Auditory Signal Processing: Physiology, Psychoacoustics, and Models.
Pressnitzer, D., de Cheveigné, A., McAdams, S.,and Collet, L. (Eds). Springer Verlag, 2004.

The enigma of cortical responses: Slow yet precise

Mounya Elhilali, David J. Klein, Jonathan B. Fritz, Jonathan Z. Simon, and Shihab A. Shamma

Institute for Systems Research & Department of Electrical and Computer Engineering,
University of Maryland, College Park MD
{mounya,djklein,ripple,jzsimon,sas}@isr.umd.edu

1 Introduction

There is a fundamental paradox lurking in the characterization of cortical response dynamics. On the one hand, it has long been accepted that cortical cells are sluggish and fail to follow sustained repetitive stimuli at rates much beyond 20 Hz (Kowalski, Depireux, and Shamma 1996; Miller, Escabí, Read, and Schreiner 2002). On the other hand, numerous studies have demonstrated a remarkable temporal precision of spike occurrences that are locked to stimulus onsets and other transients, and have considered it functionally significant (Bair and Koch 1996; Heil 1997).

These two phenomena have generally been studied separately using different stimuli that tend to highlight one phenomenon or the other; e.g., AM tones and noise, ripples, and click trains versus tone onsets and dynamic dots (Bair et al. 1996; Heil 1997). It is, however, possible to demonstrate the *coexistence* of these two response properties, and explore their limits and characteristics with stimuli that combine *both* repetitive and transient aspects. In this report, we describe how ripples (a broadband *frozen* noise or a harmonic series with various spectro-temporally modulated envelopes) can be used to elicit responses phase-locked both to the modulation envelopes and to the "texture" of the carrier. By independently manipulating these two aspects of the stimulus, it is possible to explore (1) the dependence of the precise firings on the nature of the stimulus, (2) the mechanisms that may give rise to these finely-structured responses, and (3) their functional significance.

2 Methods

Data were collected from extra-cellular cortical recordings in a total of 8 domestic ferrets (*Mostela putorius*). Five were in awake state, and the remainder were ketamine anesthetized. Details of the surgery are as in (Kowalski et al. 1996).

Stimuli included various combinations of moving ripples that last 3 seconds. Ripples are broadband complex sounds with periodically modulated spectral envelopes, explained in more detail in (Kowalski et al. 1996). We used specific combinations of ripples referred to as TORCs (Temporally-Orthogonal Ripple Combinations) to characterize the Spectro-Temporal Receptive Fields (STRFs) of cortical neurons. A TORC typically consists of 501 random-phase tones equally-spaced along the tonotopic axis, and spanning a range of 5 octaves. These tones form an instance of *frozen* noise, whose envelope is modulated by 30 different spectro-temporal waveforms with rates up to 24 Hz and spectral densities up to 1.4 cycle/octave, as described in detail in (Klein, Depireux, Simon, and Shamma 2000).

Hence, ripple stimuli have two distinct aspects, which are better described by the cochlear-like spectrogram shown in Fig. 1: (1) A prescribed spectro-temporal envelope (top trace in right panel) which allows us to estimate rapidly and accurately the STRF (using a reverse correlation technique (Klein et al. 2000)); and (2) A *Fine-structure* that carries the envelope (thick black trace in right middle panel). It is created by the interaction between the tones and can be extracted by a Hilbert transform. The fine-structure waveform depends solely on the carrier tone characteristics (frequencies, phases and amplitudes), and is limited in bandwidth to that of the cochlear filter, becoming progressively broader at higher filter frequencies. An additional variant of the TORC stimulus was used. It consisted of harmonic-TORCs whose spectro-temporal envelopes were carried by sets harmonically-spaced tones with fundamental frequencies between 25 and 200 Hz. In all other aspects, the harmonic- and regular-TORCs shared identical spectro-temporal envelope parameters.

