Local Field Potentials and Spiking Activity in the Primary Auditory Cortex in Response to Social Calls

Andrei V. Medvedev and Jagmeet S. Kanwal
Department of Physiology and Biophysics, Georgetown University, Washington, DC 20057

Submitted 23 December 2003; accepted in final form 9 February 2004

Medvedev, Andrei V. and Jagmeet S. Kanwal. Local field potentials and spiking activity in the primary auditory cortex in response to social calls. J Neurophysiol 92: 52–65, 2004; 10.1152/jn.01253.2003. The mustached bat, Pteronotus parnellii, uses complex communication sounds (“calls”) for social interactions. We recorded both event-related local field potentials (LFPs) and single/few-unit (SU) spike activity from the same electrode in the posterior region of the primary auditory cortex (Alp) during presentation of simple syllabic calls to awake bats. Temporal properties of the LFPs, which reflect activity within local neuronal clusters, and spike discharges from SUs were studied at 138 recording sites in six bats using seven variants each of 14 simple syllables presented at intensity levels of 40–90 dB SPL. There was no clear spatial selectivity to different call types within the Alp area. Rather, as shown previously, single units responded to multiple call types with similar values of the peak response rate in the peri-stimulus time histogram (PSTH). The LFPs and SUs, however, showed a rich temporal structure that was unique for each call type. Multidimensional scaling (MDS) of the averaged waveforms of call-evoked LFPs and PSTHs revealed that calls were better segregated in the two-dimensional space based on the LFP compared with the PSTH data. A representation within the “LFP-space” revealed that one of the dimensions correlated with the predominant and fundamental frequency of a call. The other dimension showed a high correlation with “harmonic complexity” (“fine” spectral structure of a call). We suggest that the temporal pattern of LFP and spiking activity reflects call-specific dynamics at any locus within the Alp area. This dynamic contributes to a distributed (population-based) representation of calls. Alternatively stated, the fundamental frequency and harmonic structure of calls, and not the recording location within the Alp, determines the temporal structure of the call-evoked LFP.

INTRODUCTION

Auditory signals are important for social interactions in many species because they provide a fast and effective means of communication between conspecifics. Frogs, dolphins, whales, bats, some nonhuman primates, and many species of birds, such as songbirds, are only a few examples of those species that use complex sounds for social communication (Brenowitz et al. 1984; Busnel 1963; Fenton 1977; Green 1974; Kanwal et al. 1994; Margoliash 1983; Margoliash and Fortune 1992; Owren 1990; Payne and McVay 1971; Rado et al. 1998; Rauschecker et al. 1995; Rowell and Hinde 1961). Neural mechanisms and coding schemes, which underlie effective transmission and processing of auditory information in animals and humans are poorly understood. While it is virtually impossible to directly study neural mechanisms for processing of communication sounds, e.g., speech in humans, animal models may provide invaluable information about common mechanisms for perception of communication sounds.

The auditory system of bats has been extensively studied by several groups of researchers, both to understand their echolocation (Casseday et al. 1994; Grinnell 1963; Suga 1965; Suga and Taniguchi 1983) and their ability to use complex communication sounds for social interactions (Fenton 1977; Kanwal 1999; Kanwal et al. 1994, 2000; Leippert 1994; Porter 1979). Even though many well-structured communication sounds (or simply “calls”) in different species of bats have been described in terms of their behavioral context, the neural mechanisms of their processing are not well understood. A detailed classification of the acoustic structure of simple syllables and composites in the mustached bat, Pteronotus parnellii, has been described previously (Kanwal et al. 1994). Single-unit (SU) studies of neural processing of the simple syllables and composites reveal that neurons in the Doppler-shifted constant frequency (DSCF) and FM-FM areas of the primary auditory cortex are dually specialized for both, processing of echolocation pulse and its echo (Suga 1965) and calls (Kanwal 1999; Kanwal et al. 1994). The stimulus-locked response of these neurons is combination-sensitive, i.e., facilitated when presented with the combination of constant frequency (CF) tones at their two best frequencies (BFlow and BFhigh). In contrast to neurons in the DSCF area, neurons in the anterior (Ala) and the posterior (Alp) parts of the primary auditory cortex of the mustached bat are considered as relatively unspecialized and arranged tonotopically along the antero-posterior axis. Our recent study has revealed that Alp neurons also show multipeaked excitatory frequency tuning to harmonic frequencies (Peng and Kanwal 2001). The best frequencies of Alp neurons, however, do not correspond to combinations of pulse and echo frequencies, as is the case for the DSCF neurons, but are harmonically related (Peng and Kanwal 2001). Distribution of the best frequencies over many Alp neurons shows prominent peaks at BF1 = 24 kHz and BF2 = 48 kHz. These frequencies are commonly present in the power spectra of many calls. Accordingly, single neurons in the Alp area respond to several different types of calls and their variants. Here we examine call specificity as revealed not in the peak response rate of single neurons, but in the temporal pattern of activity of both local field potentials (LFPs) and SU activity in the Alp area in response to the presentation of best variants of different call types. The unspecialized Alp area in mustached bats can be easily homologized to the primary auditory cortex of the mustached bat.
auditory cortical areas of other mammalian species on the basis of
topological, hodological, and functional criteria; therefore
the results obtained from this study should be generally useful
to understand call processing in the primary auditory cortex of
many other highly vocal mammals, including humans.

METHODS

The surgery, acoustic stimulation methods, and recording of elec-
trophysiological activity for these experiments are similar to those
described previously (Kanwal et al. 1999) and therefore are only
briefly presented here.

