognition of conspecific songs (von Helversen and von Helversen, 1994). However, to compute this quantity, durations of syllables and pauses need to be measured and compared with each other. The present study reveals that even a relatively small neural network such as the grasshopper auditory system (Stumpner and Ronacher, 1991; Stumpner et al., 1991) can achieve this task. The underlying computational strategy is surprisingly simple and applicable to a wide range of sensory systems that need to extract timescale-invariant stimulus features in real time.

Materials and Methods

Experimental paradigms, animals, and electrophysiology. For this study, two different sets of experiments were performed. In the first set, grasshopper mating songs (see Fig. 1A, C) were presented to six females, three Chorthippus biguttulus, and, for comparison, three Locusta migratoria. Homologous auditory receptor neurons and interneurons have been...
shown in Figure 1 were used: two examples are rising slope of the stimulus–response curve. For the natural-song paradigm, intensities were then chosen to lie approximately halfway on the level (SPL) in steps of 10 dB. For the core experiments, maximum stimulus intensities were used, each time increasing from the lowest to the highest stimulus intensity and vice versa, as shown in Figure 1A. The artificial songs consisted of six syllables with mean SD of the amplitude fluctuations of each song, 10 dB) do not obtain the correlation between spike count and signal features, the $R^2$ value (explained variance, Pearson's correlation) was calculated. It was tested for multicollinearity by computing the semipartial $R$ value for each independent variable. The mutual information between pause duration and IBSCs was calculated using the adaptive method (Nelken et al., 2005).

Song classification. AN12 responses were classified according to their IBSCs without relying on explicit temporal information: First, the IBSC of each burst was assigned to the time bin t (bin width, 2 ms) of the first spike within the burst. Second, for each bin t and song j, the IBSC number $i$ was selected, where $j$ is the highest (i.e., $j = \arg\min_{1 \leq j} (\text{diff}(i, j, k))$ where $\text{diff}(i, j, k) = \text{IBSC}(i, j, k) - \text{IBSC}(i, j, k)$. To avoid a bias, for each IBSC $(i, j, k)$, the corresponding IBSC $(i, j, k)$ was calculated without taking IBSC $(i, j, k)$ into account. Finally, for the cumulative classification, we calculated $\text{diff}_{\text{cum}}(i, j, k^*) = \sum_{k<i} \text{diff}(i, j, k)$. Here, $k^*$ denotes the burst position up to which information is used for cumulative song classification.

Let us emphasize that this analysis of responses to natural grasshopper songs is based on the stationary part of the calling songs, discarding the first nine syllable-and-pause segments of each song. We do so because, during the transient song phase, stimulus amplitudes increase slowly in time. Adaptation processes partially counteract this increase (Machens et al., 2001), leading to approximately equal burst responses on the AN12 level. However, because the duration of transients differs between songs, IBSCs calculated during the initial song phase carry additional information about the song identity and not about the large-scale structure of the song. To focus on the local syllable–pause structure, we analyze only responses to the stationary part of the songs.

Results

The response properties of a particular neuron (AN12) from the metathoracic ganglion, the first major relay station for auditory signal processing in grasshoppers, may provide an answer to the problem of timescale-invariant sound pattern representation. This neuron belongs to the ~50 auditory interneurons in the metathoracic ganglion, many of which have been morphologically and physiologically classified (Stumpner and Ronacher, 1994; Stumpner et al., 1991). Ascending neurons (ANs) project into the head ganglion. Some of them (AN1, AN2) encode directional information, whereas others (e.g., AN3, AN4, AN6, AN11, AN12) are presumably involved in pattern recognition (Stumpner and Ronacher, 1994; Krahe et al., 2002). AN12 has a large dynamic range (Stumpner and Ronacher, 1991) of ~40 dB and generates burst-like discharge patterns when stimulated by the amplitude-modulated sound patterns of grasshopper calling songs (Fig. 1C,D). Its anatomical and stimulus–response characteristics are almost identical in different grasshopper species (Stumpner and Ronacher, 1991; Stumpner et al., 1991; Neuhofer et al., 2008), indicating a highly conserved functional role. For the first part of our study, we recorded from C. biguttulus ($n = 3$) and L. migratoria ($n = 3$), as described in Materials and Methods. Unless stated otherwise, data from C. biguttulus are shown. Under stimulation with natural sounds, 67% of all spikes were contained in bursts with two or more spikes. Such bursts are mainly generated in response to syllable onsets (Fig. 1D). The IBSC (i.e., number of spikes within a burst) is highly reproducible from trial to trial but varies from syllable to syllable (Fig. 1E). Reflecting the different time courses of different songs (Fig. 1A), each song thus results in a particular IBSC
sequence (Fig. 1E). This signature can be used to discriminate among songs. For example, for a sample with eight songs from one species, each burst carries enough information to assign ~30% of the responses to the correct song, using the IBSC only (Fig. 1F, dashed line). Accumulated over time, a 90% hit rate is reached after 12 bursts (i.e., ~1 s) (Fig. 1F, solid line). This astounding discrimination performance is similar to that of grasshopper receptor neurons (Machens et al., 2003), although AN12 neurons have a far lower overall firing rate and their exact spike timing is neglected within the present analysis. Because IBSCs suffice to discriminate songs even from the same species, this measure must contain useful information about the detailed song structure and thus also help discriminate songs from different species.