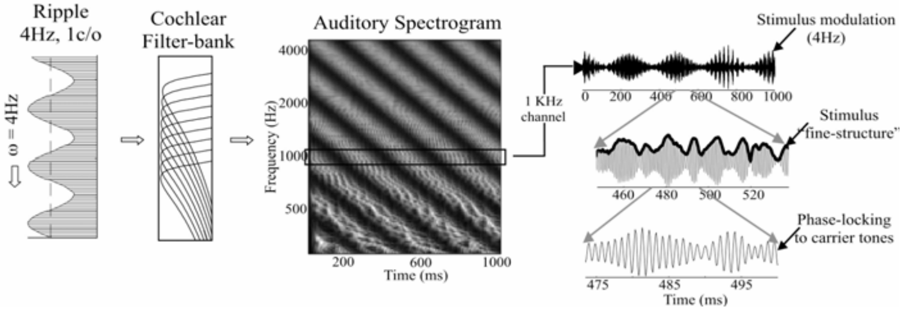


Fig. 1. Schematic of stimulus envelope and fine structure. **Left:** A ripple stimulus (4Hz and 1 c/o) is given as input to a filter-bank. **Middle:** The filters outputs show an overall pattern of a 4Hz drifting spectrogram. **Right:** The output of the 1KHz channel reveals the 4Hz envelope modulating a more dense fine structure carrier.

Neural responses to a series of TORCs are shown in the rasters of Figure 2. To quantify the precision of spiking, we computed the average cross-correlation of spike trains of different stimulus presentations (Fig. 2(B)), and then fitted it to a model of Poisson point-process cross-correlations that includes parameters to account for timing-jitter and spike deletion (Fig. 3(A)). Specifically, we assumed a Gaussian spread of the correlation peak whose variance σ represents the timing jitter and scale α (0<α<1) represents spike deletion between one trial and another. Combining these two parameters together with the rate of the process λ, we obtain:

$$R(\tau) = \lambda^2 + \frac{\alpha\lambda}{\sigma\sqrt{2\pi}} e^{-\tau^2/\sigma^2} \tag{1}$$

Finally, we employed the reverse correlation technique (Klein et al. 2000) to measure: (1) the usual STRF of the unit with respect to the spectro-temporal envelopes of the TORCs, (2) the STRF^C with respect to the *complete* cochlear filter-bank output (i.e., including both the envelopes and fine-structure); and (3) the STRF^F with respect to the fine-structure only (all shown Fig. 2(C)). In the last case, we averaged over all cochlear filter-bank outputs of the TORC stimuli to null out the spectro-temporal envelope and preserved only the fine-structure of the spectrogram to construct STRF^F.