Surgery and recording of neural activity

Mustached bats, P. pammellii, were caught in Trinidad and trans-
ported to the Animal Care Facility at Georgetown University. Under
isoflurane/air mixture (medical grade, Anaquest) anesthesia, a skin
incision was made at the mid-line on the head, and a 2-mm-diameter
metal post was affixed just behind the intersection of the sagittal and
coronal sutures with cyanoacrylic glue (Locite 411). Bats were
allowed to recover for ≥3 days before the first recording session.

During electrophysiological recordings, the head of the bat was
restrained by clamping the metal post, and the bat’s body was sus-
pended in a Styrofoam mold by elastic bands in a heated (31°C),
sound-proofed, and echo-attenuated chamber (IAC 400A). The bat
was monitored throughout the experiment with a video camera. The
electrophysiological activity was recorded with sharpened, vinyl-
coated tungsten-microelectrodes with tip diameters of ~10 μm and
impedance of ~1 MΩ inserted into the cortex perpendicular to the
skull to a depth of 300–600 μm through a small (50 μm) hole. An
indifferent tungsten-wire electrode was placed on the dura mater near
the recording electrode. To record neuronal action potentials, the
signal was amplified by an AC preamplifier and band-pass filtered
between 600 Hz and 4 kHz. LFPs were obtained from the same
electrode with a low band-pass filter (1–300 Hz). Single units were
isolated using graphical overlay of up to four time-amplitude accep-
tance/rejection windows on the recorded waveforms, allowing selec-
tion of one specific waveform. All surgeries and animal experiments
were performed with approval of the Georgetown University Animal
Care and Use Committee.

Sound stimuli used in this study were 30-ms-long CF tones and
digitized calls of the same species. CF tones were presented from two
condenser loudspeakers to determine the BFs of neurons at the re-
cording site and the corresponding auditory thresholds as described
elsewhere (Kanwal et al., 1999). Tone bursts A and B corresponded
to a low BF (20–29 kHz) and a high BF (40–58 kHz) of a neuron,
respectively, and were presented at a rate of 4/s singly as well as
simultaneously to determine whether a neuron shows nonlinear facil-
itation (“combination sensitivity”) for tonal combinations. The values
of BFs, the approximate harmonic relationship between the BFs
(within 1–2 kHz), and a lack of combination sensitivity were used to
distinguish Alp neurons from DSCF neurons. To deliver the natural
calls and their variants (ranging in duration from 4 to 89 ms) at a rate
of 1/s, another computer (Pentium PC) with an A/D-D/A board (DT
2821G) and a sound generation program SIGNAL were used. The
main set of 14 call types consisted of calls emitted by a small, captive
population of bats (Kanwal et al. 1994). Calls were assigned numbers
from 1 to 14 based on their acoustic classification via multidimen-
sional scaling (MDS) (Kanwal et al. 1994). Frequency variants of
these “standard” calls were synthesized on the computer by shifting
the spectrum of the mean waveform up and down by 1–3 SD of the
mean fundamental frequency as measured across the bat population
(Kanwal et al. 1994). Mean waveforms represent an average out of
several hundred examples of each call type obtained from a colony of
bats. CF tone bursts were used to identify the BF, and a “call-scan”
was performed to identify the best calls at each recording site. During
the first stage of this scan, a 15-s epoch of activity (no stimulus control
plus 14 calls presented at a rate of 1 call/s) was repeatedly recorded 30
times. After every 10 consecutive presentations of the same call at the
same level of loudness, the sound intensity was automatically de-
creased by 25 dB. Thus three average levels of intensity (95, 70, and
45 dB) were tested initially. Three-step intensity screening was done
using all seven sets of calls (the “standard” call and its six variants
with spectra shifted by ±1, ±2, and ±3 SD of a call’s fundamental
frequency). The SU data from this initial screening were analyzed
on-line to reveal the best variant for each call type. At the second stage
of the screening process, a newly configured set of the best variants
for each of the 14 call types was presented at six different levels of
intensity (attenuation steps of 10 dB). As a result of the two-step
screening procedure, call preference was determined for every record-
ning site; this allowed us to further analyze the response of three to five
best call variants at their optimal level of sound intensity. Each call
was also presented to the animal in the reverse order, and the corre-
sponding LFPs as well as SU responses were recorded. Background
activity was also recorded for 1 s before each repetition of the auditory
stimulus set.

Data analysis

Peri-stimulus time histograms (PSTHs) were calculated for 200 or
100 presentations of acoustic stimuli using a bin width of 1 ms. These
histograms were used to measure the neuronal response that is tradi-
tionally defined as a stimulus-locked change in the peak response
magnitude, or firing rate, and which can be detected by summation of
spike trains over repeated stimulus presentations. For each call type,
the LFPs were first analyzed within each animal to assess the degree
of variability across animals as well as at different depths within the
cortex (100–600 μm). After establishing a similar profile of the
call-evoked LFP for all the recorded Alp sites in different bats, both
types of the stimulus-locked responses (LFPs and PSTHs) were aver-
egaged across recording sites and all animals for each call type. The
call-evoked averaged responses were compared with each other using
t-tests (P < 0.01). To obtain an estimate of the duration of LFPs and
response histograms, the “center of gravity,” or median point, tmed,
for each averaged LFP and PSTH was determined as a time point where
the area under the LFP waveform or PSTH from response onset to that
point equals the area to the right of that point: LFP/PSTHarea (t <
tmed) = LFP/PSTHarea (t > tmed). Duration of the averaged LFPs or
PSTH was taken as twice the median point. Pair-wise correlations
between the averaged call-evoked LFPs were calculated within a time
window 0–250 ms, the time period of the early and middle phases of
a classical middle-latency auditory evoked potential (Barth and Di 1990).
Those cross-correlation coefficients were analyzed through the
MDS algorithm (SPSS software) to establish a relationship between
the temporal structure of the LFP waveforms and the acoustic struc-
ture of calls.