Which song features elicit burst activity? To address this question, we constructed burst-triggered stimulus averages (i.e., the mean sound wave preceding a burst with given IBSC). As illustrated in Figure 2A, large IBSC values typically occur after extended and deep stimulus excursions, long pauses followed by strong syllable onsets. To identify the song feature that is most important for burst generation, we disentangled the effects of various song parameters (Fig. 1C), such as pause duration, onset amplitude, onset slope, total duration of the preceding song element (syllable plus pause), and minimal pause amplitude. This analysis revealed that pause duration covaries with IBSC (Fig. 2B). The correlation depends on the amplitude level used to define the pause duration, but over a wide range of amplitude levels, it is rather insensitive to parameter variations (Fig. 2C).

To ascertain that pause duration correlates with IBSC under various stimulus conditions, a second set of experiments was performed (C. biguttulus, n = 9). Here, the duration of artificial song units (Fig. 1B, inset) was systematically varied (see Materials and Methods) and IBSC increased linearly with pause duration (Fig. 2D). These data also demonstrate that IBSC does not depend on syllable duration. However, IBSC increases with the onset amplitude for natural songs, but with low \( R^2 \) value (Fig. 2E, top trace). This effect is attributable to covariations of pause duration and onset amplitude in the grasshopper songs. Finally, IBSC is not systematically influenced by onset slope (Fig. 2E, bottom trace) and all other song features tested. Together, these findings support the hypothesis that the AN12 intraburst spike count is well suited to encode the duration of pauses in grasshopper communication signals.

Could the intraburst spike count be used to reliably transmit information about specific structures of the natural calling songs, such as the pause durations, or does the trial-to-trial variability blur the IBSC signal too strongly? To investigate this behaviorally relevant question, we calculated the mutual information between IBSC and preceding pause duration, using the adaptive method (Nelken et al., 2005). On average, the IBSC transmitted 0.49 ± 0.24 bits per burst about the pause duration. Hence, a single burst would not convey enough information for a binary classification of pause
durations. However, grasshoppers can integrate over several subsequent bursts, as shown when they evaluate longer song segments to estimate song quality (von Helversen and von Helversen, 1994). This is useful because trial-to-trial IBSC fluctuations are not significantly correlated from burst to burst in individual cells \(p > 0.05\), turning point test (Kendall and Stuart, 1966)]. Consequently, groups of subsequent bursts could be used to encode average pause durations.

Successful auditory discrimination between songs of different species requires the readout of the syllable-to-pause ratio to be independent of the sender’s temperature and thus be timescale invariant. Can this task be accomplished with a neural code based on IBSCs? The following consideration helps to answer this question: Any global speed up of a mating song by some factor \(X\) increases the number of syllables per second and, hence, the burst frequency by \(X\). However, because intraburst spike counts are proportional to the preceding pause durations (Fig. 2 B, D), the IBSC decreases by the same factor \(X\). As a result, the total spike count per second stays constant and does not reflect the absolute syllable duration but rather the syllable-to-pause ratio (i.e., the relevant feature for time warp-invariant song pattern recognition) (Fig. 3A). This prediction is supported by experimental data, as shown by responses from a sample cell to two conspecific grasshopper songs that differ strongly in syllable duration but have almost identical syllable-to-pause ratio (Fig. 3B). Population data from all cells investigated confirm this finding: there is no correlation between the individual timescale of each of the eight songs, given by the duration of one syllable plus one pause, and the elicited total spike count for a given cell (Fig. 3C). Note that the lower spike count values at \(\approx 80\) ms correspond to one specific song. This particular song had a less pronounced syllable–pause structure, which resulted in a lower overall spike count. Integrating the AN12 response over a time window of \(\approx 500\) ms or longer thus suffices to reliably estimate the timescale-invariant syllable-to-pause ratio.

**Discussion**

Recognition of timescale-invariant stimulus features is a general challenge for sensory systems and seems to imply that ratios are computed in the temporal domain. Here, we showed that this computationally difficult problem can be solved on the single-cell level by a bursting neuron whose output is integrated over a few hundred milliseconds. Temporal integration over such time scales is well known from various neural systems, such as the auditory cortex (Nelken et al., 2003) or the electromotor system of weakly electric fish (Oestreich et al., 2006). The counting mechanism proposed for readout is similar to that of frogs that also communicate with rhythmic patterns of sound pulses and quiet intervals. Specific neurons in the frog’s auditory midbrain integrate the number of acoustic pulses and respond only if a minimal number of appropriate pulses and interpulse intervals arrive (Alder and Rose, 1998; Edwards et al., 2002, 2007). Combined excitation and inhibition determine the dependence of the responses on the duration of the acoustic elements (Edwards et al., 2008; Leary et al., 2008). Together with these results, our findings suggest that the recognition of temporally extended acoustic objects is supported by neural integration over multiple timescales. This hierarchical integration is reminiscent of current models for invariant object recognition in vision that suggest that high-level invariances are detected by successive spatial integra-
relevance of a particular sound pattern. Using pause duration instead of syllable duration may be beneficial because this strategy lowers the overall dependency on sound intensity. Although the timescale-invariant readout mechanism proposed by this study depends in principle on the existence of “pauses” in an atonal noise signal, this mechanism can be easily extended to stimuli with rich tonal structure if the raw signal is first split into multiple frequency bands. For rhythmic stimuli in acoustic environments with background noise, separately analyzing specific frequency components using the mechanism proposed in this study will help improve the signal-to-noise ratio in frequency bands with low noise content. Likewise, in mixtures of stimuli with different rhythms and carrier frequencies, a frequency-resolved analysis may reveal otherwise hidden signal components. Computationally demanding divisions in the time domain are thus not required to extract timescale-invariant information from sensory stimuli.

**References**