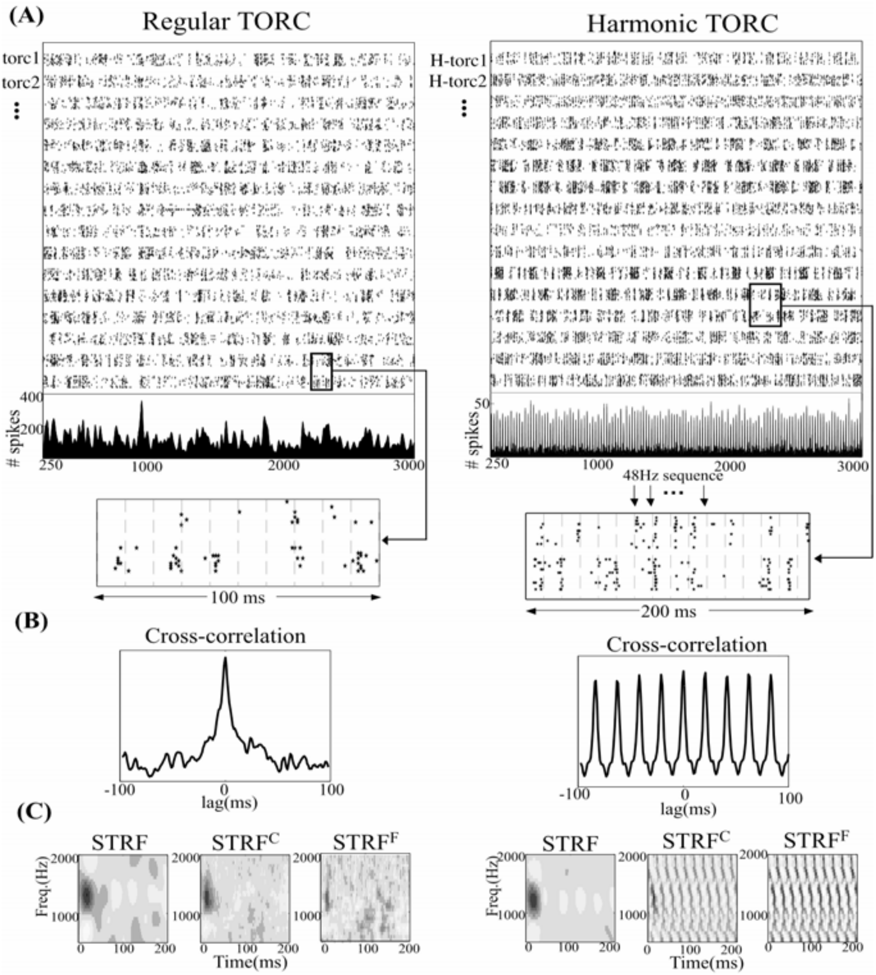


Fig. 2. Data analysis using regular (left panels), and harmonic TORCs (right panels).

3 Results

Data analyzed here were based on a total of 680 units (50% from awake ferrets). Figure 2(A) illustrates the nature of the precise spiking observed in the raster of TORC and harmonic-TORC responses. Specifically, this unit phase-locks to the fine-structure common to all the TORC stimuli with spikes that appear vertically aligned. To highlight this property, responses to all TORCs are collapsed to generate the PST histograms shown below each raster. Both histograms display strong and precise firing episodes (peaks) at numerous instants throughout the extended duration of the stimuli (3 seconds). In the case of the 48 Hz harmonic-TORC, the peaks occur regularly, reflecting the periodicity of the fine-structure.

To assess the degree of precision in the phase-locked responses, we computed the averaged cross-correlation among all TORC responses as defined by Eq 1 above. The resultant correlation functions shown in Fig. 2(B) display a sharp peak at zero lag (width of approximately 2-3 milliseconds) due to the high precision of firing from one response to another. The correlation function of the harmonic-TORC responses is periodic (right panel) demonstrating the precise phase-locking to the 48 Hz periodicity of the fine-structure. Using the Poisson model of unit responses, we estimated the distribution of σ, α, and λ for all units as shown in Fig. 3(B). Over 50% of all units exhibit relatively precise locking to the fine structure with $\sigma<10$ms. The skewness of the α distribution towards 0 suggests that spike deletion is a common property among most responses, partly because the different TORC envelopes are uncorrelated and hence spikes are suppressed differently from one TORC response to another (see discussion later). Finally, we have found that precision of spiking (σ distribution) is very similar in both the awake and the anesthetized ferret.

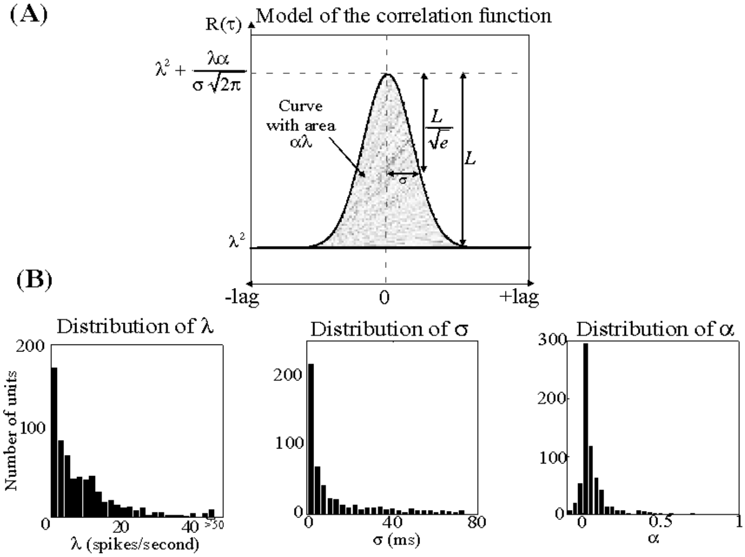


Fig. 3. **(A)** Model of spike correlation. **(B)** Population statistics for entire data set ($N = 680$).