To quantify harmonic structure of calls regardless of their fre-
cquency range, we calculated normalized spectra of call waveforms
(sn), using a normalized frequency scale (fn), such that sn(fn) = sn(fn
max ), where f is frequency in kilohertz, s(fn) is a power spectrum of
a call waveform, and sn(max) and fn(max) are the maximum
power and the corresponding predominant frequency within s(fn).
Normalized call spectra sn(fn) thus show relative distribution of power
over frequency values that are also relative to the predominant
frequency. For each normalized call spectrum sn(fn), the fol-
lowing parameters were calculated: the “effective” frequency, f eff = 2(fn
max × sn(fn max )/Σ(fn max × sn(fn max ))) and its “deviation,” dev(f eff) = [Σ(fn
max × sn(fn max )– f eff × sn(fn max ))/Σ(fn max × sn(fn max ))] that characterizes dispersion of frequencies around f eff within
the normalized spectrum sn(fn).

In total, 138 cortical sites in the left hemisphere of six bats were
studied. Figure 1 shows locations of the recorded sites within the Alp
area of all bats. As many as 10–12 penetrations were made in a bat
within an \(~1.5 \times 1\)-mm rectangular area in the central part (50–75\%) of the Alp area. For each penetration, the LFP and SU responses from one to three recording sites were used in the analysis.

RESULTS

Call-evoked LFPs and SU activity within the Alp area

Responses of local neuronal populations at each recording site were obtained by the same stimulus set of simple syllabic calls and their variants. As an example, Fig. 2A shows the LFPs and the corresponding PSTHs for SU activity for two neurons at the same recording site in response to presentation of calls 3–10. SU responses of the two neurons to several different types of calls and their variants were recorded simultaneously. These neurons responded to many calls and showed only a weak preference for call types 4–9 when scanning the response to 14 different call types (calls 1–14). Typically, the overall temporal pattern of SU responses showed a correlation to the temporal pattern of the LFP waveform. Thus the PSTHs show increases or decreases of neuronal firing coinciding with specific components of the LFP. Nevertheless, because SU activity is usually sparse, not every peak or wave in the LFP is accompanied by a distinctive change in spike activity in the individual examples of SU activity. Note that the LFPs have different waveforms for different call types. Also, responses of the SUs recorded from the same electrode showed slightly different latencies, peak rates, and durations. Thus the major feature of single unit responses and LFP’s in the Alp area is their ability to respond to many calls showing preference to more than one call. Sound intensity had a marginal effect on the response profile of Alp neurons (Fig. 3C). The best call variants and their best amplitude levels were determined from the peak response in the PSTHs via the two-step call scanning procedure. Typically, the best amplitudes for responses to calls ranged between 50 and 90 dB SPL. Bar graphs (Fig. 2B) show the best responses (call preference) for the same two units presented in Fig. 2A.

Recording of electrical activity at depths from 100 to 600 \(\mu\)m within the Alp area revealed that the call-evoked LFPs change polarity at depths between 300 and 400 \(\mu\)m (Fig. 3A).

From 400 to 600 \(\mu\)m of cortical depth, polarity did not change, and changes in the LFP were largely restricted to variations in amplitude of some positive and negative peaks, whereas the duration of those peaks remained approximately the same (Fig. 3A). Only the responses obtained at depths below 400 \(\mu\)m were taken into further analysis.

A comparison of the recordings from various Alp sites in the same bat versus from different bats showed that the generic temporal structure of the LFP for each call type within the Alp area was similar across different recording sites and different animals (Fig. 3B). Individual differences in LFPs between recording sites and animals were once again largely restricted to the amplitude of individual LFP components, whereas the temporal pattern of those components was similar (Fig. 3B). To detect some type of spatial gradient within neuronal responses over the Alp area (if any), we used a general linear model within the three-way ANOVA test. First, the most rostral recording site in each animal was taken as a reference. The responses (LFP or PSTH) from all other recording sites were correlated with the reference response, and those correlation coefficients were analyzed using the rostrocaudal and dorsoventral coordinates of recording sites as covariates and the animal number as a factor. The analysis performed for each call separately did not reveal any spatial gradient or differences between animals. For example, for call 4 (syllable sAFM): \(F = 1.74, P = 0.19, n = 19\); for call 7 (syllable dRFM): \(F = 2.69, P = 0.09, n = 14\). This allowed us to justify averaging of the LFPs recorded for each call type over all sites and all bats.

A statistical pair-wise comparison between averaged call-evoked LFPs is shown by two examples in Fig. 4. The first example shows LFPs for calls 3 and 8, which are the most different from each other as revealed by their lowest correlation coefficient (0.03) from all pair-wise comparisons between 14 call-evoked LFPs. The second example shows two LFPs with the highest correlation (0.86). All four averaged LFP waveforms have a relatively low variance (as determined by their SE), and the \(t\)-values calculated for each time point within the stimulus-locked response show significant differences \((P < 0.01)\) between the LFP waveforms. Although the high-ampli-
tude components are not observed later than 250 ms after the stimulus onset, the significant differences for pair-wise comparisons among 14 averaged call-evoked LFPs were observed for as long as 500–600 ms after the stimulus onset.