The precision and speed of phase-locking is most readily seen in the periodic fine-structure of the responses to harmonic-TORCs. We have observed phase-locked responses over a wide range of fundamental frequencies up to 200Hz, and further testing could shed more light on the upper limit of this locking, and the possible tuning of neuronal responses to different fundamentals.

The fine-structure of cortical responses at a given BF is strongly related to the "Hilbert" envelope of the cochlear filter responses near the same CF. To demonstrate this relationship, we computed the correlation between the PST histogram of each TORC response and the averaged Hilbert envelopes of the cochlear filter-bank responses to the TORCs. The results are shown in the panels labeled $STRF^F$ in Fig. 2(C). In the case of regular TORCs, the correlation maximum occurs at the cochlear CF that corresponds to the BF of the cell. It has a latency of approximately 15 ms and is temporally compact indicating precise and rapid locking to the cochlear output at that CF. This finding suggests that relatively fast temporal modulations in auditory-nerve responses are preserved through four or more synapses all the way up to the cortex. We have found correlation functions such as this in about 66% of the 340 cells that exhibited precise firings (σ<10ms). The absence of this correlation in otherwise precisely firing cells may be due to substantial convergence of cochlear channels (e.g., in broadly tuned cells), or other more elaborate linear or nonlinear transformations that alter the cochlear envelopes prior to the cortical stage. Since the auto-correlation of the cochlear modulations is concentrated around zero-lag, the correlation functions in Fig. 2(C) can also be interpreted as the effective STRF of cortical cells to these modulations (denoted by $STRF^F$ in methods).

The relation between the STRF, $STRF^C$ and $STRF^F$ is illustrated in Fig. 4. The $STRF^F$ is compared to the *regular* STRF computed from the TORC envelopes only (i.e., disregarding the fine structure; see methods), and to the $STRF^C$ computed from the cochlear outputs to each TORC (i.e., taking into account responses to TORC envelopes *and* filter-bank output fine-structure). The four examples shown illustrate the wide range of response variability observed in our experimental data. For instance, while all STRF, $STRF^C$, and $STRF^F$ of a given neuron share roughly the same BF, there is a drastic difference between the slow dynamics of the STRF and the rapid onsets of the $STRF^F$, as in units A and B. In both these units, the $STRF^C$ is intermediate, in that it combines features of the STRF and $STRF^F$. In some cases as in unit C, the $STRF^F$ is very weak or absent leading the $STRF^C$ to resemble closely the STRF. Finally, in many cases as in unit D, the STRF and $STRF^F$ may not resemble each other closely suggesting substantial transformation of processing at the cortical level.

The STRFs measured with harmonic TORCs are often virtually identical to those obtained with regular TORCs since both stimuli share the same envelope parameters (e.g. Fig. 2(C)). However, since the carrier tones of the harmonic-TORC constitute a harmonic series, the response fine-structure to the harmonic TORCs is (as expected) limited to one periodicity regardless of CF (that of the fundamental). Consequently, the fine-structure of the cortical response is well correlated with cochlear filter-bank envelopes over all CFs, and does not reflect the STRF of the cell in a meaningful way (e.g., the striped $STRF^F$ in Fig. 2(C)).

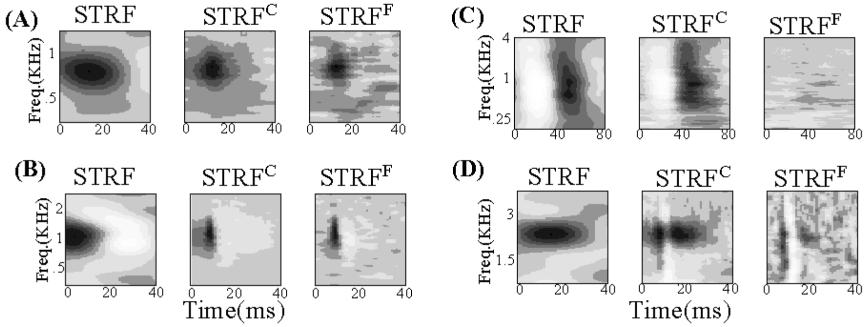


Fig. 4. Examples of STRFs. Each triplet corresponds to the STRF, STRF^C, and STRF^F of the same neuron.