The averaged LFPs and the corresponding averaged PSTHs for all 14 call types are presented in Fig. 5, along with the amplitude envelops of calls. In this and other figures, calls are arranged according to their acoustic structure from CF sounds (calls 1–3) through FM sounds (calls 4–12) to “noise-burst” (NB) sounds (calls 13 and 14), as determined by pooling various parameters of the acoustic structure of calls and scaling them down to a single dimension (Kanwal et al. 1994). All LFPs have a distinctive temporal structure and contain a number of negative and positive peaks. The polarity, amplitude, and duration of these peaks are different for different call types. Nevertheless, some common features of the call-induced LFPs can be mentioned here. As a rule, the LFP starts with a small positive deflection 10–20 ms after the stimulus onset (P1). This initial positive peak is followed by a series of other negative and positive components that are higher in amplitude and longer in duration. In some cases, a relatively slow peak/wave or a multipeak complex occurs 20–30 ms after the end of a call, which is negative in some calls (1, 7, 8, and 14) and positive in other calls (4, 12, and 13). Interestingly, this “off” component is often accompanied by an increase in neuronal firing (e.g., see calls 1, 4, 5, 7, 12, 13, and 14). The LFP is usually terminated by a slow “positive-negative wave” complex. The amplitude, duration, and time course of this late component are different for different calls. During the positive slow wave, there is a reduction in neuronal firing best seen in calls 3, 6, 7, 11, and 12. Depending on the duration of a call, the late positive-negative complex wears off at 300–500 ms after the stimulus onset. For calls with a periodic temporal structure, the LFPs show oscillations with a frequency corresponding to the rate of AM within a call (for calls 6 and 8, the AM rate is ~40 Hz; for call 10, the AM rate is 18 Hz). The periodic structure of the LFP for call 9 with a high AM rate (120 Hz), however, is obscured because of the small amplitude oscillations superimposed on other slower components of the LFP. Periodicity in the evoked activity corresponding to the periodic structure of a call is less obvious in the SU activity.

Although the LFP typically lasts longer than the corresponding call, the LFP duration does show a correlation with call duration, i.e., correlation coefficient between these values equals 0.83 ± 0.26 (SD) and is significantly different from zero ($P < 0.01$; Fig. 6A), whereas the PSTH duration does not correlate with call duration (Fig. 6B). Nevertheless, it is important to note that call duration is not the only factor determining duration of the evoked activity because the LFP duration varies significantly within groups of calls of similar duration. For example, calls 1, 5, and 13 are similarly very short (<20 ms), whereas the corresponding LFPs vary from 130 ms for call 5 to ~170 ms for calls 1 and 13. Call 6, which has the same duration as calls 12 and 14, has much shorter LFPs (170 vs. 260 ms; Fig. 6A).

**FIG. 2.** A: examples of the local field potentials (LFPs; top) and the corresponding single/few unit (SU) activity (bottom) for 2 AIp neurons isolated at 1 site and recorded during presentation of 8 calls labeled by a number and an abbreviated name (see Kanwal et al. 1994). Call waveforms and their spectrograms are shown below the electrophysiological response on the same time scale. B: bar graph of peak response values vs. call number (1–14) for the same 2 AIp neurons shown in A to the “best” statistical variants of 14 simple syllabic calls (i.e., variants eliciting greatest peak response; see METHODS for a detailed description of call variants and call scan procedure). Only responses exceeding the average background activity plus 1 SD are shown. In this example, 2 neurons respond to many variants with some preference for calls 4–9.
Effect of call reversal on the LFP

A comparison of the LFPs to the forward and time-reversed versions of the same call show that the call-evoked LFPs in the Alp area are not significantly affected by call reversal (Fig. 7). Although the amplitudes of some individual LFP components changed in a few cases, all 14 reversed calls evoked temporal patterns of activity very similar to the corresponding patterns of activity evoked by natural calls. Correlation coefficients between the corresponding LFPs in response to a natural call and a reversed call ranged between the lowest value of 0.57 (call 14) to the highest value of 0.94 (call 9). We also calculated cross-correlation functions between the call envelopes and the LFPs at time lags from −50 to 200 ms. Because time lags giving the highest “envelope–LFP” correlation vary for different call types, we took the absolute value of the highest correlation within a 0 to 100 ms time shift for each call type. Those peak values of the “envelope–LFP” correlations appeared to be preferentially in the range 0.2–0.4 for all calls and were similar for the following four comparisons, i.e., “normal call–LFP to normal call,” “reversed call–LFP to normal call,” “normal call–LFP to reversed call,” and “reversed call–LFP to reversed call.” The corresponding mean (±SD) values for four comparisons are 0.27 ± 0.08, 0.25 ± 0.09, 0.27 ± 0.08, and 0.25 ± 0.07 and are not significantly different from each other (1-way ANOVA; F = 0.56, P = 0.64, n = 56). Thus this analysis shows that the LFPs in the Alp area are relatively insensitive to call reversal. Whereas this result is not surprising...
for the CF calls whose spectral structure is invariant to time reversal, the same result for the FM calls with a well-defined and specific temporal structure was unexpected.

MDS of call responses

To establish a relationship between the unique temporal structure of the LFP evoked by a call and the acoustic structure of that call, we performed MDS of 14 call-evoked LFPs as well as 14 PSTHs averaged over many units/recording sites. Pair-wise correlations between the LFPs or PSTHs were calculated within a time window of 0 to 250 ms and were converted into “distances” by standard transformation as described by Mardia et al. (1979).

Two-dimensional scaling of the LFPs and PSTHs, based on their pair-wise correlations, captured 70 and 88% of variation, respectively. Again, this difference is due to the more uniform values of the PSTH correlations. We therefore limit our MDS analysis to two dimensions. This is equivalent to the usage of
the first two principal components (with the highest amplitudes) to capture the major features of the data pattern (Mardia et al. 1979). Segregation of call-evoked LFPs and PSTHs in the abstract representational space is shown in Fig. 8. As a result of two-dimensional scaling, all 14 call-evoked LFPs were well separated and arranged roughly along a circle (Fig. 8A), whereas the PSTHs corresponding to the same set of calls are placed closer to each other with two outliers (calls 2 and 14; Fig. 8B). It should be noted, however, that segregation of variables in the abstract space by the MDS analysis is relative and invariant to linear transformations, such as shift and rotation (Mardia et al. 1979). The LFP representation shows a better segregation of calls than that due to a smoothed PSTH representation. To quantify this result, we plotted a distribution of normalized pair-wise distances (ranging from 0 to 3) for all calls (points) from the coordinates provided by each MDS plot. An analysis of these distributions revealed a peak corresponding to a distance of >2 for the LFP representation and a peak corresponding to a distance of <1 for the PSTH representation. This provided an index of spatial segregation of calls for each representation and revealed a higher sensitivity of the LFPs compared with PSTHs to variations in the acoustic structure of different calls. Therefore a further analysis of the relationship between acoustic parameters of a call and its position in the

![Figure 7](image_url1)

**FIG. 7.** Line plots of the averaged LFPs recorded during presentation of natural calls (thick lines), and the same calls reversed in time (thin lines). For each call type, LFPs are averaged over those recording sites for which corresponding reversed calls were also presented (n = 5, 4, 10, 15, 8, 6, 16, 5, 10, 4, 4, 4, 2, and 2 for calls 1–14, respectively). Note the similar waveform and temporal profile of the LFPs to the normal and reversed calls. Amplitude is in millivolts.

![Figure 8](image_url2)

**FIG. 8.** Multidimensional scaling (MDS) representation of averaged call-evoked LFP waveforms (A) and averaged call-evoked PSTHs (B) based on their pair-wise correlations. Note a better segregation of calls in the “LFP-based” space.
Figure 9 shows the LFP-based MDS plot and the corresponding LFP waveforms placed along the perimeter of the MDS plot. On this plot, the call-evoked LFPs are arranged according to the similarity in their waveforms. For example, LFPs for calls 3, 4, 9, and 11 have a similar positive slow wave component at around 100 ms. This leads to a relatively high mutual correlation of those LFPs and their “proximity” within the MDS plot. LFPs for calls 7 and 13 also have a similar temporal profile of the slow component at 50–250 ms and thus those LFPs are placed close to each other. The LFPs for calls 1, 12, 6, and 10 in the top row of the MDS plot have high-amplitude multiple peaks at 10–100 ms, whereas the late slow-wave components are not prominent in this group of LFPs. When the LFP-based segregation of calls is analyzed in terms of the acoustic structure of calls (their spectrograms are shown in Fig. 9A), the following relationships can be delineated.

To quantify a relationship between the spectral content of a call and the placement of the corresponding LFP waveform in the MDS plot, we refer to the predominant frequency as “formant” and to the fundamental frequency as “pitch.” It should be noted that both frequencies are the same for several call types because the fundamental frequencies of these calls are also the highest in energy. In other calls, namely 3, 4, 7, 8, 9, and 11, the highest spectral peak occurs not at the fundamental frequency (by convention, called the “1st harmonic”) but at the next higher (“2nd”) harmonic. The first dimension (x values) of the averaged LFPs in the representational space correlate equally well with both the “formant” and the “pitch” frequencies of the corresponding calls (the slopes of the regression lines are significantly different from 0 by t-test, \( t = 3.0, P = 0.01 \) for both frequencies; Fig. 11A and B).

Second, the acoustic structure of calls changes along the y-axis (dimension 2) in the representational space. Calls in the lower part of Fig. 10A (low values along the y-axis) are of the CF type or have relatively weak FM components with a low number of harmonics (1–3) in their spectra (calls 2, 5, 13, 7, 3, and 4). At the midlevel along the y-axis, there are two calls with strong FM (calls 8 and 9) and a NB call 14. Finally, calls with a rich harmonic structure and periodic FM have a high value for the second dimension represented along the y-axis (calls 1, 12, 6, 10, and 11). The majority of these calls have a high number of harmonics in their spectra (“harmonic stack”). Thus, these changes in the acoustic structure of calls along the y-axis can be characterized as an increase in “harmonic complexity.” For example, CF calls with a few number of harmonics have a simpler harmonic structure compared with FM calls with a few harmonics but larger frequency variation in their spectra due to FM sweeps. Finally, calls with a high number of harmonics and/or periodic FM have the most complex harmonic structure. To quantify “harmonic complexity,” we use the normalized spectra of call waveforms \( s_n \) shown in Fig. 10B.

By definition, the maximum occurs at \( f_n = 1 \) in all spectra \( s_n(f_n) \). If the predominant frequency equals the fundamental frequency (“formant” = “pitch”), there is no spectral peak below \( f_n = 1 \) (calls 1, 2, 6, 12, 13, and 14); otherwise there is an additional peak at \( f_n = 0.5 \). “Harmonic stack” results in a “tail” of multiple small peaks at integer values of \( f_n = 2, 3, 4, \ldots \) (if “formant” equals “pitch” as it is the case for calls 1 and 6) or at values of \( f_n = 1.5, 2, 2.5, 3, \ldots \) otherwise (calls 10 and 11 in the top row of Fig. 10B). If a normalized call spectrum has only a few harmonics, the “effective” frequency
$f_{\text{eff}}$ is close to one and its dispersion $\text{dev}(f_{\text{eff}})$ is relatively low (also $\sim 1$). If a spectrum has a “harmonic stack,” the “effective” frequency $f_{\text{eff}}$ and its dispersion increase because of the contribution of multiple spectral peaks in the normalized spectrum. Therefore parameters $f_{\text{eff}}$ and $\text{dev}(f_{\text{eff}})$ can be used to quantify “harmonic complexity.” The $y$ value (dimension 2) in the