So what is the relationship between cortical cells' fine-structure responses (STRF^F) and TORC-following responses (STRF)? And why do cortical cells phase-lock well to fast cochlear envelopes (likely more than 200 Hz), yet are incapable of following envelope modulations much beyond 20 Hz?

To elucidate this issue, we examined the hypothesis that the TORC-envelope acts as a gain that *gates* the responses to the fast underlying cochlear modulations. To test this hypothesis, we compared unit responses to those predicted from its STRF and STRF^F. Figure 5 illustrates the approach for the unit already discussed in Fig. 2. Using regular TORCs, we measured the unit's STRF and STRF^F (Fig. 2(C), left panel) and then used them to predict the responses to two 48Hz harmonic-TORC stimuli. The results are depicted by three curves in each panel. The *solid black* line is a smoothed period histogram of the actual response to this harmonic-TORC. The *dashed gray* is the predicted response based only on the unit's STRF. This curve captures the broad slow fluctuations in the response due to the TORC envelope, and completely ignores the response fine-structure. The STRF^F predictions are not shown but they consist simply of a train of 48 Hz peaks whose locations are indicated by the arrows in the figure. The *solid gray* line is the *product* of the STRF and STRF^F predictions. It tracks the actual response fairly well, giving support to the hypothesis that the response is essentially a *modulated fine-structure*. That is, the fine-structure is only visible when the envelope of the response fluctuates sufficiently strongly to reveal it; otherwise it is suppressed. This conjecture is completely consistent with the known effects of *synaptic depression* in thalamo-cortical pathway (Markam and Tsodyks 1996), and may explain the paradox of a sluggish, yet precise cortical response.

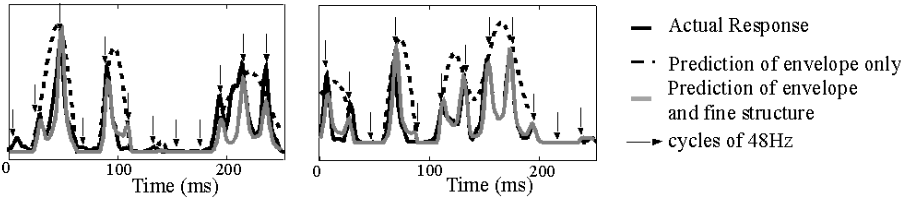


Fig. 5. Predictions of the responses of cortical neurons to a 48Hz harmonic TORC.

4 Discussion

We have demonstrated that cortical cells can phase-lock to cochlear envelope modulations with millisecond accuracy, with rates extending up to 200 Hz, and over sustained periods of time. While these response properties are common in the earlier thalamic and inferior collicular stages, they become highly constrained in the cortex. It becomes necessary to modulate the overall (spectro-temporal) envelope of the stimulus in order to reveal them. For instance, it is difficult to get rapid phase-locked responses to *sustained* simple tones, AM tones or complex tones. Rather, precise and phase-locked firings are largely confined to the onset of stimuli where the envelope rises rapidly (Heil 1997; Wallace, Shackleton, and Palmer 2002). These findings are consistent with the known properties of synaptic depression in thalamo-cortical synapses (Markram et al. 1996). To overcome its effects, it is necessary to modulate the stimulus strength (e.g., to turn it off and on) in order to allow the synapse to recover. The time-constant of this recovery is rather slow and explains why cortical responses in general are sluggish (<20 Hz) in following repetitive stimuli (Eggermont 2002). Therefore, allowing for periodic recovery, a non-depressed synapse is capable of conveying fast modulations and eliciting precise spiking over sustained periods, as is the case with the TORC stimuli.