FIG. 10. Segregation of calls based on MDS of call-evoked LFPs. Spectrograms (A) and call normalized spectra (B) are shown around the perimeter of the MDS plot. Normalized spectra ($s_n$) are calculated using the normalized frequency scale ($f_n$), such as $s_n(f_n) = s(f) / s_{\text{max}}$, $f_n = f / f_{\text{max}}$, where $f$ is frequency in kilohertz, $s(f)$ is a power spectrum of a syllable waveform, and $s_{\text{max}}$ and $f_{\text{max}}$ are the maximum power and the corresponding frequency within $s(f)$. Note that calls are arranged along a circular configuration, and a relative position of a syllable along this circle reflects its harmonic structure. Calls placed on the right-bottom portion of the circle have a relatively simple harmonic structure with the main peak at $f_n = 1$ and 2 additional smaller peaks at $f_n = 0.5$ and $f_n = 1.5$ (calls 9, 4, 3, 7, and 13). In a clockwise direction along the circle from this group, calls with a few broad spectral peaks are placed (calls 13, 5, 2, and 8). Finally, calls with a rich harmonic structure (with a high number of harmonics and relatively broad spectral peaks; calls 1, 12, 6, 10, and 11) are placed on the top of the circle. The rectangular broadband noise-burst (rBNB; call 14) is placed in the middle of the circle.
The auditory system of the mustached bat is highly developed because of the unique auditory specialization of these animals. Previous studies on the neural mechanisms for echolocation and auditory communication in bats are based on an analysis of SU activity. Different properties of single neurons, including their combination-sensitivity to specific spectral and temporal parameters, in the functionally specialized areas, such as the DSCF, CF/CF, FM-FM, etc., in the primary and secondary auditory cortices of the bat are well studied (Casseday et al. 1994; Grinnell 1963; Ohlemiller et al. 1996; Suga 1965; Suga and Taniguchi 1983; Suga et al. 1978, 1979). We describe event-related LFPs in response to natural sounds produced by conspecifics and show that the LFPs are highly specific for simple calls that compose the acoustic “alphabet” of the vocal repertoire for communication in the mustached bat (Kanwal 1999; Kanwal et al. 1994). This suggests that the full spectrum of response dynamics at a recording locus is important for understanding call representation and processing within the cortex. The present analysis is based on pair-wise correlations that are sensitive to the temporal structure of responses ignoring their amplitudes. Our analysis shows that the PSTHs have a more similar time course and therefore the lower “call-specificity” compared with the LFPs (Fig. 8). A possible interpretation of this result is that the PSTHs are more sensitive to similarities present within the calls, whereas the LFPs are more sensitive to their differences. Thus peak response rates of single neurons in the Alp area are inadequate for discerning different call types and therefore provide a relatively poor resolution of the neural response space into which calls can be mapped, but could be used for distinguishing species-specific calls from noncall stimuli.

**DISCUSSION**

The representational space of the averaged LFP is significantly correlated with both parameters $f_{\text{eff}}$ and $\text{dev}(f_{\text{eff}})$ (t-test for regression slopes; for $f_{\text{eff}}: t = 2.9, P = 0.013$; for $\text{dev}(f_{\text{eff}}): t = 3.1, P < 0.01$; Fig. 11, C and D).

Some calls with apparently different spectral-temporal structure have similar normalized spectra, and they are placed close to each other in the two-dimensional space. An example includes calls 3, 4, 7, and 9 (group 1). Call 3 is a CF type of simple syllable, whereas calls 4, 7, and 9 are FM syllables. Moreover, call 7 has a periodic FM and call 9 has a single FM sweep. Despite these differences in the spectral-temporal structure, the normalized spectra of these calls are similar and have a narrow spectral peak at the fundamental frequency ($f_0$ = 1) and two additional small peaks at $f_0 = 0.5$ and $f_0 = 1.5$ (Fig. 10B). Peak at $f_0 = 0.5$ means that the predominant frequency within these calls is higher than the fundamental frequency. Moving clockwise along the circle in the representational space from this group of calls, we find calls with a low number of harmonics but with broader spectral peaks (calls 13, 5, 2, and 8; group 2; Fig. 10B).

The circular contour line in the representation space concludes with calls having a rich harmonic structure (relatively broad and multiple spectral peaks, such as in calls 1, 6, 10, 11, and 12; call 12 does not have multiple harmonics, although it has a relatively broad spectral peak at the fundamental frequency; group 3). In the majority of calls within the second and the third group, the peak at $f_0 = 0.5$ is absent because these calls have the highest energy at a relatively low fundamental frequency. The rectangular broadband noise-burst (call 14) is placed at the center of the circle. This call has a spectral “tail” at high frequencies because of its broadband spectrum (a characteristic feature of calls with a “harmonic stack”) but, unlike the majority of other calls, it lacks a discrete harmonic structure. Thus two-dimensional scaling of call-evoked responses allowed us to segregate calls according to their harmonic structure.

**FIG. 11.** Regression line plots for MDS dimensions and call parameters. x values (dimension 1) of calls in the 2-dimensional representational space correlate with the call predominant (“formant”) frequency (frequency with the highest energy in the call spectrum; A) as well as the call “pitch” (the fundamental frequency in the call spectrum; B). y values (dimension 2) of calls correlate with 2 parameters used to quantify “harmonic complexity” of calls, namely, the “effective” normalized frequency $f_{\text{eff}}$ (C) as well as its “deviation” $\text{dev}(f_{\text{eff}})$ (D).
Temporal characteristics of call-evoked LFPs and SU activity

The call-evoked LFP is shown here to have a number of relatively high-amplitude positive and negative components of various durations that last from 10 to ~250 ms after stimulus onset (the early and middle phases of the evoked response). During the late phase of the response, there are low-amplitude slow waves with superimposed oscillations of different frequencies that gradually attenuate over 250–600 ms time window. Thus the generic temporal structure of call-evoked LFP resembles the “classical” middle-latency auditory evoked potential (MAEP) recorded in response to clicks and pure tones in the auditory cortex of many mammalian species, including rats, cats, monkeys, and humans (Eggermont and Smith 1995; Franowicz and Barth 1995; Lee et al. 1984; Makela et al. 1990; Musiek et al. 1984). The click-evoked MAEP complex has a uniform spatial distribution within the primary auditory cortex of the anesthetized rat and represents the archetypal response sequence reflecting activation of distinct subpopulations of cortical cells (Barth and Di 1990; Franowicz and Barth 1995). Similar to that, each call-evoked LFP has an archetypal waveform within the Alp area of the awake mustached bat. Nevertheless, the temporal structure of a call-evoked LFP seems to be more complex than the click-evoked or single tone-evoked MAEP complex because it contains multiple positive and negative peaks as well as slow waves and/or oscillatory components depending on the specific spectrotemporal structure of a corresponding call.

LFPs are commonly viewed as being associated with relatively slow (compared with the duration of an action potential) changes in the membrane potential. The major contributing factors to an LFP recorded anywhere in the cortex are 1) synaptic potential changes resulting from thalamocortical, cortico-cortical and other inputs at the recording locus; 2) slow membrane changes such as hyperpolarization and intrinsic oscillatory activity associated with action potentials and opening and closing of ion channels; 3) asynchronous activity related to nonspiking, subthreshold activity; and finally 4) global and/or persistent slow-wave activity across the cortex. Thus LFPs represent the summed event-driven activity at specific loci within the cortex (Eggermont and Smith 1995; Norena and Eggermont 2002). The fact that synaptic activity on multiple neurons within the Alp area contributes to a cortical representation of a complex sound in the LFP does not exclude a possibility that a subset of these neurons may converge on to single neurons at other (higher) levels of processing, such as the secondary auditory areas and/or the frontal cortex auditory field. Recordings from those levels could yield SU data that would show more specialized responses to social calls than what is observed within the Alp.

Response invariance to time reversal

Calls used in this study represent the basic elements (simple syllables) of the mustached bat’s vocal repertoire and are referred to as “syllables” by analogy with speech. Some of these syllables, especially the FM type, have a distinctive temporal structure that changes significantly with call reversal. For example, an upward FM sound becomes a downward FM sound with call reversal. An unexpected result of this study is that the LFP profile does not change significantly with call reversal. Nevertheless, it should not be concluded that neuronal clusters within the Alp area are completely insensitive to the temporal features of calls. The LFP profile remains similar when presented with the normal and reversed calls, but it is not identical in the two conditions. Small changes in the LFP waveform can be observed and they are likely to be related to the low-amplitude high-frequency components of the LFP. Because the main focus of this study is a relationship between spectral content of calls and the temporal structure of the LFP waveform, a detailed description of changes in the frequency-specific components of the LFP will be described elsewhere (unpublished data).

MDS of call responses

Calls of the mustached bat have been classified into three major types (CF, FM, and NB) and categorized on the basis of principle components and MDS analyses in the two- and three-dimensional representational space (Kanwal et al. 1994). The two-dimensional scaling of the evoked LFP and SU activity performed in this study leads to similar, although not identical, segregation of calls. Similarity in call classification between this study and the previous study is shown by the fact that the call grouping based on the LFP is also sensitive to the acoustic parameters of calls. For example, CF calls and calls with a relatively weak FM (e.g., calls 3 and 4) are segregated from calls with a pronounced FM component (e.g., calls 5 and 8) in both the LFP and PSTH representations. NB type of calls (e.g., call 14) are placed out of any cluster in either representation (Fig. 8). Our analysis of call responses based on temporal characteristics of the LFP waveforms reveals that certain aspects of the acoustic structure of a stimulus may be extracted by LFPs and SUs. Our data show that the temporal shape of an LFP evoked by a call is closely related to the spectral content of that call, in particular its overall duration and harmonic complexity, i.e., the fundamental frequency, the absolute and/or relative value of the predominant frequency, spectral discreteness, and stability.

When represented in the two-dimensional abstract space, the first dimension (x value) of the LFPs correlates both with the predominant (“formant”) frequency and the fundamental frequency of a call (“pitch”). Thus these data suggest that one of the acoustic parameters, which may be represented in the Alp area, is simply a predominant and/or fundamental frequency of a call. Another spectral parameter that shapes the LFP waveform is related to the harmonic structure of a call, namely, its “harmonic complexity.” We describe “harmonic complexity” of a call by two parameters that increase with the number of harmonics in the call spectrum and the spectral width. Those parameters are the “effective” frequency and its dispersion in the normalized spectrum of a call. The second dimension (y values) of LFPs in the two-dimensional space correlates well with both of those parameters (Fig. 11, C and D). These data suggest that the functional role of the Alp area may also include a harmonic analysis of calls.

Harmonic analysis of calls within the Alp area

Many naturally produced sounds have a clear harmonic structure. Harmonics play an important role in our ability to
distinguish different sounds. Harmonics give different sounds their quality and may help identify a sound source. Accordingly, spectral elements, such as harmonic frequencies, within calls can function as important information-bearing parameters for perception of a call type and in the detection of fine changes in the pitch of any one type of a call (Medvedev et al. 2002). Harmonic analysis of calls may also allow segregation of sounds with different fundamental frequencies when heard at the same time. Behavioral experiments show that the presence of two or three harmonics significantly improves the recognition of wriggling calls that release maternal behaviors in house mice (Ehret and Riecke 2002). Excitatory frequency tuning data in the mustached bat show that many neurons in the Alp area are tuned to harmonic frequencies with a low harmonic in the range of ~24 kHz and a high harmonic in the range of ~48 kHz (Peng and Kanwal 2001). Multipeaked tuning is also observed in the primary auditory cortex of cats, although in this case, frequencies are not as well separated and have not been shown to have a harmonic relationship (Sutter and Schreiner 1991). A recent study suggests that spectral harmonic templates emerge early in the auditory system (Shamma and Klein 2000). The model proposed by these authors demonstrates that harmonic templates can be derived from any broadband stimulus as a result of cochlear filtering followed by coincidence detection. Therefore the central auditory system is obviously exposed to sets of harmonic frequencies within complex sounds.