Acknowledgments

This work is supported by the Office of Naval Research (grant N00014-97-1-0501), NIDCD (training grant DC00046-01), and NIH (grant DC05019-01A1).

References

Bair, W. and Koch, C. (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Comput. 8, 1185-1202.

Eggermont, J.J. (2002) Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms. J. Neurophysiol. 87, 305-321.

Heil, P. (1997) Auditory cortical onset responses revisited. I. First-spike timing. J. Neurophysiol. 77, 2616-2641.

Klein, D.J., Depireux, D.A., Simon, J.Z., and Shamma, S.A. (2000) Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. J. Comput. Neurosci. 9, 85-111.

Kowalski, N., Depireux, D.A., and Shamma, S.A. (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J. Neurophysiol. 76, 3503-3523.

Markram, H. and Tsodyks, M. (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature. 382, 807-810.

Miller, L.M., Escabí, M.A., Read, H.L. and Schreiner, C.E. (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 87, 516-527.

Wallace, M.N., Shackleton T.M., and Palmer, A.R. (2002) Phase-locked responses to pure tones in primary auditory cortex. Hear. Res. 172, 160-171.

Comments

Comment by Wang:

Your paper seems to ignore the evidence from other studies that auditory cortex neurons have different discharge patterns in response to time-varying signals when animals are awake or anesthetized (e.g., Goldstein et al. 1959; Lu and Wang 2000; Lu et al. 2001). In particular, we have shown in recent studies that neurons in the awake marmoset auditory cortex exhibit prominent discharges that are not time-locked to stimulus waveform or envelope at rapid modulate rates (Lu et al. 2001; Liang et al. 2002). Such rate responses may not be adequately captured by your analyses that lumped together data from both anesthetized and awake preparations. In your data collected from awake ferrets, what is the percentage of units from which you were able to produce significant STRFs? For those units that did not yield STRFs, were they driven by your stimuli?

Goldstein M.H. Jr, Kiang N.Y.S., and Brown R.M. (1959) Response of the auditory cortex to repetitive acoustic stimuli. J Acoust Soc Am 31, 356–364.

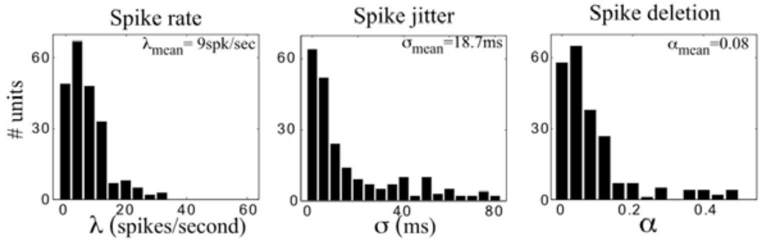
Lu, T. and Wang, X. (2000) Temporal discharge patterns evoked by rapid sequences of wide- and narrow-band clicks in the primary auditory cortex of cat. J. Neurophysiol. 84, 236-246.

Lu T., Liang L. and Wang X. (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci. 4, 1131-1138.

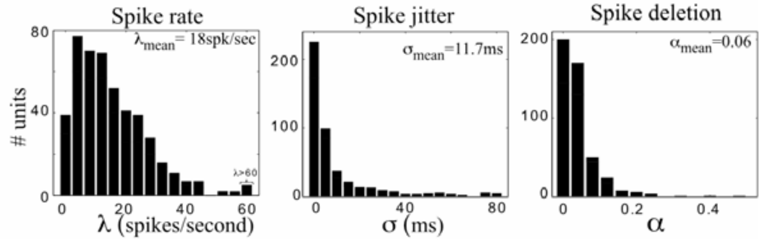
Liang, L., Lu, T. and Wang, X. (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J. Neurophysiol. 87, 2237-2261.

Reply:

It is indisputable that neuronal responses in A1 exhibit various differences under awake and anesthetized conditions. Other than a more vigorous spike rate (Fig. A1, leftmost panel), awake recordings exhibit a more complex spectro-temporal response pattern captured by their spectro-temporal receptive fields (STRFs). Nevertheless, there are substantial similarities between the two populations. In our data set, we observe that TORC stimuli possess a strong response-driving capability under both anesthetized and awake conditions. This property enables us to successfully derive STRFs from over 2/3 of the cells. Particularly, using a basic estimate of reliability of cells' responsiveness to TORC spectro-temporal modulations, we find that a reliable response (and STRF measurement) can be obtained from 70% of awake and 69% of anesthetized units (from a total of 473 awake and 341 anesthetized single units). These findings of comparable STRF reliability under the two conditions translate also to the analysis of temporal precision in cortical responses. The distribution of precision parameters in awake and anesthetized populations are shown below:



A. Anesthetized Experiments



B. Awake Experiments

Fig. A1. Distribution of correlation model parameters.

The population distributions, shown in Fig. A1, exhibit apparent similarities between the anesthetized and awake experiments. The distribution of spike deletion (α) is strikingly similar in the two populations (rightmost panels). Additionally, spike jitter (σ) is very comparable, despite a slightly heavier tail in the anesthetized compared to the awake condition (middle panels). This slight bias in the awake condition does not refute our overall finding of remarkable accuracy of phase-locking to the stimulus fine-structure. The similarities between two distributions are indeed more profound, with a mean of 18.7 vs. 11.7 ms in a range of 0–80 ms.

Overall, the striking similarities between the anesthetized and awake populations justify our analysis of temporal precision under the two conditions without distinguishing between experimental setups.

Comment by de Cheveigné and Pressnitzer:

Maass (1997, 1998) has shown that formal neural networks based on spike times are more powerful (in terms of the number of neurons required for a given task) than networks that use spike rate to represent data. The representation is also efficient in that the amount of information carried by a spike is limited only by its temporal accuracy. Thorpe et al. (1996) argue that such a representation is used in the visual system, and the same principle makes sense in hearing, an eminently temporal sense. One could make the conjecture that sound properties are coded as relative spike times within a group of neurons. For example, a reference spike might be triggered within a neuron sensitive to the *occurrence* of events (regardless of their nature) while specific properties would be coded by the times of subsequent spikes within the group. If a stationary property such as periodicity is coded in this way, the code need not be repeated unless the property changes. For example one does not expect the cortical response to a periodic tone to be periodic. Interspike time intervals within the group must be accurate, and the emission of the group of spikes is expected to follow reliably (and possibly with fixed latency) the sound event that it codes. If so, this coding principle might produce a pattern of cortical responses similar to that reported in your paper. The sparse nature of a temporal code fits well also with Barlow's (1961) principle of "redundancy reduction" (see comment to Nelken et al., this volume).

Barlow, H.B. (1961) Possible principles underlying the transformations of sensory messages. In Rosenblith WA (Ed) *Sensory Communication.* Cambridge Mass: MIT Press, 217-234.
Maass, W. (1997) Networks of spiking neurons: the third generation of neural network models. Neural Networks 10, 1659-1671.
Maass, W. (1998) On the role of time and space in neural computation. Lecture Notes in Computer Science 1450, 72-83.
Thorpe, S., Fize, D., and Marlot, C. (1996) Speed of processing in the human visual system. Nature 381, 520-522.

Comment by Nelken on the comment by de Cheveigné and Pressnitzer:

In general, I agree with this comment. In the responses of cortical neurons to some sounds, there are components that are highly precise but which may appear at a considerable latency after the onset of the sound (e.g. Bar-Yosef et al. 2002). These could serve the role of the spike groups you mention in your comment. However, the situation is really more complicated than that. Whereas some cortical spikes are highly precise, others seem to be much more variable in their timing. This is true even when using "frozen" stimuli. Thus, sound coding in auditory cortex have both highly precise and highly variable spikes, and both are informative (when decoded at the appropriate time scales). The multiplicity of time constants apparent in cortical activity is probably one of its most difficult aspects, both for experimental and theoretical treatment (e.g. Nelken et al. 2003).

Bar-Yosef, O., Rotman, Y., Nelken, I. (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J. Neurosci. 22 8619-8632.
Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., Farkas, D. (2003) Primary auditory cortex of cats: feature detection or something else? Biol. Cybern. 89,397-406.