The harmonic structure of environmental and communications sounds is intimately related to one of the major psycho-physical characteristics of auditory perception, namely, the perception of pitch. Pitch is a function of the fundamental frequency of a complex sound and can be perceived even when the fundamental frequency itself is absent in the spectrum of a sound (case of the “missing fundamental”) (Schouten et al. 1962; Shamma and Klein 2000). The phenomenon of perception of the missing fundamental has been shown in bats as well (Preisler and Schmidt 1995). One of the most influential concepts regarding pitch perception is a theory for the central formation of pitch (Goldstein 1973). This theory is based on a hypothetical optimum processor performing statistical estimation of the best matching template from the periodic spectrum of a complex sound. Using an associative neural network to model perception of pitch, we have recently suggested that spectral templates in the central auditory system may be formed as a result of Hebbian-like associations between harmonically related frequency channels (Medvedev et al. 2002). Multi-frequency tuning of neurons in the Alp area to harmonically related frequencies may provide the neural substrate for such a mechanism. Our data on correlation of the LFP in the Alp area with the fundamental frequency of a social call are also suggestive that this part of the auditory cortex may be involved in the recognition of the fundamental frequency (“pitch”) of calls. This is not surprising in light of the recent finding of a topographic map of periodicity pitch in the brain of the Mongolian gerbil (Schulze et al. 2002).

Another acoustic parameter of calls defined here as “harmonic complexity” depends on the number of harmonics within the call spectrum and the width of the individual spectral peaks. “Harmonic complexity” can be related to the “fine” spectral structure of sounds, or timbre (Hartmann 1997). As evident from our MDS analysis, the LFP waveform shows a high correlation with the “harmonic complexity” of a call. This finding indicates that synaptic activity within the Alp area is most likely also involved in the computation of timbre of complex sounds.

Population coding of calls?

Single neurons within the unspecialized Alp area fire scarcely and probabilistically in coincidence with specific components of the LFP. In the absence of call-specialized neurons, it seems logical to look for neural representation of calls in the synaptic activity within small populations of neurons. The ability of Alp neurons to respond to many calls is probably related not to their individual properties but rather to their collective properties within a network. In this case, a network is able to respond to many calls. The “call-specificity” of the LFP waveform, however, is based on the network’s response to the unique spectral-temporal structure of a call. Thus in the Alp area, calls may be encoded in the activity of small populations of neurons and are relatively independent of the topographic location of this activity. This result corresponds well with the established sensitivity of the scalp-recorded evoked activity to spectrotemporal characteristics of speech sounds in humans (for a review, see Eggermont and Ponton 2002). In the visual domain, sparse population coding was demonstrated using MDS analysis of SU responses to faces in the inferotemporal cortex of monkeys (Young and Yamane 1992). This elegant result was obtained by summing the SU response of a population of neurons in the inferotemporal cortex. In our study, we show a similar coding of species-specific calls based on LFP data, which represents the response of a sparse population of neurons in the primary auditory cortex. This is exciting because it implicates a common model for processing complex visual and acoustic stimuli by cortical areas (Kanwal et al. 2004).

Previously proposed schemes for the representation of auditory information include rate coding and temporal coding at the level of single neurons and spatial representation (mapping) of various stimulus parameters (Cariani 1999; deCharmes and Zador 2000). An important example of spatial mapping is tonotopic representation of frequencies and pitch by anatomically distinct channels within the auditory system (Lukhonen 2003; Pantev et al. 1989). Our data with complex sounds invoke population activity that suggests that stimulus qualities such as pitch and timbre can be represented within the same or highly overlapping local networks of neurons and therefore do not require absolute spatial segregation of the encoded parameters. Instead, these parameters are encoded within specific temporal patterns of synaptic activity that can be extracted most effectively by the LFP waveform. Thus an LFP represents a rapid succession of synaptic activity (within 1–200 ms) in response to a call and leads to less robust, stochastic spiking activity within single neurons at the recording site. A mechanism for spatiotemporal integration of complex sounds based on transient synchrony within neural clusters has been suggested recently (Hopfield and Brody 2000, 2001). Transient synchronization of neuronal firing patterns in a set of neurons is achieved through temporal convergence of synaptic currents across those neurons as determined by the acoustic structure of a call. Our results showing a “call-specific” temporal structure of the LFP are consistent with this mechanism.
In conclusion, the results of this study suggest that call processing within the AIp area involves determination of the “fine” spectral structure of a call. This information may be used to distinguish calls from noncall stimuli as well as to discriminate between different simple syllabic calls. Most importantly, these functions appear to be carried out in the AIp area of the mustached bat without invoking specialized neural mechanisms that are so important for processing echolocation information (Suga 1965; Suga and Taniguchi 1983). Further studies are needed to elucidate the different roles of the highly specialized versus unspecialized cortical mechanisms for call processing in the mammalian brain.

ACKNOWLEDGMENTS

We thank the Ministry of Agriculture, Land and Marine Resources in Trinidad for permitting us to export mustached bats and P. Muradali for assistance with the collection and exportation procedures.

GRANTS

This research was supported by National Institute on Deafness and Other Communication Disorders Grant DC-02054 to J. S. Kanwal.

REFERENCES


Lutkenhoner B. Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. Audiol Neurootol 8: 222–233, 2003.


J Neurophysiol • VOL 92 • JULY 2004 • www.jn.org


